The EMBO Journal vol.6 no.6 pp.1677–1683, 1987

Primary structure of the gene for the murine Ia antigen-associated invariant chains (Ii). An alternatively spliced exon encodes a cysteine-rich domain highly homologous to a repetitive sequence of thyroglobulin

Norbert Koch, Wolfgang Lauer[1], Jüri Habicht and Bernhard Dobberstein[1]

Institute of Immunology and Genetics, German Cancer Research Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, and [1]European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heldelberg, FRG

Communicated by B.Dobberstein

The gene for murine Ia-associated invariant (Ii) chains (Ii31 and Ii41) was characterized by sequence analysis. The gene extends over ~9 kb and is organized in nine exons. Exon 1 encodes the 5′ untranslated region and the cytoplasmic segment, exon 2 the membrane spanning segment and adjacent amino acids and exons 3–8 the extracytoplasmic portion of Ii31. Putative promoter sequences were found upstream of the start of the coding sequence. Between exons 6 and 7 an additional, alternatively spliced exon 6b has been identified. This exon is spliced into the mRNA coding for the Ii-related Ii41 protein. Exon 6b encodes a cysteine-rich domain of 64 amino acids. It shows a remarkably high homology to the repetitive elements in thyroglobulin, a precursor for thyroid hormone. Based on this homology, it is suggested that this domain (TgR) in Tg and in Ii41 may play a role either in hormone formation or as a carrier in the transport of molecules (thyroid hormone or processed antigen respectively) between intracellular compartments.

Key words: invariant chain gene/class II histocompatibility antigens/alternative-splicing/thyroglobulin/antigen processing

Introduction

Invariant chain (Ii) is a protein which is associated intracellularly with the murine and human class II histocompatibility (MHC) antigens, Ia and HLA-D respectively (Jones *et al.*, 1978; Charron *et al.*, 1983; for review see Long, 1985). This assembly occurs shortly after insertion into the membrane of the rough endoplasmic reticulum (ER) (Kvist *et al.*, 1982). After transport to the Golgi complex and the addition of sialic acid residues, Ii dissociates from the class II MHC antigens (Machamer and Cresswell, 1982; Rudd *et al.*, 1985). Class II MHC antigens are transported to the cell surface where they are involved in the presentation of foreign antigens to T-cells (Unanue, 1984). The fate and function of Ii is unclear.

The expression of Ia antigens and Ii is coregulated even though the respective genes are located on different chromosomes (Claesson-Welsh *et al.*, 1984; Koch and Harris, 1984; Yamamoto *et al.*, 1985a; Momburg *et al.*, 1986). Interferon-γ (IFN-γ) and B-cell stimulating factor (BSF-1) induce expression of both Ia antigens and Ii (Collins *et al.*, 1984; Koch *et al.*, 1984; Polla *et al.*, 1986). Proteins homologous to Ii are found in all species which are known to express class II MHC antigens (Sung *et al.*, 1982; Quill and Schwartz, 1983).

Murine Ii chain is a 31-kd type II membrane glycoprotein which spans the membrane once and exposes the 29 amino-terminal residues on the cytoplasmic and the glycosylated carboxy-terminal portion on the extracytoplasmic side of the membrane (Lipp and Dobberstein, 1986). The single cysteine residue on the cytoplasmic side is fatty acylated (Koch and Hämmerling, 1986). Ii shares the membrane orientation and the site of fatty acylation with the transferrin receptor (TR) which internalizes transferrin and recycles it from an intracellular compartment to the plasma membrane (Omary and Trowbridge, 1981; Schneider *et al.*, 1984). Because of its association and co-regulation with class II MHC antigens, and its structural similarity to receptors, it has been proposed that invariant chain may be involved in intracellular transport or recycling of Ia and/or processed antigen complexes (Kvist *et al.*, 1982; Claesson and Peterson, 1983; Cresswell, 1985; Koch and Hämmerling, 1986; Miller and Germain, 1986; Sekaly *et al.*, 1986).

