Molecular cloning and nucleotide sequencing of human immunoglobulin ε chain cDNA

Masaharu Seno, Tsutomu Kurokawa, Yoshitaka Ono, Haruo Onda, Reiko Sasada, Koichi Igarashi, Masakazu Kikuchi, Yukio Sugino, Yasuyoshi Nishida* and Tasuku Honjo*

Biotechnology Laboratories, Central Research Division, Takeda Chemical Industries Ltd., Osaka 532, and *Department of Genetics, Osaka University Medical School, Osaka 530, Japan

Received 22 November 1982; Revised and Accepted 7 January 1983

ABSTRACT

DNA complementary to mRNA of human immunoglobulin E heavy chain (ε chain) isolated and purified from U266 cells has been synthesized and inserted into the PstI site of pBR322 by G-C tailing. This recombinant plasmid was used to transform E.coli χ1776 to screen 1445 tetracycline resistant colonies. Nine clones (pGET1 - 9) containing cDNA coding for the human ε chain were recognized by colony hybridization and Southern blotting analysis with a nick-translated human IgE genome fragment. The nucleotide sequence of the longest cDNA contained in pGET2 was determined. The results indicate that the sequence of 1657 nucleotides codes for 494 amino acids covering a part of the variable region and all of the constant region of the human ε chain. Most of the amino acid sequence deduced from the nucleotide sequence is in substantial agreement with that reported. Furthermore a termination codon after the -COOH terminal amino acid marks the beginning of a 3' untranslated region of 125 nucleotides with a poly A tail. Taking this into account, the structure of the human ε chain mRNA, except a part of the 5' end, is conserved fairly well in the cDNA insert in pGET2.

INTRODUCTION

Considerable attention has been paied to immunoglobulin E (IgE) since antibodies of this minor class of immunoglobulin were found to carry the biological, immunochemical, and physical characteristics of reaginic antibodies (1-7).

The IgE molecule basically consists of two light chains and two heavy chains, joined to each other by disulfide bonds as is found in other immunoglobulins. The molecular weight of the light chain is 22,500 and that of the heavy chain 72,500, so IgE has a molecular weight of about 190,000 (6,7). The amino acid sequence of IgE (ND) has been determined by Dorrington and Bennich (7).

In spite of the success in cloning human, mouse, and rat IgE genomic DNA (8-13), cDNA cloning of human IgE has been difficult because of the scarcity of mRNA. It is widely known that myeloma U266 cells subcloned from 266B1 (14,15) produce human IgE (ND), and these cells are considered to be a good source for molecular cloning of IgE mRNA. We have isolated and purified mRNA coding for the human ε chain from these U266 cells (16). The scarcity of human IgE coupled with its potential value in medicine makes the possibility of producing the whole or a part

of the polypeptide by recombinant DNA techniques attractive. Artificially modified human IgE polypeptide produced in this way is expected to compete with natural human IgE in binding to specific membrane receptors on basophils and mast cells so that allergic reactions of the immediate type will be modified.

In this paper we report the synthesis, cloning, and nucleotide sequencing of the human ε chain cDNA. In the course of the preparation of this manuscript, Kenten et al. also reported the cDNA cloning from myeloma 266Bl (17).

MATERIALS AND METHODS

Enzymes

Avian myeloblastosis virus (AMV) reverse transcriptase was purchased from Life Science Inc.; calf thymus deoxynucleotidyl terminal transferase, Sl nuclease, and T4 ligase from Bethesda Research Laboratories; bovine pancreas DNase I from Worthington; E.coli DNA polymerase I from Boehringer; and restriction endonucleases from Takara Shuzo. Restriction endonucleases were used under the conditions recommended by the supplier.

Preparation of labelled probe

Double-stranded DNA was labelled with [^{32}P] by the nick-translation method described by P.Rigby et al. (18) and used as the hybridization probe. The IgE genomic DNA fragment partially coding for CH2 and CH3 (19) was used as a probe.

Purification of mRNA

mRNA for human ε chain was isolated and purified from U266 cells as described by Onda et al.(16).

