Volume 9 Number 18 1981 Nucleic Acids Research

Human immunoglobulin heavy chain genes: evolutionary comparisons of Cμ, Cδ and Cγ genes and associated switch sequences

T.H.Rabbitts, A.Forster and Celia P.Milstein*

Laboratory of Molecular Biology, The MRC Centre, Hills Road, Cambridge CB2 2QH, UK

Received 26 June 1981

ABSTRACT

Human immunoglobulin heavy chain constant region genes have been characterised in isolated clones. The human Cμ gene comprises discrete domains for Cμ1, Cμ2, Cμ3 and Cμ4 + tp separated by short intervening sequences. The Cδ gene has been located about 5 kb downstream of Cμ4. Furthermore, the coding segments for the membrane form of μ have been located 1.9 kb downstream of Cμ4. Tandemly repeated sequences implicated in the heavy chain class switch occur upstream of the Cμ and the Cγ genes, but none were detected near the Cδ gene. These tandem repeats are very homologous to those of mouse. Particularly common is the sequence G-A-G-C-T. These data suggest that the μ to γ switch in humans involves DNA rearrangements of the CH-genes and subsequent deletion of DNA, but that the coexpression of Cμ and Cδ genes results from different mechanisms.

INTRODUCTION

Antibodies are composed of light (L) and heavy (H) chains which can themselves be different types or classes. In humans there are five classes of H-chain defined by the constant (C) region sequences designated Cμ, Cδ, Cγ, Cα and Cε, which give rise to IgM, IgD, IgG, IgA and IgE respectively. During the differentiation of B-lymphocytes the form and class of immunoglobulin H-chain synthesised alters. The first heavy chain constant (CH) gene to be expressed is Cμ which gives rise to a membrane form of IgM [1]. This class can also be secreted into the serum in the form of a pentameric antibody. Individual B-cells can also express IgM together with a second CH-gene product, surface IgD [2,3] and, furthermore, the cells are able to switch the class of immunoglobulin to IgG or IgA [1,4,5]. The salient feature of these processes of combined or switched expression is that the same VH-gene is expressed with the Cμ, Cδ or a subsequent CH-gene (e.g. C) in the progeny of single cells. Further, cells can change from the production of membrane immunoglobulin to the secreted form. There are a number of reports on these processes in mouse cells. It is clear that μ_m and μ_s chains result from differential splicing of a nuclear RNA precursor molecule which encompasses

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

coding segments for both μ_m and μ_s [6,7]. The class switch from μ to γ or α occurs by a quite different mechanism involving chromosomal DNA rearrangement. As suggested by Honjo [8], this rearrangement results in deletion of the DNA segment between the expressed VH-gene and the next CH-gene to be activated [9-11]. The site at which this deletion event occurs in the class switch was shown to be located within the VH-Cμ large intervening sequence and the region flanking the 5' end of, for example, the Cα gene [12]. Sequence analysis of different mouse CH-genes shows that tandemly repeated sequences (particularly G-A-G-C-T) are located adjacent to CH-genes [13,14] and these have been implicated in the mechanism of the switch.

No information has hitherto been available on the analogous gene systems of man. Here we describe human genomic clones containing the Cμ, Cδ, and a Cγ gene, as well as some relevant sequences occurring within the intervening sequences.

MATERIALS AND METHODS

Selection and analysis of human immunoglobulin clones

Human genomic clones were isolated from a foetal liver phage library in lambda Charon 4A. Plaque hybridisation was carried out [15] to the library plated on E. coli 803 using 23 x 23 cm Nunc bioassay plates [16]. The hybridisation probes were pμ/118, C75p1.2 and pγ1/A5 which were nick-translated to a specific activity of approximately 4×10^7 c/m/ g [17,18]. Positively hybridising clones were selected for bulk growth. Phage from a plaque were taken in to LS broth + 10 mm $MgSO_4$, 10 mM Tris-HCl pH 7.5 and plated onto 15 cm diameter dishes. After overnight growth, the top agar was scraped into a Petri dish, the plates washed with a minimal volume of LS broth and finally the top agar + broth was stirred for 30 min in the presence of 1/10 volume of $CHCl_3$. The top agar was pelleted and the supernatant titred. One litre of LS broth + Mg and Tris was infected with an overnight growth of E. coli 803 (1/100 dilution), the cells were grown until OD_{600} 0.15 and infected with phage at a multiplicity of infection = 0.5. The culture was grown for 12 hr at 37° with vigorous aeration. Complete cell lysis was achieved by adding 1/10 vol $CHCl_3$ for 10 min at 37°. Phage were purified by CsCl centrifugation and DNA prepared by phenol extraction.

For recloning the EcoRI fragments of the lambda phage DNAs the unique EcoRI site of pACYC184 was utilised [19]. Lambda DNA was digested completely with EcoRI-cut pACYC184 in a molar ratio of 3:1. Recombinant clones were assayed for the presence of inserts by the rapid alkali/acid procedure [20].

Nucleotide sequencing was carried out by recloning fragments in the M13 phage vector [21] and direct priming on to the single-stranded phage DNA [22, 23]. Where necessary DNA was isolated from low melting temperature agarose as described [23].

Screening for recombinant M13 phage with various probes was conducted in one of two ways. The recombinant phage DNA was plated on the JM101 bacterial host, cellulose nitrate filter replicas of the plates made, and hybridised as described [15]. Alternatively, white plaques were picked into 1 ml of an exponential culture of JM101, grown for 4 hr at 37° and the cells centrifuged to leave the phage particles in the supernatant. 1 µl of the phage stock was transferred to a cellulose nitrate grid which was subsequently treated with 0.5 N NaOH (7 min), 1 M Tris-HCl pH 7.4 (2 x 1 min) and 1.5 M NaCl, 50 mM Tris-HCl pH 7.4 (5 min). The filter was allowed to air-dry, baked for 2 hr at 80° in vacuo and hybridised as above.

