Cloning and sequence of the cDNA corresponding to the variable region of immunoglobulin heavy chain MPCH1

Rina Zakut, Justus Cohen and David Givol

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot, Israel

Received 5 June 1980

ABSTRACT

Poly(A)-containing mRNA from mouse myeloma MPCll was transcribed into cDNA which was cloned in the PstI site of the plasmid pBR322. The transformants were screened by hybridization with a cDNA fragment, derived from plasmid $py(11)^7$, corresponding to the 5' portion of the constant region of MPCll heavy chain. Several positive transformants were found to contain various lengths of the variable region of the heavy chain. We describe the structure and sequence of one of these clones, $pV(11)^2$, which contains CDNA corresponding to the entire variable region of MPCll heavy chain and extends to codon 248 in the constant region. The protein sequence deduced from the DNA sequence indicates that the variable region of MPCll heavy chain contains 121 amino acids and belongs to subgroup II of mouse heavy chains. Comparison of this sequence with other heavy chain sequences suggests a J (joining) segment of 16 residues which overlaps five residues of the third hypervariable region. The cDNA sequence shows that there is no discontinuity between the end of the variable region and the beginning of the constant region.

INTRODUCTION

Immunoglobulins are composed of heavy and light chains. Evidence has been presented to show that the light chain is encoded by four distinct gene segments, leader (L), variable (V), joining (J) and constant (C), which are separated by intervening sequences in the germline DNA (1,2). A similar organization is probably present in the heavy chain gene (3,4). During lymphocyte differentiation DNA rearrangement juxtaposes V and J segments which together encode the entire V region. The same V region can join different C segments at different stages of lymphocyte differentiation to form the complete heavy chain gene (3,4).

Mouse myeloma MPC11 produces an immunoglobulin heavy chain of the $\gamma_{2}b$ subclass. The structure of the mRNA and the embryonic gene for this heavy chain constant region havebeen reported (5-8). A number of variants in heavy chain production are known for this cell line. Some of them show short heavy chains (9) or a recombinant heavy chain which contains a portion of the γ_2 a

constant region (10). Other variants lost the ability to synthesize heavy chain (11). It is also likely that variants or mutants in the variable region can be obtained like in other myeloma lines (12).

The study of both V and C genes is essential for our understanding of lymphocyte differentiation and formation of variants in V or C region. Schibler et al. (13) cloned the cDNA of MPCll heavy chain constant region in plasmid pMB9. This plasmid, $py(11)^7$, contains cDNA from codon 150 to the Cterminal residue and includes also the ³' nontranslated region (5,7). We used a purified fragment of the cDNA insert of $py(11)^7$ to probe several hundred clones which were obtained by cloning cDNA of MPCll mRNA in pBR322 and found several clones which contained the entire V region of the heavy chain including part of its leader sequence. We report here the sequence of the variable region of MPCll heavy chain cDNA which was derived from one of these clones.

MATERIALS AND METHODS

Preparation of cDNA clones. Mouse myeloma MPCll was obtained from Dr. M. Potter (NIH). Poly(A) rich mRNA was prepared from tumors by the method of Kirby (14) followed by oligo(dT)-cellulose chromatography and sucrose gradient centrifugation. The 16S fraction was pooled and used for the preparation of cDNA. 5 µg of mRNA was added to a 0.1 ml incubation mixture containing 50 mM Tris-HCl (pH 8.3) 9 mM $MgCl₂$, 0.5 mM dithiothreitol (DTT), 60 mM KC1, 5 µg/ml oligo(dT)₁₂₋₁₈ (P.-L. Biochemicals), 1 mM of the unlabeled deoxynucleoside triphosphates dATP, dCTP, dGTP and dTTP in the presence of 200 μ Ci $3H$ -dATP (29 Ci/mmole), 4 mM sodium pyrophosphate and 54 u of AMV reverse transcriptase (obtained from Dr. Beard, Life Science Inc., St. Petersburg, Fla.). After 60 min of incubation at 42°, NaOH was added to a concentration of 0.3 N and the reaction mixture was heated at 70° for 20 min, neutralized and extracted with phenol. Sodium dodecyl sulphate (SDS) was added to 0.5% and the solution was passed through a Sephadex G-50 column equilibrated in 10 mM Tris-HCl (pH 8) and ¹ mM EDTA. The excluded fraction was made 0.3 M in sodium acetate (pH 6) and precipitated by ³ volumes of ethanol without the addition of carrier. The yield of cDNA was 1.1 µg. This cDNA was used as a template for a second strand synthesis in a 0.1 ml reaction mixture containing 50 mM Tris-HCl (pH 7.8), 9 mM MgCl₂, 0.5 mM DTT, 1 mM of the four unlabeled deoxynucleoside triphosphates, 30 μ Ci 32 P-dATP (300 Ci/mmole) and 105 u of AMV reverse transcriptase. After 2 h at 42° the reaction was terminated by addition of SDS to 0.5% and EDTA to ¹ mM, followed