Isolation of cDNA clones

Double-stranded cDNA was synthesized from mRNA by oligo(dT)-primed reverse transcription (20), Sl treated, and tailed with dC residues. The tailed cDNA of more than 700 nucleotides was selected by agarose gel electrophoresis and electroeluted (21). The isolated cDNA was annealed to dG tailed Pstl digested pBR322 and used to transform E.coliχ1776 (22). Tetracycline resistant transformants were picked onto nitrocellulose filters (Schleicher & Shuell) and colony hybridization was done by the method of Grunstein and Hogness (23). Southern blotting analysis (24) was applied to the positive clones to ensure the hybridizability of the inserts to the probe DNA.

DNA sequence determination

Plasmid DNA was prepared from E.coliχ1776 by the alkaline-SDS procedure (25) followed by equilibrium density gradient centrifugation, digested with restriction endonucleases (Sau 3A, Hpa II), and fractionated by gel electrophoresis. The DNA fragments in the gels were individually eluted (21) and cloned into M13mp7 RF DNA

(26,27). Using this recombinant phage, the nucleotide sequence was determined by the 'dideoxy' chain termination DNA method of Sanger et al.(28,29). The M13 kit was supplied by Bethesda Research Labratories. The resulting fragments were resolved on 0.4 mm thick 6 or 8 % acrylamide gels run for 40 or 90 cm.

The nucleotide sequence determined as described above was edited and analyzed with programs written in FORTRAN IV on an IBM 3031 computer.

RESULTS

cDNA synthesis and transformation of E.coli χ1776

We have isolated and partially purified the mRNA for the human ε chain from IgE producing human myeloma cell line U266 (14), and synthesized cDNA for human ε chain with this mRNA. We obtained 80 ng of double-stranded cDNA starting with 10μg of mRNA. The product synthesized here is longer than 700 nucleotides after size selection by 1.3% agarose gel electrophoresis .

This cDNA (40ng) was annealed to dG tailed PstI digested pBR322 (150ng) in order to transform E.coli χ1776. The number of tetracycline resistant colonies was 1445, which leads to the transformation frequency of $3.6X10^4$ colonies/μg cDNA. Ninty five percent of the transformants are ampicillin sensitive, and almost all of them are considered to have recombinant plasmids.

Isolation of cDNA

The ε chain specific clones were identified by colony hybridization. A nick-translated IgE genome fragment, which was 580 nucleotides of an Ava II digested fragment partially coding for CH2 and CH3 (19), was used as the probe; 9 out of the 1445 transformants hybridized to the probe. The recombinant plasmids were named pGET1 - 9. These plasmids were digested by Pst I and fractionated on 1.0% agarose gel to determine the size of each cDNA insert as shown in Fig.1(a). As the result of size selection at the time of cDNA synthesis, all of the inserts were longer than 900 nucleotides. Among them, pGET2 was shown to have the longest insert of 1650 nucleotides. On southern blotting analysis, all the cDNA inserts of pGET1-9 strongly hybridized to the same probe that was used in colony hybridization (Fig.1(b)), and it was asertained that these inserts were complementary to the IgE genome fragment. Restriction maps of the cDNA inserts in pGET2 and pGET4 are shown in Fig.2. The alignment of the maps was determined by double-digestion with Hinc II, Hinf I, Hpa II, and Sau 3A. It is evident that both of the cDNA inserts cover almost the same region of the ε chain mRNA derived from U266 cells.

Nucleotide sequence determination

As it was clear that pGET2 covered the largest part of the human ε chain mRNA among all of the 9 clones, the nucleotide sequence of the cDNA insert in

pGET2 was determined. The sequencing strategy using the 'dideoxy' chain termination DNA method is shown in Fig.2 and the nucleotide sequence determined is presented in Fig.3. This sequence consists of 1614 nucleotides derived from the mRNA of the human ε chain with G-C tails of 14-15 residues and a poly A tail of 16 residues specific to the 3' end of mRNA. A termination codon at nucleotide no. 1499-1501 after the -COOH terminal amino acid marks the beginning of the 3' untranslated region of 124 nucleotides, which contains one further in-phase termination codon at nucleotide no. 1587-1589. It also includes the AATAAA sequence thought to be the signal for polyadenylation (30), 14 nucleotides upstream from the beginning of the poly A tail. The 5' untranslated region plus about 200 nucleotides downstream from the initiation codon of ATG triplet were not contained in the cDNA insert of pGET2, because the first 62 amino acids reported by Dorrington and Bennich (7) are not coded in our sequence. That is, about a half of the variable region is missing. The number of amino acids coded in our sequence is 494 and the sequence is compared with that reported (7) in Fig.3. Each amino acid is represented by one letter code proposed by IUPAC-IUB (31). Our sequence