Preparation of RNA and blotting analysis

RNA was prepared from DAUDI cells as described [24] and the poly(A)$^+$ fraction isolated by oligo(dT) cellulose chromatography. cDNA was prepared from the poly(A)$^+$ fraction as previously described using pT_{10} as the primer [24]. Southern filter hybridisation [25] was carried out as described [10] using 10^6 c/m/ml ^{32}P-cDNA or 5 x 10^5 c/m/ml nick-translated plasmid DNA.

RESULTS

Characterisation of human Cµ constant region gene segments

A genomic clone (designated λC75) containing Cµ coding segments has been isolated from a human phage library [26] using as a probe a clone of mouse Cµ cDNA, pµ/118 [27]. The coding region of the human Cµ gene occurs in two EcoRI fragments of 1.2 kb and 0.92 kb. Nucleotide sequencing of the 1.2 kb fragment (see below) showed the presence of the Cµ1 domain, and restriction mapping (data not shown) allowed the ordering of fragments shown in Figure 1. Thus in this clone, λC75, the Cµ coding segment is flanked by EcoRI fragments of 5.6 kb on the 5' side and of 10 kb on the 3' side. All the EcoRI fragments were subcloned in the plasmid pACYC184 and a clone containing the 1.2 kb fragment (C75p1.2) was used to reprobe the human phage library. A second clone was obtained (λC76), the restriction map of which is also indicated in Figure 1. In clone λC76 the 3' side of the Cµ gene only extends for 1.5 kb whilst the 5' side consists of an 8 kb fragment.

The location and orientation of the Cµ coding segments within λC75 was determined by nucleotide sequencing using the M13 cloning/sequencing

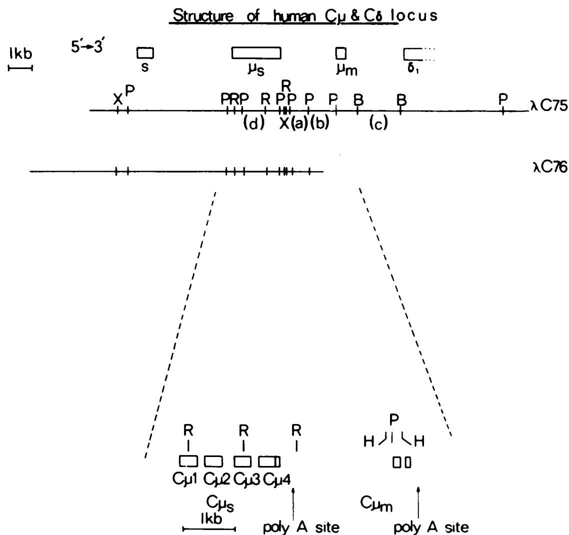

Figure 1. A restriction map of isolated clones containing the human Cμ and Cδ gene segments. The orientation of the Cμ coding region is that the 5' end is shown on the left hand side of the drawing. Restriction sites are: R = EcoRI, P = PstI, X = XbaI, B = BamHI or BstI and H = HpaII. (a) to (d) denote particular PstI restriction fragments. S refers to sequences related to the H-chain switch. Sites shown in the lower portion are incomplete.

techniques [23]. Figure 2 displays the partial nucleotide data which demonstrates that the human Cμ gene consists of four domains (Cμ1, Cμ2, Cμ3 and Cμ4 + tp) which are split by short intervening sequences. The sequence of the Cμ4 domain and tp are continuous segments in the genomic DNA. At a distance of 101 residues from the TGA chain termination codon, the sequence A-A-T-A-A-A occurs which characterises areas close to the poly(A) addition site [28]. This therefore approximates to that site of poly(A) addition in the the μ_s production. Table 1 enumerates the terminal amino acid codons of the human Cμ domains at the DNA level.

To investigate the presence of segments encoding membrane μ in the genomic clones, we hybridised Southern filters of digests from λC75 with cDNA made on mRNA prepared from DAUDI cells which are known to produce μ_m [29]. Figure 3 compares this hybridisation with a parellel filter hybridised to the cDNA plasmid pμ/118 which contains transcripts of only the μ_s form of the gene [27]. Both probes detect the 1.2 and 0.9 kb EcoRI fragments containing the $C\mu_s$ segments (panels A and B, slot 2) in λC75 but DAUDI cDNA also hybridised to the 10 kb EcoRI fragment to the 3' side of the 0.9 kb fragment.

Table 1. Domain structure of the human $C\mu_s$ gene

Genetic domain	First residue	Last residue
Cμ1	Gly(114)	Pro(216)
Cμ2	Val(217)	Pro(328)
Cμ3	Asp(329)	Lys(433)
Cμ4 + tp	Gly(434)	Tyr(565)

The first codon of each domain is created by RNA splicing.
The number in brackets refers to the amino acid number according to sequence GAL [31].

Furthermore the DAUDI cDNA hybridised two Pst fragments (6 kb and 1.5 kb) whilst the pμ/118 probe detects only the smaller of these fragments (the fragment containing the $C\mu_s$ sequences) (panels A and B, slot 1).
The additional hybridisation properties of DAUDI cDNA reflects the detection of μ_m sequence in λC75. The presence of μ_m sequences was confirmed by nucleotide sequencing. HpaII and HaeIII digests of the 10 kb EcoRI fragment were cloned in the M13mp7 vector and sequenced [21-23]. Nucleotide sequences were converted to protein sequences in all six reading frames and these were compared to the sequence of mouse μ_m [7]. This facilitated the identification of clones containing the first human μ_m coding segment. Figure 4 displays the nucleotide sequence of this first μ_m coding segment and its derived amino acid sequence, as well as comparing the area of human and mouse Cμ4 from amino acid codon 539 through to the TGA chain termination codon: this area contains the internal splicing site, at codon 546, for linking the $C\mu_s$ gene with the μ_m coding segments.