by phenol extraction and Sephadex G-50 fractionation. The excluded fraction contained 2 µg DNA which was precipitated by ethanol.

The double-stranded cDNA was digested with the single strand specific S1 nuclease (Boehringer) to generate blunt ended molecules. The cDNA was dissolved in 0.1 ml containing 30 mM sodium acetate (pH 4.5), 0.25 M NaCl, 1 mM ZnSO₄ and 25 u of S1 nuclease. After 20 min at 37° the reaction was terminated by adding Tris-HCl (pH 8.0) to 100 mM, extracted with phenol followed by Sephadex G-50 filtration, and ethanol precipitation.

The double stranded cDNA (300 ng) was dissolved in water and adjusted to 0.14 M sodium cacodylate (pH 7.6), 30 mM Tris, 100 μ M DTT, 1 mM CoCl₂ and 40 μ M 32 P-dCTP (2 Ci/mmole) in a 50 μ l reaction mixture and 10 u of calf thymus terminal deoxynucleotidyl transferase (P-L Biochemicals) was used to add 30-SO dC residues per 3' end. pBR322 DNA was cleaved with PstI and similarly treated except that $32P$ -dGTP was used to add 30-50 dG per 3' end. The tailing reactions were incubated at 37° for 90 sec and terminated by adding SDS to 0.1% followed by 4 min centrifugation through a 3 ml Sephadex G-50 column in a syringe (Isolabs). The recovered DNA was adjusted to 0.3 M sodium acetate and precipitated by ethanol.

The dG-tailed plasmid (75 ng) and dC-tailed cDNA (15 ng) were mixed together in 0.1 ml of annealing buffer (10 mM Tris-HCl pH 7.8, 100 mM NaCl, 1 mM EDTA), heated at 70° for 10 min and incubated at 48° for 2 h to allow for annealing of the tailed DNA, followed by ethanol precipitation. The DNA was dissolved in 0.1 ml of 0.1 M CaCl₂ and used to transform CaCl₂ treated E. coli HB101 to ampicillin-sensitive, tetracycline-resistant colonies according to Enea et al. (15).

Screening of cDNA clones. Tetracycline-resistant, ampicillin-sensitive transformants of HB101 were transferred to nitrocellulose filters which were processed according to Grunstein and Hogness (16). The filters were hybridized to a ²²P-labeled cDNA fragment of the γ_2 b heavy chain recombinant plasmid p $\gamma(11)^7$, kindly provided by Dr. R. Perry. This plasmid contains cDNA of most of the constant region of γ_{2} b heavy chain and was previously characterized by us (7). Bam H1-HpaII digestion of the insert of $py(11)^{7}$ produces a fragment containing the DNA sequence between codons 158 and 303 of the heavy chain constant region (see Fig. ¹ and ref. 7). This fragment was isolated by electrophoresis on 1% polyacrylamide gel and labeled by filling in the ³' ends with DNA polymerase n-Klenow (Boehringer). Labeling was in a 20 μ l reaction mixture containing 6 mM Tris-HCl (pH 7.4), 6 mM NaCl, 6 mM MgCl₂, 6 mM 2-mercaptoethanol, 10 µCi of each of the four 32 P-deoxynucleoside

triphosphates and 1 u of polymerase. After 15 min at 37° the volume was increased to 50 μ 1 and the reaction mixture was extracted with phenol and passed through a Sephadex G-50 column.

Positive clones detected with this probe were picked and grown in 1L of medium containing tetracycline (25 μ g/ml). When the culture reached A₅₅₀=0.5, chloramphenicol was added to 15 μ g/ml and the culture was incubated with shaking overnight. The plasmid DNA was prepared by the clear lysate method as described (16).

All experiments with transformants were performed according to NIH guidelines in a P3 laboratory.

