Volume 10 Number 19 1982 Nucleic Acids Research

Structure and evolution of the heavy chain from rat immunoglobulin E

Lars Hellman[1], Ulf Pettersson[1], Åke Engström[2], Torbjörn Karlsson[2] and Hans Bennich[2]

[1]Department of Medical Genetics and [2]Department of Medical and Physiological Chemistry, The Biomedical Center, S-75123 Uppsala, Sweden

Received 24 August 1982; Accepted 21 September 1982

ABSTRACT

The nucleotide sequence of the rat ε-chain mRNA has been determined by sequencing cloned cDNA copies of the mRNA. The established sequence covers the coding region, the 3'-non coding region and most of the 5' non-coding region. A comparison with the nucleotide sequence of the human ε-chain constant region reveals that C_3 and C_4 are the most highly conserved domains. The rat ε-chain contains a C-terminal decapeptide which is not present in the human counterpart.

INTRODUCTION

A vast amount of knowledge has accumulated during the past decade concerning the structure of immunoglobulin molecules. The primary sequences of numerous heavy and light chains have been determined either by amino acid or nucleotide sequencing and it is now well established that heavy chain genes are assembled by joining V-, D-, J- and C-segments (for a review see 1). Immunoglobulin E (IgE) is a minor but important component of the immune system, being of great medical interest, since IgE is a mediator of allergic reactions of the immediate type (2). Studies of the human IgE system have been hampered by the lack of IgE-secreting myelomas and less than twenty IgE myelomas have so far been identified, one of which, IgE(ND), has been studied in detail (3). An animal model to aid the study of the biology of the IgE system is provided by the IgE-secreting immunocytomas, which occur spontaneously in LOU/c/Wsl rats (4). We have prepared cDNA copies of the ε-chain mRNA from the rat immunocytomas IR2 and IR162, and one cDNA clone was recently described (5). In the present communication we have constructed and sequenced cDNA copies which cover almost the entire ε-chain mRNA from the IR2 immunocytoma.

MATERIALS AND METHODS

Cells and mRNA

Tumors were produced and mRNA was extracted according to already published

0305-1048/82/1019-6041$2.00/0

procedures (5).

Preparation of cDNA clones

The clone IR2:11 was prepared as described (5). The clone IR2:4 is a cDNA clone prepared by using fragments of the cDNA present in the clone IR2:2:11 as primers. The primer fragments were generated by cleavage with endonucleases AvaI, PvuII and HphI. The second strand of clone IR2:4 was prepared by attaching (dC) tails of the first strand followed by priming the second strand with oligo(dG) as described by Land et al. (8). The cDNAs were inserted into the PstI cleavage site of the pBR322 plasmid after dG/dC tailing.

DNA sequencing

The method of Maxam and Gilbert was followed (13).

RESULTS AND DISCUSSION

Construction of cDNA clones covering the rat ε-chain mRNA:

In order to establish the complete structure of the rat ε-chain, a cDNA library was constructed as described (5). One clone (pIR2:2:11) which contained an approximately 1.1 kilobases (kb) long insert was partially characterized, and sequence analysis of the clone revealed that it contains sequences from the 3' part of the ε-chain. Analysis of RNA from the immunocytomas showed, however, that the complete ε-chain mRNA is approximately 2 kb long (5) and none of the clones in the original library contained cDNAs approaching this size. We therefore constructed an additional cDNA library using fragments from the original clone pIR2:2:11 as primers. The primer fragments were generated by cleavage with a mixture of endonucleases AvaI and PvuII and the purified fragments were recleaved with endonuclease HphI to increase the molar concentration of fragments. Using this approach several clones which include the 5' part of the ε-chain mRNA were obtained. One of these, designated pIR2:4, contains a cDNA insert which is 1,406 nucleotides long. This clone together with the original clone pIR2:2:11 were used to sequence the coding region together with the entire 3' and most of the 5' non-coding regions of the ε-chain mRNA.