The location of the μ_m coding segments in the human DNA was determined by restriction mapping. Figure 5 displays the hybridisation properties of an M13 clone with the HpaII fragment (M13H5) spanning the first μ_m coding segment plus the 5' flanking sequence. This probe hybridises to the 10 kb EcoRI fragment of λC75 (slot 3) and a large 25 kb BamHI fragment (slot 2) originating from the left hand end of λC75 (including 14.3 kb of λ4A vector left arm). This hybridisation was confirmed by the ability of M13H5 to react with a 4.5 kb Bam fragment of the plasmid C75p10 (Fig 5, slot 4) which contains the left hand RI-Bam fragment from the 10 kb RI band plus a piece of pACYC184 DNA. Furthermore Figure 5, slot 1, shows that the M13H5 probe detects two Pst fragments (6 kb and 1.2 kb) of which the 6 kb fragment was

```
CH1

A.   G S A S A P T L F P L V S C E N S
     TCAGGGAGTGCATCCGCCCCAACCCTTTTCCCCCTCGTCTCCTGTGAGAATTCC
        ▲                                            R

CH2

B.   V I A E L P P K V S V F V P P R D G F F G N P R K S K L I C Q A T G F S P R Q I
     CAGTGATTGCTGAGCTGCCTCCCAAAGTGAGCGTCTTCGTCCCACCCCGCGACGGCTTCTTCGGCAACCCCCGCAAGTCCAAGCTCATCTGCCAGGCCACGGGTTTCAGTCCCCGGCAGA  120
        ▲       AI

     Q V S W L R E G K Q V G S G V T T D Q V Q A E A K E S G P T T Y K V T S T L T I
     TTCAGGTGTCCTGGCTGCGCGAGGGGAAGCAGGTGGGGTCTGGCGTCACCACGGACCAGGTGCAGGCTGAGGCCAAAGAGTCTGGGCCCACGACCTACAAGGTGACCAGCACACTGACCA  240

     K E
     TCAAAGAG

CH2

C.   D H R G L T F Q Q N A S S M C V P
     GATCACAGGGGCCTGACCTTCCAGCAGAATGCGTCCTCCATGTGTGTCCCCGGTGAGTGACCTGTCCCCAGGGGCAGCACCCACCGACACACAGGGGTCCACTCGGGTCTGGCATTGCCA  120
     SA                                              H2▲

     CCCCGGATGCAGCCATCTACTCCCTGAGCCTTGGTTCCCAGAGCGGCCAAGGGAGGGGCTCGGGCGGCAGGACCCCTGGGCTCGGCAGAGGCAGTTGCTAC/CTTTGGGTGGGAACCATG  240
       H2                                        H3

                                                   CH3
                                                      D Q D T A I R V F A I P P S F A S I F L T K S
     CCTCCGCCCACATCCACACCTGCCCCACCTCTGACTCCCTTCTCTTGA-TCCAGATCAAGACACAGCCATCCGGGTCTTCGCCATCCCCCCATCCTTTGCCAGCATCTTCCTGACCAAGT  360
                                                           ▲SA

     T K L T C L V T D L T T Y D S V T I S W
     CCACCAAGTTGACCTGCCTGGTCACAGACCTGACCACCTATGACAGCGTGACCATCTCCTGG

CH3

D.   N S G E R F T C T V T H T D L P S P L K Q T I S R P K
     GAATTCGCGGGAGAGGTTCACGTGCACCGTGACCCACACAGACCTGCCCTCGCCACTGAAGCAGACCATCTCCCGGCCCAAGGGTAGGCCCCACTC
      R                                                                                 ▲

CH4

E.   E Q L N L R E S A T I T C L V T G I S P A D V F V Q W M Q R G Q P L S P E K Y V
     GAGCAGCTGAACCTGCGGGAGTCGGCCACCATCACGTGCCTGGTGACGGGCTTCTCTCCCGCGGACGTCTTCGTGCAGTGGATGCAGAGGGGGCAGCCCTTGTCCCCGGAGAAGTATGTG  120
                             H3                                                                                      H2

     T S A P M P E P Q A P G R Y F A H S I L T V S E E L W N T G E T Y T C V A H D A
     ACCAGCGCCCCAATGCCTGAGCCCCAGGCCCCAGGCCGGTACTTCGCCCACAGCATCCTGACCGTGTCCGAGGAGGAATGGAACACGGGGGAGACCTACACCTGCGTGGCCCATGACGCC  240
                                 H3    H3                                                                        H3

     L P N R V T E R T V D K S T G K P T L Y N V S L V M S D T A G T C Y *
     CTGCCCAACAGGGTCACCGAGAGGACCGTGGACAAGTCCACCGGTAAACCCACCCTGTACAACGTGTCCCTGGTCATGTCCGACACAGCTGGCACCTGCTACTGACCCTGCTGGCCTGCC  360
                                                                                                                     H3

     CACAGGCTCGGGGCGGCTGGCCGCTCTGTGTGTGCATGCAAACTAACCGTGTCAACGGGGTCGAGATGTTGCATCTTATAAAATTAGAAATAAAAGATCCATTCAAAAGATACTGGTCC  480

     TGAGTGCACGATGCTCTGGCCTACTGGGGCGGCGGCTGTGCTGCACCCACCCTGCGCCTCCCCTGCAGAACACCTTCCTCCACAGCCCCCACCCCTGCCTCACCCACCTGCGTGCCTCAG  600
                      H3                                             P

     TGGCTTCTAGAAACCCCTGAATTCC
                         R
```

detectable with DAUDI cDNA (Fig. 3). The sequence of the μ_m region shown in Figure 4 demonstrates the presence of a PstI site just to the 5' side of the first μ_m coding segment, indicating that the 6 kb and 1.2 kb fragments are contiguous in λC75. The location of these two Pst fragments relative to the 0.9 kb EcoRI fragment was determined by probing a PstI digest of λC75 with the 3' terminal 1.5 kb EcoRI fragment of λC76. This fragment detects the 1.2 kb Pst fragment plus a smaller fragment about 0.5 kb (data not shown) and additionally the M13H5 clone does not detect the 1.5 kb EcoRI fragment located at the 3' end of λC76 (Fig. 5, slot 5). The order of Pst fragments in the λC75 10 kb fragment appears to be (5'→3') 0.5 kb, 1.2 kb, and 6 kb, with the μ_m coding segments being located at the 5' end of the 6 kb fragment. The distance from the end of Cμ4/tp domain to the start of the μ_m domain is about 1.9 kb.