Sequencing of recombinant DNA. Two methods were used. In the first, end labeled fragments were sequenced as described by Maxam and Gilbert (17), except that the A reaction was performed in 0.1 M pyridine formate, pH 2.0. Fragments were labeled at the 5' end by kinase (Boehringer) and $32P$ - γ -ATP (Amersham), or at the 3' end by filling in protruding restriction sites with DNA polymerase n-Klenow using two or three 32 P-a-deoxynucleoside triphosphates (Amersham).

In the second method DNA fragments were cloned in bacteriophage M13 and sequenced by the dideoxy method (18) as described by Schreier and Cortese (19). Since none of our cDNA inserts contained HhaI restriction sites we have isolated the largest fragment produced by HhaI digestion of the plasmid from 1% agarose after electrophoresis. This fragment was digested with PstI and cloned in a derivative of M13 mp2 (20) containing a single PstI restriction site, kindly provided by D.L. Bentley (MRC, Cambridge). The white plaques (insert-containing phage) were picked, grown in ¹ ml culture and the single-stranded DNA was isolated from the supernatant by polyethylene glycol precipitation and sequenced as described (19).

RESULTS AND DISCUSSION

Construction and selection of hybrid plasmids containing the variable region of MPCll heavy chain. Poly(A) rich mRNA was transcribed into cDNA with a yield of 20% using reverse transcriptase. The cDNA was isolated by gel filtration and was converted without size selection into double stranded DNA using reverse transcriptase with a yield of 80%. After S1 nuclease digestion,the DNA was tailed with oligo(dC) and the PstI cut pBR322 was tailed with oligo(dG) using terminal deoxynucleotidyl transferase. The sequence of the insert in one plasmid showed homopolymer tails of 35 and 45 nucleotides.

The tailed cDNA and plasmid were annealed and the mixture was used to transform $CaCl₂$ -treated HB101. 420 transformants were screened with the labeled Bam Hl-HpaII fragment derived from $py(11)^7$. This fragment (Fig. 1) contains cDNA corresponding to codons 158-303 of MPCll heavy chain and detects clones containing sequences within this region.

Five positive clones were identified, the bacteria were grown and the plasmids isolated. None of the inserts in these plasmids contains a HhaI restriction site. Hence the plasmids were digested with HhaI and the largest fragment of each clone was isolated on a 1% polyacrylamide gel. These fragments contain the cDNA insert, 23 bp and 313 bp of plasmid DNA at the 5' and ³' ends of the insert. Digestion of two of these inserts with Bam Hl or HpaII provided information on their structure (Table I). As is shown in Fig. 1 the py(11)⁷ insert contains one Bam H1 restriction site at codon 158 and two HpaII restriction sites at codons 303 and 447. The insert of $pV(11)^{1}$ contains one Bam H1 site and one HpaII site whereas the insert of $pV(11)^2$ contains one Bam Hl site and no HpaII site. This indicates that the insert of $pV(11)¹$ ends before codon 447 whereas the insert of $pV(11)²$ ends before codon 303 (Fig. 1). Hence the fragments produced by Bam Hl digestion indicate the

Fig. 1. Schematic representation of the cDNA inserts in $p\gamma(11)'$, according to ref. 7, and in $pv(11)^{1}$ and $pv(11)^{2}$. The Bam H1-HpaII fragment used for screening the transformants is hatched. The positions of restriction enzyme sites are given as the corresponding amino acids in the lower line.

 $py(11)$ ^T was derived from pMB9 and the HhaI insert contains 0.62 and 0.42 kb of plasmid sequence at the 5' and ³' end of the insert. pV(ll) plasmids were derived from pBR322 and the HhaI inserts contain 23 bp and 0.32 kb of plasmid at the ⁵' and 3' end of the insert. HpaII produces also a 0.25 kb fragment of plasmid DNA by cutting 50 bp from the PstI site in pBR322. This fragment is not given in the table. The size of the fragments is given in parentheses in kb.

extension of the cDNA towards the 5' end of the mRNA. Digestion of the inserts with Bam Hi and double digestion with Bam Hl-HpaII indicates that the 0.3 kb fragment of $pV(11)^{1}$ (Table I) extends from codon 158 (Bam Hl site) to codon 62 and sequence analysis of this fragment confirms this. On the other hand the 0.6 kb fragment produced by Bam H1 digestion of the $pV(11)^2$ insert extends from codon 158 (Bam Hi site) to the beginning of the V region, including part of the leader sequence. The other 0.6 kb fragment (Table I) of $pV(11)^2$ contains the cDNA sequence from codon 158 to codon 248, the dC-dG tail and 322 bp of plasmid DNA. Hence plasmid $pV(11)^2$ was selected for sequencing of the variable region.