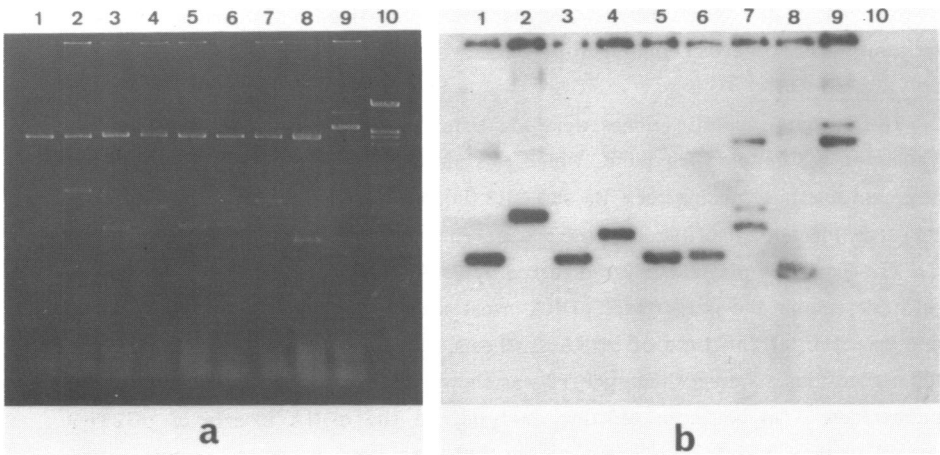

Figure 1. Agarose gel electrophoresis and Southern blotting analysis
Plasmids of 9 clones positive in colony hybridization were digested by PstI and fractionated by 1% agarose gel electrophoresis. DNA fragments in this gel were transferred onto nitrocellulose filter and fixed by the procedure of Southern (24). Pretreatment of the filter was done with 3X SSC for 30 min at 65° C and then with 5X Denhardt's solution(35)/ 3X SSC for 1 hr at 65° C and continued for 30 min at 65° C in the hybridization solution (5X Denhardt's solution/ 50mM Tris-HCl (pH 7.5)/1M NaCl/ 10mM EDTA/ 0.1% SDS/ 10mg/ml salmon sperm DNA). The DNA on the filter was hybridized to the probe in the hybridization solution for 16 hrs at 65° C. (a) 1.0% agarose gel electrophoresis, (b) Southern blotting analysis :Lane; 1 pGET1, 2 pGET2, 3 pGET3, 4 pGET4, 5 pGET5, 6 pGET6, 7 pGET7, 8 pGET8, 9 pGET9, 10 Marker (Eco RI, Hind III digested λ phage)

coded in the cDNA insert of pGET2 is partially different from that reported (7,9), furthermore, the unordered sequence is predicted. In this cDNA insert the number of nucleotides coding for the human ε chain is 1487 including the termination codon of TGA, and the rest of the 124 nucleotides is 3' untranslated region.

DISCUSSION

The nucleotide sequence of our cDNA synthesized against the human ε chain mRNA codes for the whole constant region (CHl-CH4) and a half of the variable region (VH) of the human ε chain after 15 residues of the G-tail. None of the plasmids so far examined had the sequence for the 5' end untranslated region and the first half of the variable region. Our amino acid sequence deduced from the nucleotides is partially different from that reported by Dorrington and Bennich (7). The different parts and the newly assigned amino acids are marked by asterisks in Fig.3 . Though the sequence of A-S-T-E-S-Q was reported to repeat in the residues 229-234 and 238-243 as underlined in Fig.3, there exists no such repetition in our sequence. Some of cysteines and aspargines in the polypeptide are known to be the important positions for disulfide bond formation and glycosylation, respectively (7). In our sequencing, the locations of these amino acids are completely coincident with those reported (7).