Location of the Cδ gene adjacent to the human Cμ gene

We have investigated the location of the human Cδ gene within the λC75 clone by adopting the following strategy. The 10 kb EcoRI restriction fragment lying adjacent to μ_s was purified from a low melting temperature agarose gel and digested with HinfI. These fragments were cloned in the M13 phage and randomly derived clones were sequenced. Nucleotide sequences were

Figure 2. Partial nucleotide sequences encoding the human genomic μ_S segment. (a) 5' end of the CH1 domain. EcoRI-HaeIII fragments from the plasmid C75p5.6 were cloned in M13mp7 and sequenced using the Bam fragment primer [30]. R = EcoRI. (b) 5' end of the CH2 domain. A PstI fragment of 1.5 kb was cloned for sequencing in M13mp2/Pst and AluI fragments cloned in M13mp7. The relevant AluI site is marked in the sequence (A1) and residues 13 to the 3' end of the fragment were sequenced with the PstI cloned fragment. (c) 3' end of the CH2 domain-intervening sequence-5' end of the CH3 domain. Sequencing of this area of the 1.2 kb EcoRI was carried out using the following enzyme digests to generate clones in M13mp7 (residue numbers indicate that area sequenced by a particular clone and the direction of reading). Sau3A (SA) residue 1-195, 296-10; HpaII (H2) 51-121, 124-225; EcoRI (R) 422-304; HaeIII (H3) 166-394, 422-387 (latter site not in Figure). (d) 3' end of the CH3 domain. The area shown was sequenced after inserting the 0.9 kb EcoRI (R) fragment into M13mp7. (e) 3' end of CH4/tp plus 3' untranslated region of μ_S.

The sequenced regions using various enzymes were: HpaII (H2) 1-97, 110-156; HaeIII (H3) 24-125, 150-25, 228-155, 147-159, 353-230, 232-355, 380-498, 498-380; EcoRI (R) 620-340; PstI (P) 543-620, 620-543.

In this Figure arrows indicate RNA splicing positions.

These data show that the 1.2 kb EcoRI fragment contains 258 bases of Cμ1, IVS between Cμ1 and Cμ2 is about 130 bases, Cμ2 is 330 bases, IVS between Cμ2 and Cμ3 is 242 bases and 237 bases of Cμ3 occur in this fragment. The 0.9 kb EcoRI fragment contains the remains of Cμ3 (80 bases) followed by an IVS between Cμ3 and Cμ4 of about 150 bases. CH4/tp region is 387 bases and the 3' untranslated region is 101 nucleotides to the A-A-T-A-A-A (boxed) sequence.

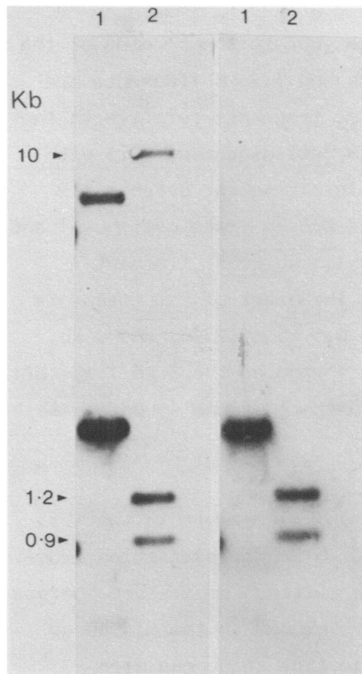

Figure 3. Detection of μ_m coding segments in λC75 using DAUDI cDNA as the probe. λC75 DNA (0.5 μg) was digested to completion with EcoRI or PstI and the digests fractionated on 0.8% agarose gel. The gel was stained to allow visualisation of the DNA fragments and molecular weight markers (phage lambda DNA cut with EcoRI and phage G4 cut with HaeIII), transferred to a cellulose nitrate filter [25]. After transfer the filter was cut into two, one half being hybridised to DAUDI ^{32}P-labelled cDNA (panel A) and the other to nick-translated pμ/118 (panel B). Slots 1 = PstI; slots 2 = EcoRI. The salt conditions for hybridisation were 6 x SSC 0.2% Ficoll 400, 0.2% PVP 360, 0.2% BSA, 0.1% SDS, 10 μg/ml poly(A) and 50 μg/ml sonicated, denatured salmon sperm DNA at 65°C for two days and post-hybridisation washes were carried out with 1 x SSC 0.1% SDS at 65°C (6 x 45 min).

converted to amino acids in all six reading frames using a computer program designed also to compare these derived protein sequences (R. Staden, personal communication) with the known sequence of a human IgD (Er1) myeloma heavy

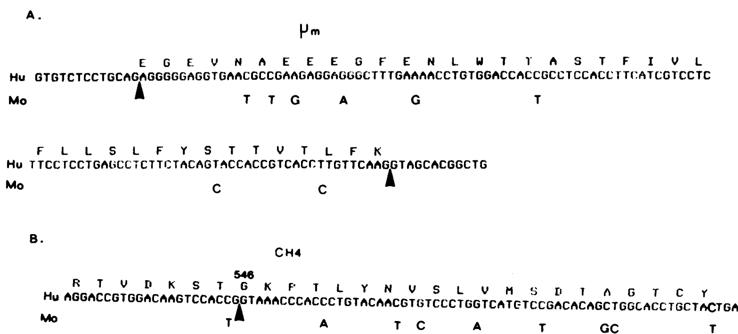

Figure 4. Comparison of nucleotide sequences of human and mouse μ_m coding segments. The nucleotide sequence of (a) the first coding segment of human μ_m and (b) the intra-C 4 splicing position (codon 546) are compared between the human (this paper) and mouse [7]. The arrows signify splicing positions. HU = human sequence, MO = bases different in mouse.