Immunochemical analysis has demonstrated that several forms of murine Ii chain are associated with Ia antigens (Zecher *et al.*, 1984). Recently two mRNA species were described coding for 31-kd and 41-kd Ii-related proteins, Ii31 and Ii41 respectively (Yamamoto, 1985b; Strubin *et al.*, 1986b). Ii31 is expressed in amounts 5–10 times higher than Ii41. After transfection of the Ii gene into rat fibroblasts, both Ii31 and Ii41 were expressed (Yamamoto *et al.*, 1985b). Proteins with similar mol. wts (In33

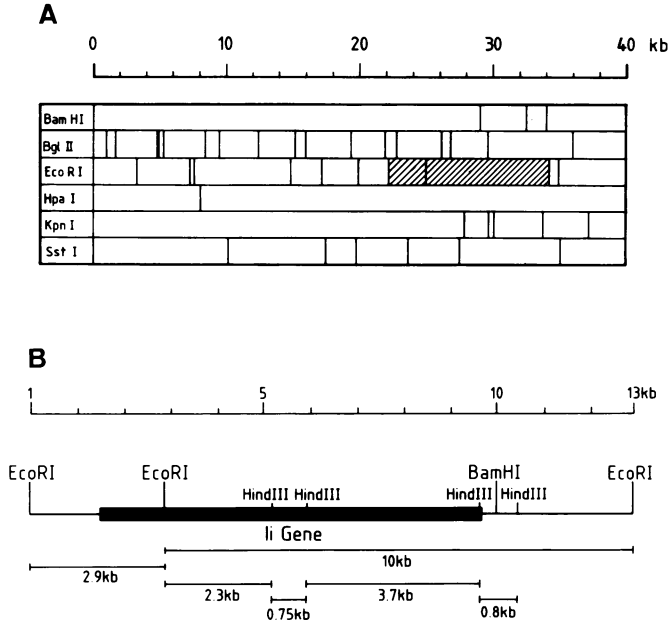

Fig. 1. Physical map of the mouse Ii gene. (**A**) The 40-kb insert in cos 10.7 containing the mouse Ii gene was partially digested with six restriction enzymes and a restriction map constructed. The Ii gene was located by using mouse or human cDNA probes on two adjacent *Eco*RI fragments of 2.9 and 10 kb in length. They are indicated by slanted lines. (**B**) Sequencing strategy for the Ii gene. Fragments as indicated in the figure were subcloned into plasmid vectors, deletions were introduced by DNase I treatment and selected plasmids sequenced. The 5′ end of the gene is to the left and the 3′ to the right.