Recently the nucleotide sequence of the genomic DNA for the ε chain from

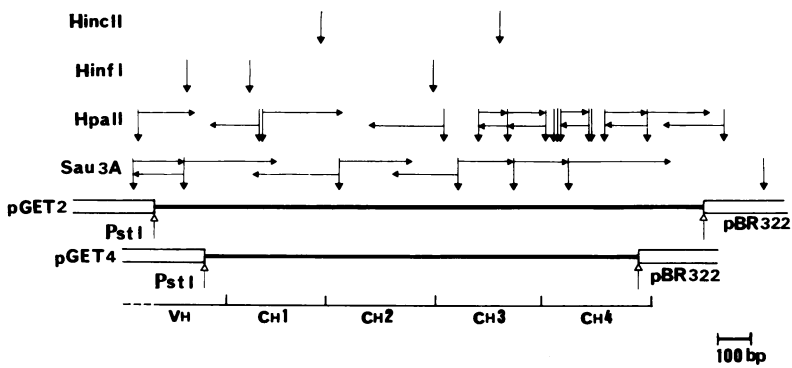

Figure 2. Restriction map and sequencing strategy
The cDNA inserts of pGET2 and pGET4 were digested by Hinc II, Hinf I, Hpa II and Sau 3A and tentative restriction maps were constructed. The longest cDNA inserted into pGET2 was used to determine the nucleotide sequence by the procedure of Sanger et al. (28,29). Each horizontal arrow and its length indicate the direction and extent of the sequence determined from the appropriate fragment by Sau 3A and Hpa II. And each vertical arrow indicates the restriction site of each restriction endonuclease. The variable region (VH) and the constant region (CHl-CH4) are also indicated from the comparison between our nucleotide sequence and the one reported (9).