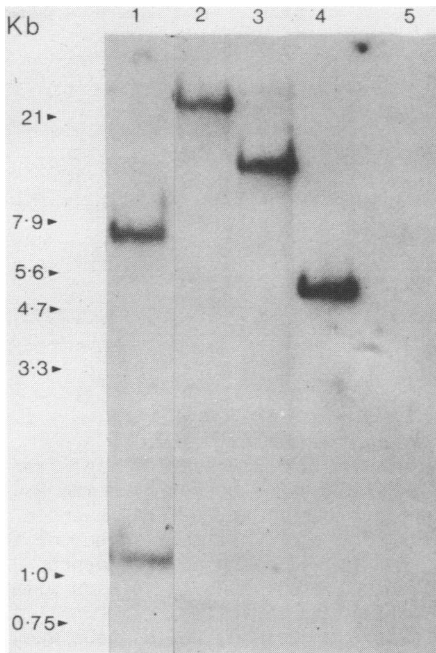

Figure 5. Evaluation of the distance between μ_s and μ_m segments in human DNA. Various restriction enzyme digests of λC75, λC76 and C75p10 were separated on 0.8% agarose blotted and hybridised (as in Fig. 3) with M13H5 DNA. Size estimations were made using lambda cut with EcoRI and phage G4 cut with HaeIII.

Slot 1 - λC75 cut with PstI
Slot 2 - λC75 cut with BamHI
Slot 3 - λC75 cut with EcoRI
Slot 4 - C75p10 cut with BamHI
Slot 5 - λC76 cut with EcoRI

chain protein [32]. One of the Hinf clones (M13HF28) was found to contain the 5' end of the Cδ1 domain plus part of the intervening sequence between μ_m and Cδ. Figure 6(a) shows the nucleotide sequence of M13HF28 plus the derived peptide sequence compared with the sequence of the protein Er1. The Cδ1 sequence of the DNA ends at the junction between JH and Cδ1 and at this point in the DNA the sequence characteristic of RNA splicing point (CAG) occurs [33]. Thus, like the Cμ1 domain, the 5' end of the Cδ1 domain can splice to the 3' end of a JH segment in nuclear RNA.

The linkage between the Cμ and Cδ genes in the human genomic clones was investigated by restriction mapping using M13HF28 (i.e. Cδ1) as the probe in Southern hybridisation (Fig. 6b). The plasmid C75p10 which contained the 10 kb EcoRI fragment of λC75 was digested with BamHI plus EcoRI and hybridised with the Cδ1 probe (4.5 and 2.6 kb) allowing the location of Cδ1 in the restriction map (Fig. 1) about 5.5 kb downstream of $C\mu_s$.

Tandemly repeated sequences in regions important for the class switch

In the area of DNA where the class switch occurs tandemly repeated sequences of the type G-G-G-G-T$(G-A-G-C-T)_n$ are located near mouse Cμ and Cγ1[13] and their presence in these regions implies a role in the switch. To deter-

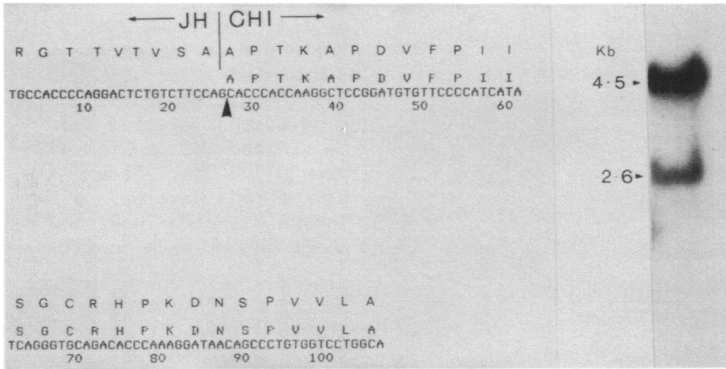

Figure 6. Identification and mapping of the Cδ1 domain with λC75.
LHS: Sequence of NH_2-terminus of Cδ1. The nucleotide sequences of Hinf fragments (from λC75) in M13mp7 were compared to the known sequence of human Cδ. The sequence of one clone M13HF28 which contains the 5' end of Cδ1 is shown compared to the J/Cδ1 junction of human Er1. The top line is Er1 sequence in single letter code and the second line is the derived amino acid sequence from M13HF28. The arrow refers to the splice point which could join JH with Cδ1 and the boundary of JH with Cδ in Er1 protein is indicated.
RHS: Mapping of 5' end of Cδ1. The clone C75p16 containing the Cδ gene was digested with EcoRI + BamHI to generate the mixed digest fragments displayed in Figure 1. The digest was fractionated on 0.8% agarose, transferred to nitro cellulose and the filter hybridised with nick-translated M13HF28 RF (Cδ1 probe).

mine the location of tandem repeats of this type near human CH-genes we have investigated the cloned Cμ genes (λC75 and λC76) and a cloned Cγ gene (λHγ4-6) using the S-probe from mouse or a cross-hybridising fragment from λC75 (Fig. 7). Panel A shows the hybridisation of the mouse S-probe (pγ1/IF2P1.8) to EcoRI digest λC75 demonstrating the hybridisation to the 5.6 kb EcoRI from the 5' side of Cμ. No hybridisation was observed with the 10 kb EcoRI fragment containing the Cδ gene. Clearly an S sequence complementary to the mouse probe occurs within 5 kb of the human Cμ gene. The 5.6 kb EcoRI fragment was subsequently cloned in pACYC184 (clone C75p5.6) and used as a human S sequence probe to verify the absence of a similar sequence in the locality of the Cδ gene. Figure 7, panel B, shows that the C75p5.6 detects sequences in the 8 kb left hand fragment of λC76 either in EcoRI-digested λC75 (slot 2), in the isolated 8 kb EcoRI fragment (slot 3), and in the homologous 5.6 kb EcoRI fragment from λC75 (slot 1). However, this probe does not detect sequences in the 10 kb EcoRI fragment from λC75 either in the whole digest or the the isolated fragments (slots 1 and 4). The strength of the hybridisation