The structure of the rat ε-chain mRNA and its protein product

The strategy used to sequence the two cDNA clones is shown in Fig. 1 and the established 1,866 nucleotides long sequence is shown in Figs. 2 and 3. Sequences corresponding to the signal peptide as well as those corresponding to the V-, J-, and C-segments were identified as shown in Fig. 1. The D-segment could not be mapped precisely since no germline V-region, D- or J-segment sequences of the rat have been published. The approximate junction bet-

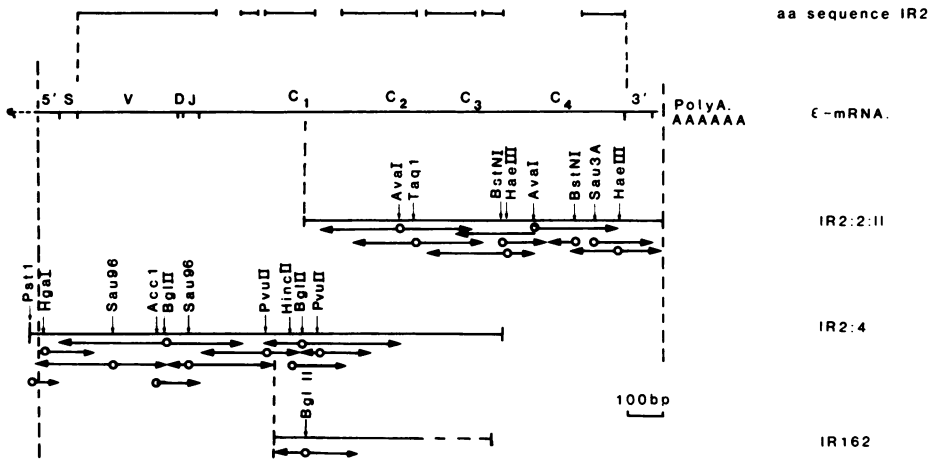

Figure. 1. The structure of the rat ε-chain mRNA and the strategy used to determine its sequence. Two cDNA clones, IR2:2:11 and IR2:4 were used to establish the sequence and their structures are shown in the figure. The cleavage sites which were used for sequencing are shown, and the arrows indicate the extent of sequence information obtained from individual cleavage sites. The brackets at the top of the figure indicate regions of the ε-chain which were determined by chemical amino acid sequencing using automated (Beckman 890 C Sequencer) Edman degradation of CNBr fragments (Engström, Å. et al., unpublished results). Nucleotide sequence information was also collected from a cDNA clone prepared with mRNA from the IR162 rat immunocytoma as template. The arrows indicate the amount of sequence information collected from this cDNA clone. In the mRNA structure, the signal peptide (S), the variable region (V) and the D- and J-segments are indicated as well as the four constant (C) domains C_1-C_4.

ween V- and D-regions was found by a comparison with the published mouse germline V-gene sequences (6). A slight homology with the mouse germline D-segment $D_{SP2.1}$ (7) was found in the expected area between the V- and J-regions (Fig. 2). Approximately two-thirds of the predicted amino acid sequence has been verified by amino acid sequence analysis (Engström et al., unpublished results) as illustrated in Fig. 1, and the results obtained by the two methods were found to be consistent. A corresponding cDNA clone, originating from the IR162 immunocytoma was partially sequenced (Fig. 1) and the results were found to be identical with those obtained for the IR2 ε-chain, suggesting that identical C-genes have been translocated in the two tumors. The junction between the signal peptide and the V-segment was deduced from the chemically determined amino acid sequence of the N-terminal end of the IR2 ε-chain. By amino acid sequencing the same carboxy-terminal sequences was identified as could be predicted from the nucleotide sequence,