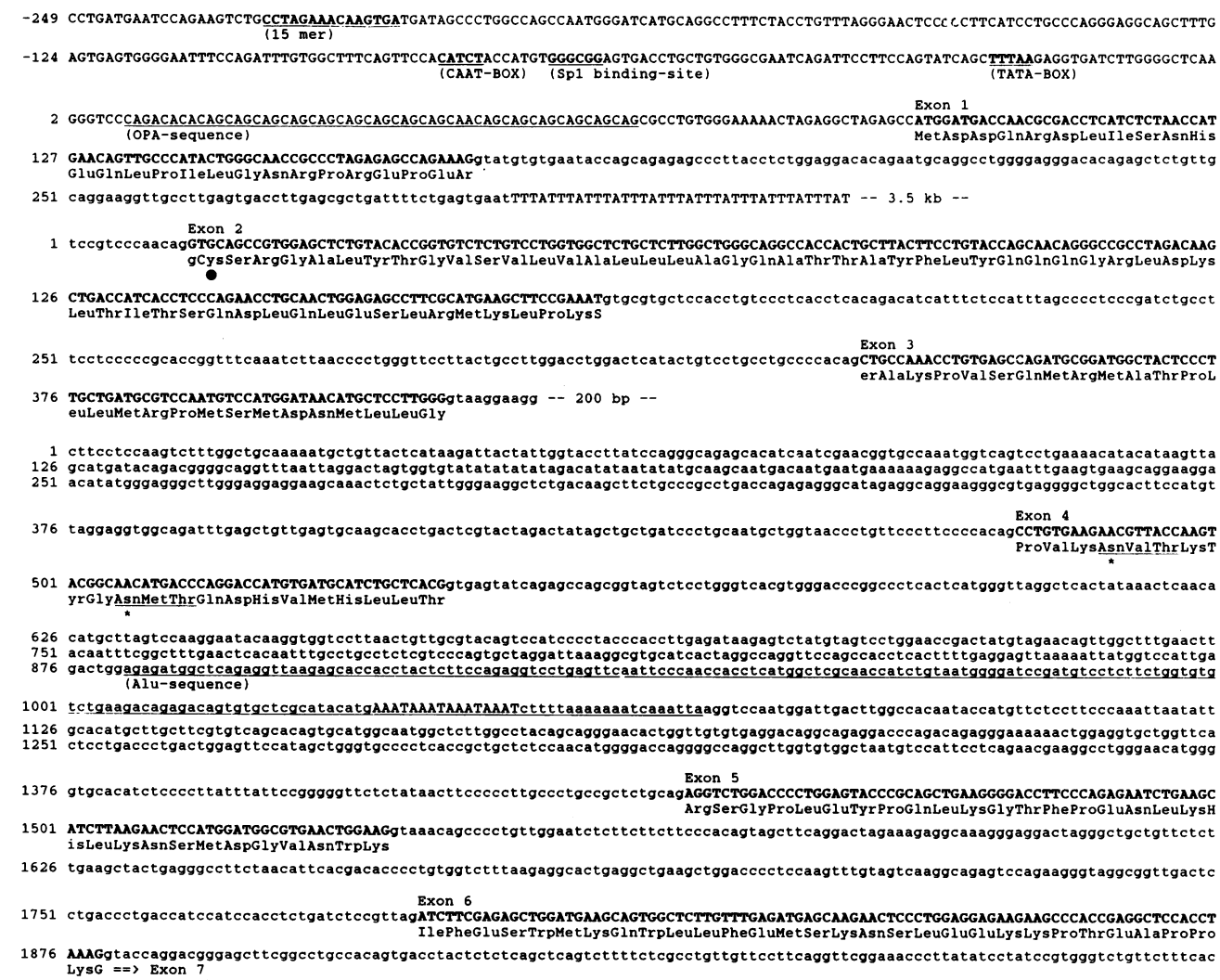

Fig. 2 Sequence of the mouse Ii gene. The nucleotide sequence is shown together with the predicted amino acid sequence of nine exons. Exon 6b is an alternatively spliced exon used to generate the Ii41 protein. Potential regulatory sequences, the CAAT box, one consensus SpI protein binding site (GGGCGG), and the TATA box are indicated at the 5′ end of the gene. A 15-mer segment at −220 is indicated which is highly homologous to a 15-mer segment found in mouse and human class II MHC antigen genes. Downstream of the putative transcription start site, an alternating CAG structure OPA-sequence is found. Alternating TTTA, AAAT and GT nucleotides are indicated by capital letters. An Alu-type repeat element in the fourth intron and the polyadenylation signal AATAAA in the 3′ non-coding region are underlined. In exon 6b the sequence highly homologous to the repetitive element in thyroglobulin (TgR) is also underlined. Potential glycosylation sites in the deduced amino acid sequence are indicated by an * and the cysteine residue to which palmitic acid is bound (exon 2) by a dot.