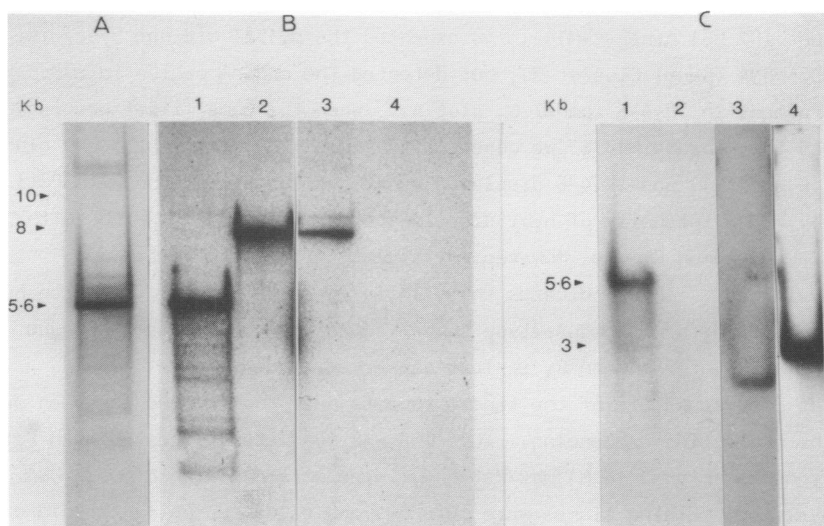

Figure 7. Hybridisation of mouse and human tandem repeat sequences to genomic clones containing human Cμ, Cδ and Cγ genes and mouse Cγ. The digests of the clones or isolated EcoRI fragments were blotted from 0.8% agarose gels onto cellulose nitrate and hybridised to nick-translated plasmids: Panel A, pγ1/IF2P1.8, Panel B, C75p5.6, Panel C, slots 1 and 3, M13/63, slots 2 and 4, pγ1/A5. The DNA components on each panel were:
Panel A - EcoRI cut λC75
Panel B - Slot 1, EcoRI cut λC75
Slot 2, EcoRI cut λC76
Slot 3, isolated 8 kb EcoRI fragment from λC76
Slot 4, isolated 10 kb EcoRI fragment (Cδ) from λC75
Panel C - Slots 1/2, EcoRI cut λC75
Slots 3/4, PstI cut λHγ4-6

signals between λC75 and either the homologous probe (C75p5.6) or the mouse S-probe (pγ1/IF2P1.8) are very comparable, indicating that the sequences detected by the two probes are highly homologous. Even long term autoradiography of hybridised filters, however, failed to reveal any signal to the 10 kb EcoRI Cδ-containing fragment, showing that no homology exists in this region to the human or mouse S-probes. Further, Figure 7 shows the hybridisation properties of a mouse S-probe with a genomic clone containing a human Cγ2 gene (λHγ4-6) [U. Krawinkel and T.H. Rabbitts, unpublished]. Figure 7, panel C, shows the result of hybridisation experiments of this clone with pγ1/A5 (a cDNA clone containing CH_2, CH_3 and 3' UT region of mouse Cγ1) or with M13/63 (a 677 base-pair Sau3A fragment cloned in M13mp2 containing the region of the tandemly repeated sequence G-G-G-G-T$(G-A-G-C-T)_4$ plus 1½ copies

of a 49 residue repeat sequence originating from the 5' flanking sequence of the mouse IF2 Cγ1 gene - γM14). As expected the pγ1/A5 did not hybridise with λC75 DNA (panel C, slot 2), but detected the coding region in a 3.1 kb PstI fragment in λHγ4-6 (panel C, slot 4). When the same filter was rehybridised to the S-probe M13/63 we observed hybridisation with the 5.6 kb clone (panel C, slot 1) and λHγ4-6 displayed a new hybridisation band with PstI (2.5 kb). The intensity of hybridisation between the mouse S-probe with human Cμ and human Cγ was, however, different in that the latter was relatively weak. This indicates that, while the S sequence adjacent to human Cμ is conserved for the mouse-type tandem repeat sequence, the analogous sequence adjacent to human Cγ is less conserved.

The precise nature of the tandem repeats near the Cμ gene has been determined by nucleotide sequencing. λC75 DNA was completely digested with PstI and fragments, cloned in M13mp2/Pst [34], were screened with C75p5.6 DNA. Positively hybridising clones were selected and sequenced using the dideoxy chain termination method. An area of sequence from one clone, M13/P27, is shown in Figure 8: this sequence shows that a tandemly repeated sequence of the basic form G-G-G-X-G$(G\text{-}A\text{-}G\text{-}C\text{-}T)_n$ occurs within 100-200 residues of the PstI site (i.e. about 3 kb from the 5' end of the Cμ gene) and is repeated in the sequence more times than could effectively be read from the sequencing gel. The human repeat sequence is compared with that found near mouse μ. In the mouse the basic unit of repeat is the 20 base long sequence G-G-G-G-T$(G\text{-}A\text{-}G\text{-}C\text{-}T)_3$. Although very homologous to this basic unit, the human repeat shows greater heterogeneity of unit length and divergence of the fourth G