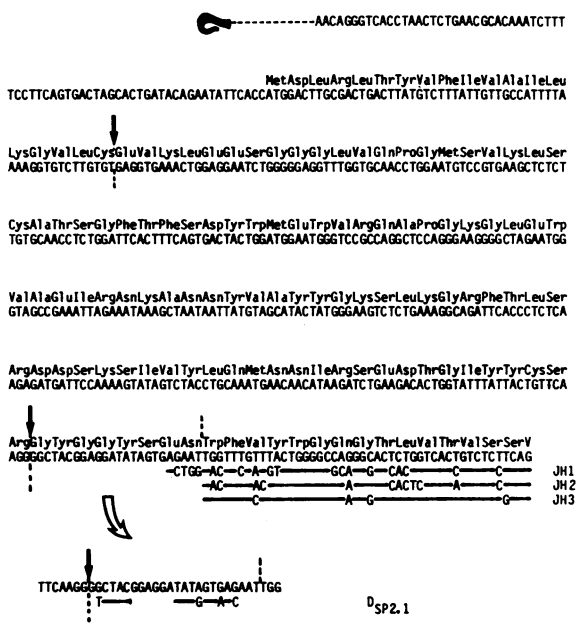

Figure.2: The nucleotide sequence of the 5' non-coding region of the rat ε-chain mRNA and regions encoding the signal (S) peptide and the variable (V) segment. The coding part of the sequence is translated into the predicted amino acid sequence. The junction between the S- and V-regions was deduced from the amino acid sequence as determined by chemical sequencing of the amino terminal end of the mature rat ε-chain. The junction between V- and D-regions was found by comparing the sequence with published mouse germline V-region sequences (6). The junctions between S and V as well as V and D are indicated by arrows. As a comparison, the sequence of three mouse J-segments (9), JH1, JH2 and JH3 and a germline mouse D segment $D_{SP2.1}$ is included. Underlined nucleotides are homologous.

thus proving that no processing occurs at this end of the ε-chain. The signal peptide consisting of 19 residues ends with a cysteine residue, and like other signal peptides it is rich in hydrophobic amino acids. The length of the processed ε-chain is 552 amino acids as deduced from the cDNA sequence, and the length of the 3' non-coding region of the mRNA is 86 nucleotides. In this part of the mRNA the poly(A) addition signal AAUAAA is found 10-15 nucleotides upstream from the poly(A) addition site. The structure of the entire 5' non-coding region of the ε-chain mRNA cannot be deduced from cDNA clones since 5' sequences usually are lost when cDNA clones are prepared. However, as a primer was used to prepare the second strand of the cDNA (8) we expect that only a few nucleotides are missing from the 5' end, which makes the entire length of the rat ε-chain mRNA about 1,900 nucleotides

excluding the poly(A) tail.

The mouse heavy chain J-segments have recently been sequenced (9) and a comparison with the IR2 -sequence is shown in Fig. 2. The J-segment present in the IR2 sequence resembles the mouse JH3 segment; four differences are apparent, possibly generated by somatic mutations.

Analysis of the codon usage reveals nothing striking as compared to other eucaryotic genes (10) and codons ending with C (40%) or G (26%) are clearly overrepresented.

A comparison between rat and human ε-chains

Max et al. (11) have recently established the nucleotide sequence for the gene segments encoding the constant (C) domains of the human ε-chain. The predicted amino acid sequence for human and rat ε-chains were aligned for comparison as shown in Fig. 4 and Table 1 lists the degree of homology obtained in terms of amino acids and nucleotides, respectively. The highest degree of sequence homology is found within the CH3(50%) and CH4(54%) segments, both at the DNA and the amino acid levels. Particularly striking is the very high degree of homology present within the 3' non-coding region (Fig. 3). The relatively high degree of homology as displayed by the amino acid sequence, particularly in the CH3 and CH4 domains, is interesting in view of the fact that human and rat IgE neither cross-react immunologically using conventionally produced antisera, nor bind to mast cells or basophilic leucocytes of the opposite species (4).