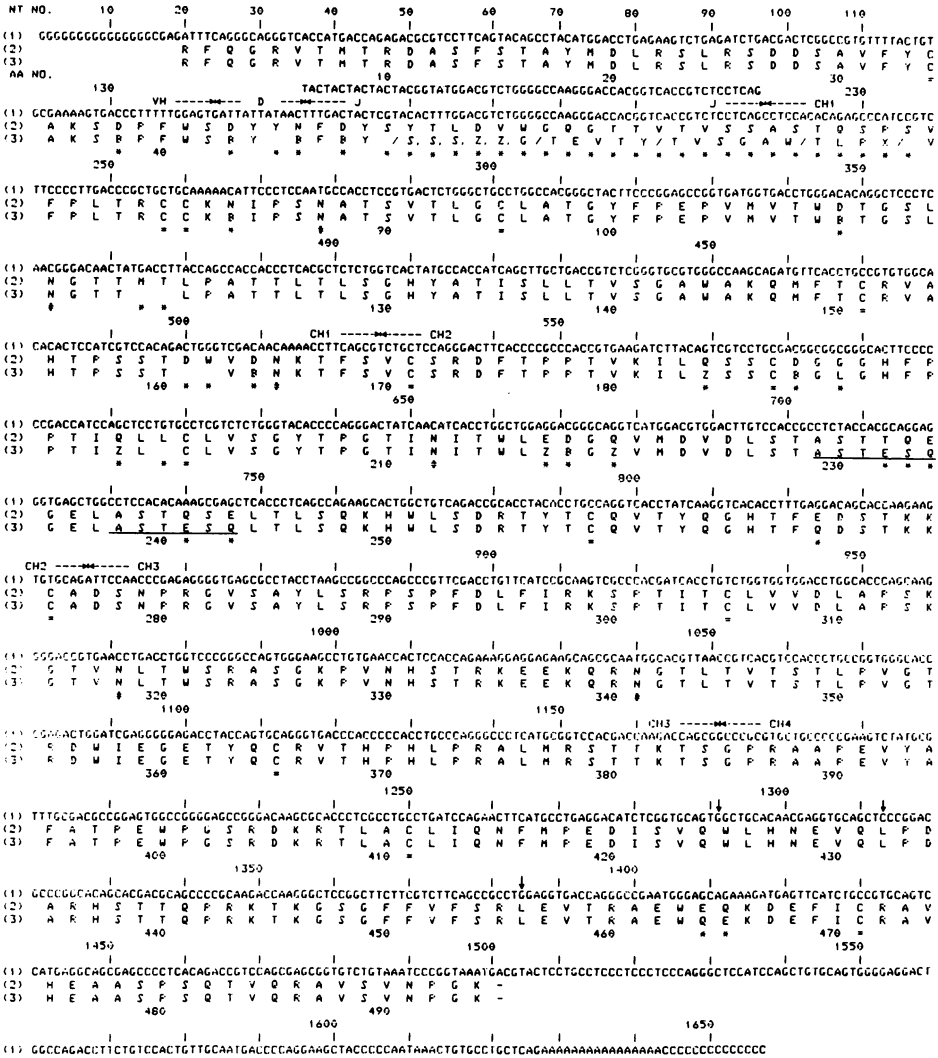

Figure 3. Complete nucleotide sequence of the pGET2 cDNA insert

The amino acid sequence (row(2)) deduced from the nucleotide sequence (row(1)) is compared with the amino acid sequence of Dorrington and Bennich (row(3))(7). Asterisks mark the amino acids in row(2) that are different from those reported by them and newly ordered. There exists repetition of the amino acid sequence in the reported one (row(3); single-underlined). One-letter codes follow the rule by IUPAC-IUB (31). The nucleotide sequence of J6 by J.Ravetch et al. (32) is proportionally indicated. The location of VH, D, J, and CH1-CH4 are also estimated by the comparison of our nucleotide sequence with the one reported (9,32,33,34). Double-underlines and sharps mark disulfide-bound-cystein and oligosaccharide-located-arginine in intact IgE protein respectively. The nucleotides that differ from the reported sequence by E.Max et al. (9) are indicated by vertical arrows.

the human placenta was reported by Max et al. (9). The functional genomic DNA including introns codes for whole constant region (CH1-CH4). Comparison of our nucleotide sequence with theirs shows complete homology except for 3 nucleotides (indicated by vertical arrows in Fig.3). These nucleotides contained in our CH4 coding sequence agree with the ones in the sequence of the duplicated ε pseudogene for CH4 (9). It is interesting to note that one of these 3 nucleotides, which makes the codon of TGG corresponding to W for 424th residue (Fig.3), supports the genetic polymorphism proposed by Max et al. (9). It is not the TTG in their functional sequence that is retained in the gene expressed by patient N.D. Our views on the human ε chain cDNA do not contradict the recent report by Flanagan and Rabbitts (8).

From the comparison of our nucleotide sequence with the published sequence data on the human heavy chain J-region genes by Ravetch et al. (32), the ND protein appears to utilize the J6 sequence (Y-Y-Y-Y-Y-G-M-D-V-W-G-Q-G-T-T-V-T-V-S-S) for residues 47-66, followed by the alanine residue coded by GCC, which is located at the junction of the J6 and CH1 (Fig.3). Five of the first 7 amino acids in this region are different from the ones coded in the J6 sequence. These differences result from 8 nucleotide discrepancies between the two sequences shown in Fig.3.

The -COOH terminal of the VH region is supposed to associate with the D (diversity) segment (33). It is difficult to determine the sequence derived from the D-segment gene from our sequence only. Considering other reports of myeloma proteins for the D-segment (32-34), the Y-Y sequence appears one of the characteristic features for several proteins and the D-segment may lie among the nucleotides at around 144-155 coding D-Y-Y-N for residues 43-46 in the case of the ND protein.

Recently Kenten et al. reported a cDNA clone from myeloma 266Bl (17). Their views on the VH, D and J regions are almost identical to ours, however, further discussion is not to be made because they have not determined all the nucleotide sequence for the CH regions in their cDNA.

The primer extension experiment for the cDNA cloning of the 5' end of mRNA and studies on the expression of the human ε chain gene in E.coli are now in progress.