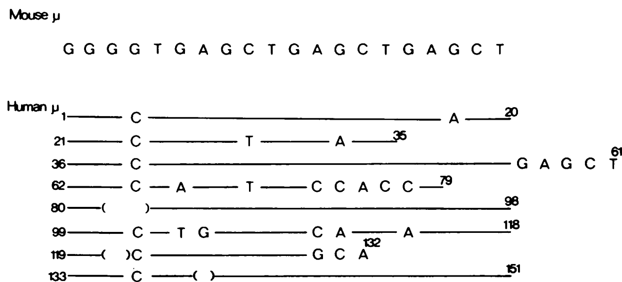

Figure 8. Comparison of the nucleotide sequence of the tandemly repeated units found adjacent to the human Cμ with those found in mouse DNA. The nucleotide sequence of M13P27 is compared with a repetitive unit [G-G-G-G-T $(G\text{-}A\text{-}G\text{-}C\text{-}T)_4$] occurring near mouse Cμ. The continuous line indicates sequence identity and brackets represent the positions of base deletions in the human sequence relative to that of mouse.

residue (i.e. to give G-G-G-C-T). On the whole the most common feature is the G-A-G-C-T sequence which recurs over a wide area of the M13/P27 clone.

DISCUSSION

Structure of human μ_s and μ_m segments

The human Cμ gene as previously shown[35,36], is very similar to that of mouse and we now also describe strong similarities in the μ_m of both DNAs. Each domain of the human Cμ gene is separated by a short intervening sequence and there is no intervening sequence between Cμ4 and tp. The first and last amino acids of each domain of Cμ are summarised in Table 1. The 5' end of the Cμ1 domain is found at the codon for residue 114 at which point in the nucleotide sequence an RNA splicing signal occurs. We have previously shown that the 3' end of human VHIII genes occurs at the codon for residue 94 [16] so that the DH/JH region of human DNA encompasses codons 95-113. The structural features of the human $C\mu_s$ gene segment and the VH segments show that mechanistically VH/D/J joining and $VDJC\mu_s$ expression are likely to occur by similar means to those genes of mouse. From the data presented in this paper it also appears that formation of membrane IgM occurs by analogous means in the two species.

In mouse, coding segments for μ_m have been described about 2 kb downstream of μ_s [7]. We have found homologous sequences in human DNA at roughly the same distance (about 1.9 kb) from μ_s by hybridisation to cDNA made from DAUDI cells which produce membrane-bound IgM (Fig. 3). The nucleotide sequence of the first human μ_m coding segment (Fig. 4) shows that the derived amino acid sequence is identical with that of the mouse, whilst silent base changes have allowed the drift of 7% at the nucleotide level. The intervening sequences on either side of the coding segment, however, show much greater sequence drift. Comparison of the tp shows that there are two amino acid changes (8% difference in the area compared) and eight base changes (9.6% difference). In addition the RNA splicing signals at both ends of the sequence are conserved between species, implying the existence of a second μ_m coding segment in man. The evolutionary conservation of the μ_m coding segment shows a strong selective pressure to maintain this peptide sequence: it has been postulated that this peptide resides in the membrane as an anchor [7] but also the necessity for functional interaction with membrane components may dictate the need for conservation. Therefore, the transcription of the human μ gene, like the situation first described in mouse, involves two alternative strategies from a single nuclear RNA precursor molecule. In

order to secrete IgM, the precursor molecule is spliced to join the Cμ1, Cμ2, Cμ3 and Cμ4 + tp domains using a proximal poly(A) addition site whilst to make surface IgM the precursor molecule is spliced to join Cμ1, Cμ2, Cμ3 and Cμ4 but in this case the splicing site within Cμ4 is utilised to join the 3' end of Cμ4 to the first μ_m coding segment. This segment itself would then be spliced to the second μ_m segment to generate the μ_m mRNA configuration.

Cμ and Cδ genes are closely linked in human DNA

In the results section we described the identification of a clone containing the 5' end of the Cδ1 domain as well as its location at about 5 kb downstream of Cμ4. The sequence of the region showed that the terminus of the Cδ1 chain is Pro codon CCC; adjacent to this the nucleotide sequence C-A-G-C-A occurs. This represents an RNA splicing site between G-C and thus would create GCA (Ala) as the codon resulting from the splicing with JH. These data show that the Cδ1 domain is very analogous to that found for the mouse [37] and, in addition, the homology to mouse $C\mu_m$ and $C\mu_s$ further strengthens the similarity between the two species in this area of DNA.

Switch recombination signals near human CH-genes

There are tandemly repeated sequences in the intervening sequences between JH and Cμ and flanking regions of Cγ genes which have been implicated in the H-chain switch possibly by mediating staggered sequence alignments which could facilitate homologous recombination (e.g. by sister chromatid exchange [38]. Thus the integrated VH-gene can switch between different CH-genes. Accompanying this switch is a deletion of DNA residing between the switch points. An argument in favour of a functional significance for the tandemly repeated sequences is the presence of similar sequences adjacent to human and mouse CH-genes[35,36]. We have identified a rather well conserved sequence [G-G-G-C-G(G-A-G-C-T)$_n$] about 3 kb to the 5' side of the human Cμ gene which is very homologous to a sequence [G-G-G-G-T(G-A-G-C-T)$_n$] previously located near mouse CH-genes. Our data also suggest that a similar but more divergent type of sequence occurs near to human Cγ genes. The presence of this conserved sequence between species in an area of functional significance argues for a functional role of these sequences in the H-chain class switch, i.e. the switch from μ to γ in human lymphocytes occurs by the switch recombination deletion as in the mouse.