The positions for cysteinyl and tryptophanyl residues are highly conserved as expected, and it is noteworthy that those cysteinyl residues involved in the formation of the two inter-epsilon disulfide-bridges, both are located within the exon coding for the CH2 domain, which in ε- and μ-chains is analogous to the hinge region of α-, δ- and γ-chains. The extra intra-chain disulfide loop present in the CH1 domain of the human ε-chain and the rabbit γ-chain is apparently absent in the rat ε-chain.

The predicted amino acid sequence for the rat ε-chain (IR2) contains as expected two invariant tryptophanyl residues in the CH4 segment, which is consistent with the amino acid sequence data for CH4 domain of the human ε-chain (ND) (3). In contrast, the nucleotide sequence of human ε-chain shows that one of the tryptophane residues is replaced by a leucine. However, a tryptophane codon is present in this position in a duplicated ε-pseudogene, thus indicating polymorphism in the human ε-chain locus (11).

The human ε-chain (ND) contains six glycosylation sites each having a glucosamine residue attached to the asparaginyl residue in the configuration

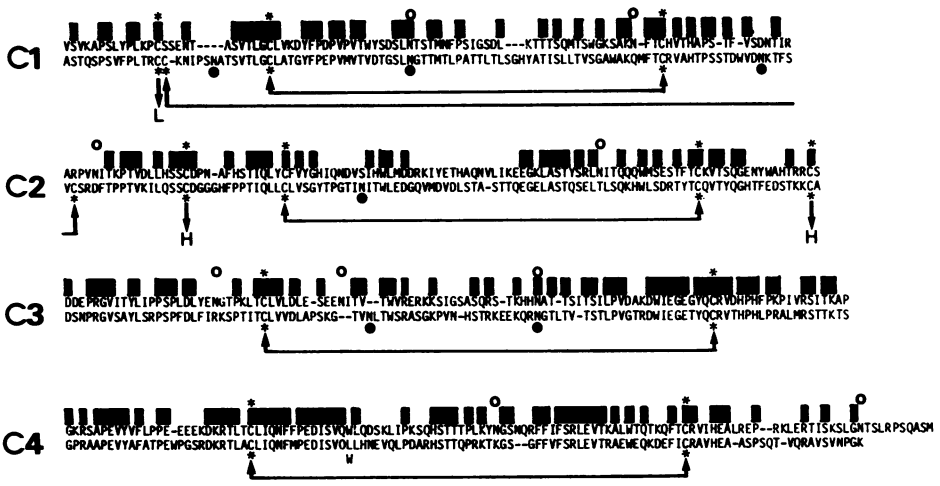

Figure 4: Comparison of the primary structure of human and rat C-region domains 1 to 4 as predicted from the nucleotide sequences of the rat (top line) and human (11) (bottom line) ε-chains. The sequences were aligned to maximize homology. Black boxes above the sequences indicate homologous amino acids. Known glycosylation sites in the human ε-chain (3) are indicated by filled circles, and putative glycosylation site in the rat ε-chain are indicated by open circles. Known disulfide bridges in the human sequence are indicated; cysteine residues are marked with a star (*), and the location of a polymorphic site (Leu/Trp) in the human CH4 domain is also indicated.

-ASN-X-THR. The rat ε-chain (IR2) contains 9 putative glycosylation sites (-ASN-X-SER- and ASN-X-THR) as shown in Fig. 4. Three of these sites, one in CH1 and two in CH2, are homologous to the corresponding sites identified in the human ε-chain (3). Concerning the remaining putative locations of carbohydrate side chains, it is noteworthy that one site is located in the carboxy-terminal deca-peptide, which is absent in the human ε-chain (ND) (Fig. 4). The finding of an additional C-terminal segment comprising 10 amino acid residues, which appears to be unique for the rat ε-chain, is intriguing in terms of its possible contribution to the biological functions of IgE and its evolutionary relationship to the extended C-terminal segments present in the α-, δ- and μ-chains. In contrast to the highly conserved structure of

Figure 3: The nucleotide sequence of the C-region of the rat ε-chain (IR2) and the 3' non-coding region of the ε-chain mRNA. For comparison the corresponding nucleotide sequence of the human ε-chain (11) is also shown. Nucleotides, which are identical in the two sequences, are underlined. The junctions between the 4 domains, as predicted from the human ε-sequence, are indicated by vertical dashed lines. The termination codons are indicated by stipled boxes.