REFERENCES

1. Ishizaka,K. and Ishizaka,T. (1967) J.Immunol. 99, 1187-1198
2. Stanworth,D.R., Humphrey,J.H., Bennich,H.H. and Johanson,S.G.O. (1967) Lancet II, 330-332
3. Bennich,H.H., Ishizaka,K., Ishizaka,T. and Johanson,S.G.O. (1969) J.Immunol. 102, 826-831

4. Bennich,H.H. and Johanson,S.G.O. (1971) in:Adv.Immunol., vol.13,eds. F.I.Dixon and H.G.Kunkel (Academic Press New York) 1-55
5. Banziger,J. and Kornfeld,S. (1974) J.Biol.Chem. 249, 1889-1903
6. Bennich,H.H. and Von Bahr Lindstrom,H. (1974) in:Prog.Immunol. II, vol 1,eds. L.Brent and J.Holborrow (North-Holland Publishing Company) 49-58
7. Dorrington,K.J. and Bennich,H.H. (1978) Immunol.Rev. 41, 3-25
8. Flanagan,J.G. and Rabbitts,T.H. (1982) EMBO J. 1, 655-660
9. Max,E.E., Battey,J., Ney,R., Kirsch,I.R. and Leder,P. (1982) Cell 29, 691-699
10. Ueda,S., Nakai,S., Nishida,Y., Hisajima,H. and Honjo,T. : EMBO J. (in press)
11. Nishida,Y., Kataoka,T., Ishida,N., Nakai,S., Kishimoto,T., Bottcher,I. and Honjo,T. (1981) Proc.Natl.Acad.Sci.USA 78, 1581-1585
12. Hellman,L., Petterson,U. and Bennich,H.H. (1982) Proc.Natl.Acad.Sci.USA 79, 1264-1268
13. Hellman,L., Petterson,U., Engstrom,A., Karlsson,T. and Bennich,H. (1982) Nucleic Acids Res. 10,6041-6049
14. Han,T., Dadey,B., Minowada,J. and Golde,D.W. (1979) Immunology 38,63-68
15. Nilsson,K. (1971) Int.J.Canser 7,380-396
16. Onda,H. et al. : in preparation
17. Kenten,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 79, 6661-6665
18. Rigby,P.W.J., Dickmann,M., Rhodes,C. and Berg,P. (1977) J.Mol.Biol. 113, 237-251
19. Nishida,Y., Miki,T., Hisajima,H., and Honjo,T. (1982) Proc.Natl.Acad.Sci. USA 79, 3833-3837
20. Taniguchi,T., Palmieri,M. and Weissmann,C. (1978) Nature 274 , 223-228
21. Yang,R., Lis,J. and Wu,R. (1979) Methods in Enzymology 68, 176-182
22. Enea,V., Vovis,G.F. and Zinder,N.D. (1975) J.Mol.Biol. 96, 495-509
23. Grunstein,M. and Hogness,D.S. (1975) Proc.Natl.Acad. Sci.USA 72, 3961-3965
24. Southern,E.M. (1975) J.Mol.Biol. 98, 503-517
25. Ish-Horowicz,D.B. and Burke,J.F. (1981) Nucleic Acids Res. 9, 2989-2998
26. Schreier,P.H. and Cortese,R. (1979) J.Mol.Biol. 129, 169-172
27. Sanger,F., Coulson,A.R., Barrel,B.G., Smith,A.J.H. and Roe,B.A. (1980) J.Mol.Biol. 143, 161-178
28. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc.Natl.Acad.Sci.USA 74, 5463-5467
29. Sanger,F. and Coulson,A.R. (1978) FEBS Lett. 87, 107-110
30. Proudfoot,N.T. and Brownlee,G.G. (1976) Nature 263, 211-214
31. IUPAC-IUB (1969) Biochem.J. 113, 1-4
32. Ravetch,J.V., Siebenlist,U., Korsmeyer,S., Waltmann,T. and Leder,P. (1981) Cell 27, 583-591
33. Schilling,J., Clevinger,B., Davie,J.M. and Hood,L. (1980) Nature 283, 35-40
34. Rao,D.N., Rudikoff,S., Kurtzsch,H. and Potter,M. (1979) Proc.Natl.Acad.Sci. USA 76, 2890-2894
35. Denhardt,D.T. (1966) Biochem.Biophys.Res.Commun. 23, 641-646