The problem of the stable coexpression of μ and δ CH-genes by human B-lymphocytes does not seem to fall in the same category as the μ to γ switching. The Cδ gene is located only about 5 kb from the human Cμ gene whilst the most adjacent Cγ genes are a minimum of 15 kb apart. Further we

cannot detect any sequence related, to S sequences, adjacent to the δ gene using either the homologous human S sequence probe or the heterologous mouse probe (Fig. 7). The implication is that, if these sequences are important for the μ to γ class switch, the μ/δ switch cannot be operated by this switch recombination method. There are three principal ways in which μ/δ coexpression could be attained. Firstly a double integration could bring a VH segment into the transcription unit of Cμ and a separate transcription unit of Cδ. Secondly, μ to δ switching could occur by means of deletion in exactly analogous fashion to other CH-gene switching but accompanied by stable mRNA for μ to allow continued production of IgM. Thirdly it is possible that Cμ and Cδ coding regions are cotranscribed into one precursor nuclear RNA molecule in which differential splicing allows coexpression of these genes. The first possibility seems unlikely since we have been able to detect JH sequences upstream of Cμ but not upstream of Cδ or Cγ [T.H.R. and A.F., unpublished]. In addition double integration implies copying of a VH-gene since the Cμ and Cδ coexpress the same VH-gene. There are no precedents for this situation in the immunoglobulin genes. As regards deletion as a mediator of switching to δ expression we have found no evidence for switch recombination type of tandemly repeated sequences near the Cδ gene although they are present near Cμ and a Cγ gene. Thus if deletion of Cμ is a mechanism by which the VH-gene is brought to Cδ, the S sequence of Cδ must be radically different from that of Cμ and Cγ. The sequence difference must be sufficient great, in fact, (or involve a very short segment) that no sequence occurs anywhere between JH and Cμ which hybridises to the region of Cδ. A likely possibility for Cμ/Cδ coexpression, therefore, is the contranscription of the entire VH-Cμ-Cδ region into one nuclear precursor molecule. This would provide a third splicing alternative to the cell so that production of μ_s, μ_m and δ (δ_s and δ_m probably) would be within a single transcription unit.

*Permanent address: Institute of Animal Physiology, Babraham, Cambridge, UK

REFERENCES

1. Cooper, M.D., Kearney, J.F., Lydyard, P.M., Grossi, C.E. and Lawton, A.R. (1976) Cold Spring Harbor Symp. Quant. Biol. 41, 139-191.
2. Knapp, W., Bolhuis, R.L.H., Radl, J. and Hijmans, W. (1973) J. Immunol. 111, 1295-1298.
3. Pernis, B., Forni, L. and Luzzati, A.L. (1976) Cold Spring Harbor Symp. Quant. Biol. 41, 175-183.
4. Sledge, C., Fair, D.S., Black, B., Krueger, R.G. and Hood, L. (1976) Proc. Nat. Acad. Sci. USA 73, 923-927.

5. Wang, A.C., Wilson, S.K., Hopper, J.E., Fudenberg, H.H. and Nisonoff, A. (1970) Proc. Nat. Acad. Sci. USA 66, 337-343.
6. Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimore, D. (1980) Cell 20, 306-311.
7. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Cell 20, 313-319.
8. Honjo, T. and Kataoka, T. (1978) Proc. Nat. Acad. Sci USA 75, 2140-2144.
9. Cory, S. and Adams, J.M. (1980) Cell 19, 37-51.
10. Rabbitts, T.H., Forster, A., Dunnick, W. and Bentley, D.L. (1980) Nature 283, 351-356.
11. Coleclough, C., Cooper, D. and Perry, R.P. (1980) Proc. Nat. Acad. Sci. USA 77, 1422-1426.
12. Davis, M.M., Calame, K., Early, P.W., Livant, D.L., Joho, R., Weissman, I.L. and Hood, L. (1980) Nature 283, 733-739.
13. Dunnick, W., Rabbitts, T.H. and Milstein, C. (1980) Nature 286, 669-675.
14. Kataoka, T., Kawakami, T., Takahashi, N. and Honjo, T. (1980) Proc. Nat. Acad. Sci. USA 77, 919-923.
15. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
16. Matthyssens, G.E.A.R.and Rabbitts, T.H. (1980) Proc. Nat. Acad. Sci. USA 77, 6561-6565.
17. Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Nat. Acad. Sci. USA 72, 1184-1188.
18. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
19. Chang, A.C.Y. and Cohen, S.N. (1978) J. Bacteriol. 134, 1141-1156.
20. Birnboim, H.C. and Doty, J. (1979) Nucleic Acids Res. 7, 1513-1529.
21. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
22. Schreier, P.H. and Cortese, R. (1979). J. Mol. Biol. 129, 169-172.
23. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
24. Rabbitts, T.H., Forster, A., Smith, M. and Gilham, S. (1977) Eur. J. Immunol. 7, 43-48.
25. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
26. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) Cell 15, 1157-1174.
27. Matthyssens, G.E.A.R. and Rabbitts, T.H. (1980) Nucleic Acids Res. 8, 703-713.
28. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
29. Klein, E., Klein, G., Nadkarni, J.S., Nadkarni, J.J., Wigzell, H. and Clifford, P. (1968) Cancer Res. 28, 1300-1310.
30. Anderson, S., Gait, M.J., Mayol, L. and Young, I.G. (1980) Nucleic Acids Res. 8, 1731-1743.
31. Watanabe, S., Barnikol, H.U., Horn, J., Bertram, J. and Hilschmann, N. (1973) Z. Physiol. Chem. 354, 1505-1509.
32. Milstein, C.P. and Deverson, E.V. (1980) Immunology 40, 657-664.
33. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Nat. Acad. Sci. USA 75, 4853-4858.
34. Bentley, D.L. and Rabbitts, T.H. (1980) Nature 288, 730-733.
35. Ravetch, J.V., Kirsch, I.R. and Leder, P. (1980) Proc. Nat. Acad. Sci. USA 77, 6734-6738.
36. Takahashi, N., Nakai, S. and Honjo, T. (1980) Nucleic Acids Res. 8, 5983-5991.
37. Liu, C-P., Tucker, P.W., Mushinski, J.F. and Blattner, F. (1980) Science 209, 1348-1353.
38. Rabbitts, T.H., Hamlyn, P.H., Matthyssens, G.E.A.R. and Roe, B.A. (1980) Can. J. Biochem. 58, 176-187.