DNA sequence of MPCll heavy chain variable region. Two methods were used for sequencing $pV(11)^2$. In the first, end-labeled restriction fragments produced by Bam Hl or Hinfl digestion were sequenced as described by Maxam and Gilbert (16). In the second a PstI digest of the insert-containing HhaI fragment of $pV(11)^2$ was cloned in Ml3mp2/Pst and sequenced as described by Schreier and Cortese (19) using the dideoxy method (20).

Fig. 2 gives the details of the restriction fragments sequenced by both methods. In the M13 cloning-sequencing method we expected 5 PstI fragments. Hence 20 white clones were picked, grown in ¹ ml medium and their DNA was isolated and sequenced; Fig. 2 shows the distribution of the single stranded clones obtained in M13mp2 and their orientation. No clone was obtained for the segment between the plasmid PstI site and the PstI site at codon 4 (Fig. 2), inspite of the fact that sequencing by the Maxam and Gilbert method showed that the PstI site in the plasmid was preserved. This segment

Hha I fragment of pV (II)²

Fig. 2. The restriction sites used in sequencing $pV(11)^2$. The PstI restriction fragments were cloned in the phage M13mp2/Pst and sequenced by the dideoxy method (19) as described in Materials and Methods. Arrows pointing to the right represent the coding strand and arrows pointing to the left, the non-coding strand. The numbers above and below the arrows represent the number of clones obtained by screening 20 clones. The other restriction sites were used in sequencing by the Maxam and Gilbert method (17).

contains 45 bp of dC-dG tail and approximately 45 bp of insert down to codon 4. Perhaps this type of structure is difficult to clone in the M13 vector because of the tail. The sequencing of the M13 clones was performed in the two orientations and the information from both methods provided the necessary overlaps.

The DNA sequence of $pV(11)^2$ showed that the CDNA insert extends to codon 248 in $C_{U}2$. The sequence of the constant region (data not shown) is in agreement with that published by Tucker et al. (5). Fig. 3 shows the V-C junction in MPCll heavy chain. The ³' end of the V region (Fig. 4) is followed by the sequence:

> GGC AAA ACA ACA CCC CCA TCA GTC Ala Lys Thr Thr Pro Pro Ser Val

which is the beginning of the C_H1 domain (5). This demonstrates that there is no discontinuity in the mRNA between the V and C domains (Fig. 3).

The sequence of the variable region is given in Fig. 4. This sequence indicates that V_H of MPC11 belongs to subgroup II (21) like the V_H of MOPC21 and MOPC104E. The homology between MPCll and MOPC104E is shown in Fig. ³ and

Fig. 3. The DNA sequence at the V-C junction. The sequencing was performed on a Bam Hi fragment labeled at the ³' end by filling in the restriction site with DNA polymerase n Klenow. The codons should therefore be read from right to left. Codon 122, GCC is numbered 114 in ref. 5. \star , marks the beginning (A) of the indicated sequence.

demonstrates that most of the replacement (19, including deletions, out of 36) are in complementarity determining regions (CDR, see ref. 21). The CDR3 in MOPC104E is four residues shorter than that in MPCll whereas CDR1 and CDR2 are similar in size. MPC11 V_H contains 121 residues and the J segment overlaps five residues of the third hypervariable segment (CDR3). The assignment