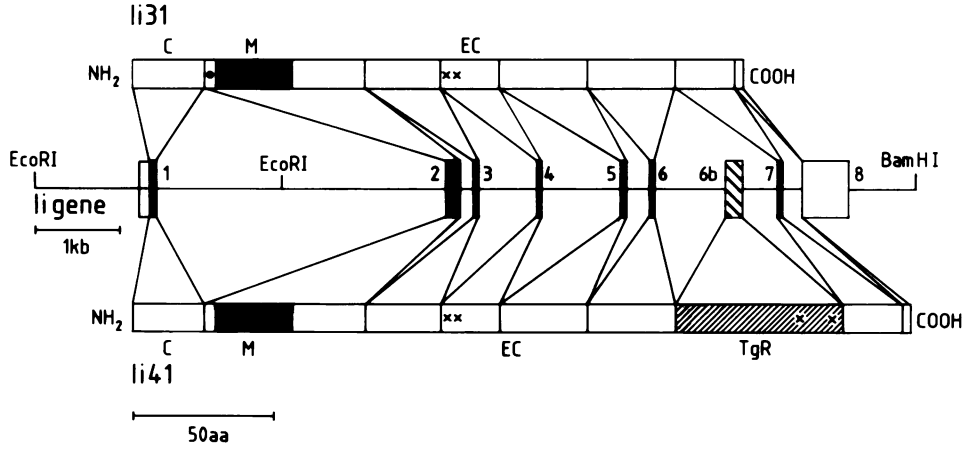

Fig. 3. Outline of the exon−intron structure of the murine Ii gene. Eight exons encode Ii31. Closed boxes indicate exons for coding sequences, open boxes for non-coding sequences. One exon encodes the cytoplasmic and one the membrane-spanning segment. Six exons encode the extracytoplasmic part. An additional exon, 6b, is used in mRNA coding for Ii41. This exon encodes a domain highly homologous to a repetitive domain in thyroglobulin (TgR). The site for fatty acid acylation is indicated by a dot, sites for potential N-glycosylation by ×. C: cytoplasmic, M: membrane-spanning; EC: extracytoplasmic segment.

and In41) were also found in human lymphoblastoid cell lines (Quaranta *et al.*, 1984; Strubin *et al.*, 1986b). Strubin *et al.* concluded from sequence analysis of In41 cDNA that the 41-kd protein results from differential splicing of an invariant chain gene transcript (Strubin *et al.*, 1986b). Other forms of invariant chain (In35 and In43) were found in human cells. These were shown to be translated from an AUG initiation codon upstream of that used for the production of the 33-kd and 41-kd major forms of human invariant chain (Strubin *et al.*, 1986a).

The structure of the human invariant chain gene has recently been determined (Kudo *et al.*, 1985; O'Sullivan *et al.*, 1986). It was shown to be organized in nine exons. One exon codes for the 5′ untranslated region and the cytoplasmic segment, one for the membrane spanning segment and seven for the extracytoplasmic portion.

We describe here the nucleotide sequence of the murine Ii gene. Several consensus sequences with possible regulatory functions are found in the 5′ untranslated region. Comparison with the analogous sequence of the human gene reveals a strong homology in all exons including the exon 6b which by alternative splicing gives rise to the 41-kd Ii41 protein.

Results

Structure of the Ii gene

The isolation and expression of a genomic clone coding for Ii chains has recently been described (Yamamoto *et al.*, 1985b). The 40-kb genomic clone cos 10.7 was shown to contain the complete gene coding for Ii31 and Ii41. We mapped the genomic clone cos 10.7 by restriction analysis. A restriction map of the Ii gene and its flanking regions is shown in Figure 1A. Hybridizations with 5′ and 3′ invariant chain cDNA probes revealed that the entire Ii gene is contained on two *Eco*RI fragments, a 2.9-kb fragment with the 5′ end and a 10-kb *Eco*RI fragment with the 3′ end of the gene.*Eco*RI or *Hin*dIII fragments as shown in Figure 1B were subcloned and all the exons and several of the introns were sequenced. The sequences are shown in Figure 2. Comparison of the Ii31 cDNA sequence with the genomic sequence revealed that the Ii gene is composed of eight exons (Figure 3). Exon 1 encodes the 5′ untranslated region and the amino-terminal portion of the cytoplasmic segment. Exon 2 encodes the three amino acids located on the cytoplasmic side, the membrane-spanning segment and 23 amino acid residues on the extracytoplasmic side of the membrane. Six exons (3−8) encode the extracytoplasmic portion. The two sites for the addition of N-linked carbohydrate side chains are encoded by exon 4. Comparison of the murine Ii gene sequence with the human one revealed the same exon−intron structure. The homology between the exons was found to be 72−84% and between the introns ~50% (Table I).