TABLE 1. Sequence homology between the constant regions of rat and human ε-chains.

Region	percent homology	
	amino acids	nucleotides
CH1	43	60
CH2	38	60
CH3	50	63
CH4[x]	54	64
3' non-coding[xx]	-	71

[x]excluding the tail peptide of rat ε-chain.
[xx]from the termination codon in the rat ε-chain.

the α- and μ- nonadeca-peptides (12), there is no direct homology between these and the deca-peptide of the rat ε-chain except for the methionyl residue at position 10. Similarly, the deca-peptide shows no homology with the C-terminal octa-peptide of the δ-chain except that both peptides terminate with a methionine.

The biological significance of an extended CH4 domain is obscure. Though it seems unlikely that the deca-peptide contributes to the unique features of IgE to bind to mast cells and basophils, it might well be of importance for some yet unknown biological functions of rat IgE.

ACKNOWLEDGEMENTS

We thank Christina Sjöholm and Marianne Gustafson for secretarial aid and Agneta Lissmats for technical assistance. We are also grateful to Doctors Philip Leder and Edward Max for making sequence information available prior to publication. This investigation was supported by grants from the Swedish Medical Research Council and the Swedish National Board for Technical Development.

REFERENCES

1. Honjo, T., Nakai, S., Nishida, Y., Kataoka, T., Yamawaki-Kataoka, Y., Takahashi, N., Obata, M., Shimizu, A., Yaoita, Y., Nikaido, T. and Ishida, N. (1981) Immmunol.Rev. 59, 33-67.
2. Ishizaka, T. and Ishizaka, K. (1975) Prog.Allergy 19, 60-121.
3. Bennich, H., Johansson, S.G.O. and von Bahr-Lindström, H. in Modern Concepts and Developments in Hypersensitivity (ed. Bahc, M.K.) 2-36 (Marcel Dekker, New York, 1977).
4. Bazin, H. and Beckers, A. in Nobel Symposium 33: Molecular and

Biological Aspects of the Acute Allergic Reaction (eds Johansson, S.G.O. Strandberg, K. and Uvnäs, B.) 125-152 (Plenum Press, New York & London, 1976).
5. Hellman, L., Pettersson, U. and Bennich, H. (1982) Proc.Natl.Acad.Sci. USA 79, 1264-1268.
6. Cohen, J.B., Effron, K., Rechavi, G., Ben-Neriah, Y., Zakut, R. and Givol, D. (1982) Nucleic Acids Res. 10, 3353-3370.
7. Kurosawa, Y., von Boehmer, H., Haas, W., Sakano, H., Trauneker, A. and Tonegawa, S. (1981) Nature 290, 565-570.
8. Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schütz, G. (1981) Nucleic Acids Res. 9, 2251-2266.
9. Newell, N., Richards, J.E., Tucker, P.W. and Blattner, F.R. (1980) Science 209, 1128-1132.
10. Kawakami, T., Takahashi, N. and Honjo, T. (1980) Nucleic Acids Res. 8, 3933-3945.
11. Max, E.E., Battey, J., Ney, R., Kirsch, I.R. and Leder, P. (1982) Cell, in press.
12. Lind, L.-C. and Putnam, F.W. (1981) Proc.Natl.Acad.Sci. USA 78, 504-508.
13. Maxam, A.M. and Gilbert, W. (1977) Proc.Natl.Acad.Sci. USA 74, 560-564.