10 20 GAG GCT CAG CTG CAG CAG TCT GGA GCT GAG CTG GTA AGA CCT GGG ACT TCA GTG AAG ATC Glu Val Gln Leu Gln Gln Ser Gly Ala Glu Leu Val Arg Pro Gly Thr Ser Val Lys Met - - - - - - - - Pro - - - Lys - - Ala 30 40 TCC TGC AAG GCT GCT GGA TAC ACC TTC ACT AAC TAC TGG ATA GGT TGG GTA AAG GAG AGG Ser Cys Lys Ala Ala Gly Tyr Thr Phe Thr Asn Tyr Trp Ile Gly Trp Val Lys Glu Arg Ser - - - - - Asp - Tyr Met Lys - Gln Ser CDRl 50 60 CCT GGA CAT GGC CTT GAG TGG ATT GGA GAT ATT TAC CCT GGA GGT GGC TTT ACT AAC TAC Pro Gly His Gly Leu Glu Trp Ile Gly Asp Ile Tyr Pro Gly Gly Gly Phe Thr Asn Tyr His - Lys Ser $-$ - Asn - Asn Asn - Gly - Ser \sim \sim \sim $-$ CDR2 70 80 AAT GAC AAT TTG AAG GGC AAG GCC ACA CTG ACT GCA GAC ACA TCC TCC AGC ACA GCC TAC Asn Glu Asn Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Thr Ser Ser Ser Thr Ala Tyr- Gln Lys - - - - - - - - Val - Lys - - - - 90 100 ATC CAG CTC AGC AGC CTG ACA TCT GAG GAC TCT GCC ATC TAC CAC TGT GCA AGA GGG ATT Met Asp Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Ile Tyr His Cys Ala Arg Gly Ile - Gln - Asn - - - - - - - - - - - - - Yal - Tyr - - - - Asp Tyr $***$ 110 120 TAC TAC AAT AGT AGC CCC TAC TTT GAC TCC TGG GGC CAA GGC ACC ACT CTC ACA GTC TCC Tyr Tyr Asn Gly Asp Pro Tyr Phe Asp Ser Trp Gly Gln Gly Thr Thr Leu Thr Val Ser) - Trp - - - Val - - Ala - - - Val (D J CDR3 **TCA** Ser

 \overline{a}

Fig. 4. The nucleotide sequence of the cDNA insert in $pV(11)^2$ corresponding to the variable region of MPCll heavy chain. Only the coding strand is shown with the encoded amino acid sequence. This sequence was determined by the method of Maxam and Gilbert (17) on 5' or 3' end labeled Bam Hi, Hinfl or BstNl labeled fragments and by the dideoxy method (18) on PstI restriction fragments cloned in phage M13 (19). CDR, is complementarity determining, or hypervariable region (21). The sequence of MOPC104E (23) which is of the same subgroup is given for comparison under the amino acid sequence of MPC11. The assignment of D (****) and J (----) are according to ref. 22 and 24. This assignment is made by homology and the precise borders of D and J should be determined by analysis of embryonic genes.

of J is from a comparison with amino acid sequence data according to Schilling et al. (22) and from DNA sequence data of Early et al. (24). In their recent publication Early et al. (24) determined the sequence of V_H and J_H segments of a germline gene coding for S107 V_H . They showed that the V_H segment encodes amino acids 1-101 and the J_H segment encodes amino acids 107-123. By comparing these sequences to that of a rearranged gene from a myeloma, they concluded that an additional separate genetic element, D (diversity), which codes for amino acids 102-106 must be present between V_H and J_H . In MPC11, as deduced by homology, J segment includes residues $106-121$ like J_{H603} (24) or residues 107-121 like J_{H315} (24) and the D segment may include residues 100-105 or 100-106. The precise borders of these segments in MPCll must await analysis of embryonic V_H genes.

The D segment in MPCll probably includes 6 amino acids Ile-Tyr-Tyr-Asn-Gly-Asp (Fig. 4) whereas in MOPC104E it includes only ² amino acids (22). These two residues (Tyr,Asp) are present at the borders of D in MPCll (TyrlOl and Asp105, Fig. 4). The J segments of MPC11 and MOPC104E differ at two positions which are outside CDR3 and at two positions which are within CDR3. It is also noted that J segment of MPCll begins with Pro (residue No. 106) which is not present in any other heavy chain at this position (21). Hence the two V_H segments (of MOPC104E and MPC11, although belonging to the same subgroup join different D and J segment. Since J and D contain a significant part of CDR3, a combinatorial joining at the DNA level of a V segment with various J and D segments increases antibody diversity and may generate new antibody specificity as discussed by Early et al. (24).

ACKNOWLEDGENENTS

We thank S. Wylder and D. Ram for excellent technical help, Dr. R. Perry for the gift of plasmid $py(11)^7$, Dr. D.L. Bentley for phage M13mp2/Pst and Dr. M. Gorecky for advice in the cloning. D.G. is grateful to Dr. G. Brownley and S. Fields from MRC, Cambridge, for a short visit to their laboratory in order to learn cloning and sequencing in M13. This work was supported by the NCRD, Israel, and the DKFZ, Heidelberg, Germany.

REFERENCES

² Seidman, J.G., Leder, A., Edgell, M.H., Polsky, F., Tilghman, S.M., Tiemeier, D.C. and Leder, P. (1978) Proc. Natl. Acad. Sci. USA 75, 3881- 3885.