Table I. Per cent homology between murine and human invariant chain exons and introns

% homology	Number of exon										
	5′NC	1	2	3	4	5	6	6b	7	8	3′NC
Exon	84	76	84	72	81	77	80	88	83	75	53
Intron			56		50	50	52	56	55		

Potential regulatory sequences in the 5′ and 3′ non-coding regions

Of the 5′ flanking region, 342 bp were sequenced (Figure 2). The sequences CATCT and TTTAA were found upstream of the ATG initiation codon (underlined in Figure 2). They show strong homology to the 'CAAT' and 'TATA' consensus sequences which are indispensable for specific initiation of transcription (Breathnach and Chambon, 1981). A consensus SpI protein binding site (GGGCGG) was found upstream of the TATA box (Gidoni *et al.*, 1984). The cap site of Ii gene transcription has not yet been determined. In the human Ii gene the start of transcription has been determined to be located 22 bp downstream of the proposed TATA box (Strubin *et al.*, 1984a). The analogous position in the murine gene is arbitrarily assigned +1. The sequence between the TATA box and the ATG initiation codon shows a repetitive CAG sequence characteristic for so-called OPA elements previously found in homoeotic and other genes (Wharton *et al.*, 1985). Its functional relevance remains to be shown.

As the expression of Ii gene is induced by IFN-γ we compared its 5′ sequence with those of other IFN-γ inducible genes such as the class II histocompatibilty antigens. Two elements, a 15-mer and a 8-mer, were previously suggested to be involved in the transcriptional regulation by IFN-γ (O'Sullivan *et al.*, 1986). Only the 15-mer sequence could be identified in the Ii gene (−228 to −213) (Figure 2). Control of Ii gene expression

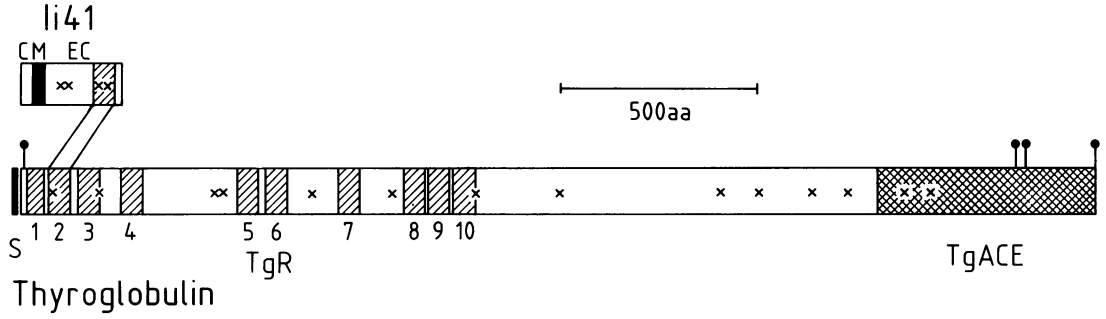

Fig. 4. Alignment of the homologous regions between Ii41 and thyroglobulin. The segment in Ii41 encoded by exon 6b and the 10 times repeated homologous regions in thyroglobulin (TgR) are indicated by boxes with slanted lines. Segments with the highest homology are connected by lines. Potential N-glycosylation sites are indicated by ×. C: indicates the cytoplasmic segment; M, membrane-spanning segment; EC, extracytoplasmic segment; S, signal sequence; TgACE, the segment homologous to acetylcholinesterase (Swillens *et al.*, 1986); ♀, hormonogenic tyrosines.

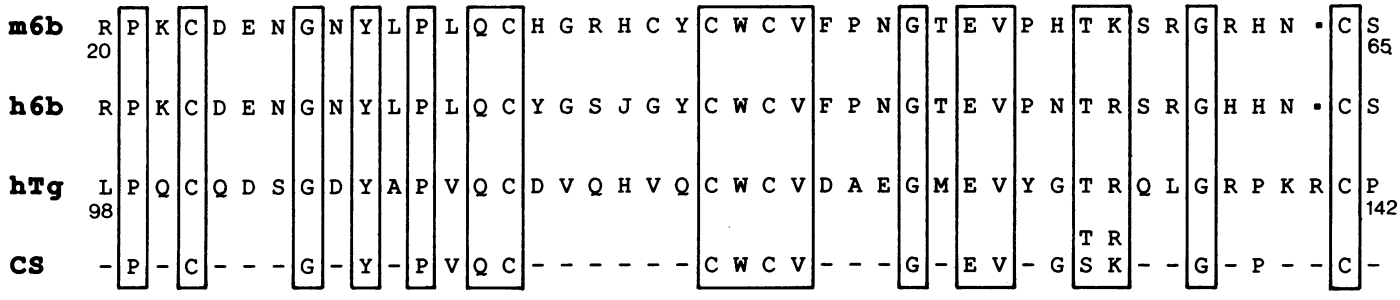

Fig. 5. Alignment of part of the amino acid sequence encoded by exon 6b of mouse (m) and human (h) invariant chain gene with the second repetitive units of human thyroglobulin (TgR) and a consensus TgR sequence. The consensus sequence (CS) has been established from a comparison of five of the human Tg repetitive units (Malthiery and Lissitzky, 1985). Dots indicate a gap, dashes indicate variable amino acid residues.

is also coupled to non-proliferation. The Ii gene and the metallothionine gene are inducible by the arrest of proliferation (Rahmsdorf *et al.*, 1983; Angel *et al.*, 1986). Between the TATA and the ATG start site for translation a region of homology was found.

Exon 6b encodes a cysteine-rich domain homologous to part of thyroglobulin

After transfer of the murine Ii gene into rat fibroblasts, two related invariant chain proteins were identified, one of 31 kd, Ii31, and one of 41 kd, Ii41 (Yamamoto *et al.*, 1985b). It has been suggested that the mRNAs coding for these two proteins are produced by differential splicing. In order to search for the sequence in the Ii gene that encodes Ii41 protein we used its known biochemical properties and the homology to the recently sequenced cDNA for the human 41-kd form of the invariant chain, In41 (Strubin *et al.*, 1986b). Ii41 should contain at least three sites for N-linked glycosylation. Ii31 and Ii41 should have common amino- and carboxy-terminal portions; and the additional segment in Ii41 should have a mol. wt of ~5 kd and be rich in cysteins (Yamamoto *et al.*, 1985b; Lipp and Dobberstein, 1986). Based on this information, a sequence between the sixth and seventh exons was found which fulfilled all the criteria for an additional exon used in the mRNA coding for the 41-kd protein. This exon 6b contains 192 bp between consensus splice sites. The resulting reading frame and the deduced amino acid sequence is shown in Figure 2. Exon 6b encodes 64 amino acid residues of which seven are cysteines. It has two potential sites for the addition of N-linked carbohydrate side-chains. The exon 6b shows 88% homology to the human p41-1 cDNA (Table I and Strubin *et al.*, 1986b).

The deduced protein sequence from exon 6b was compared to the sequences in the protein data base maintain by the National Biomedical Research Foundation, Washington, DC by using the search program FASTP (Lipman and Pearson, 1985). A significant homology of 38% was found by Patrick Argos (EMBL) to the repetitive sequence close to the amino terminus of thyroglobulin (Tg). This sequence centres around the tetrapeptide Cys-Trp-Cys-Val and is 10 times repeated in Tg (Figure 4) (Malthiery and Lissitzky, 1985; Mercken *et al.*, 1985a,b). A consensus sequence for the Tg repetitive elements (TgR) has been derived showing conserved positions for Cys, Pro and Gly residues (Malthiery and Lissitzky, 1985). When the amino acid sequence deduced from exon 6b was compared with this consensus sequence, nearly all positions were found to be conserved (Figure 5). Mouse and human 6b sequences were identical in all positions to the TgR consensus sequence (Figure 5). It is interesting to note that the deduced cysteine at position 39 of the mouse 6b exon is not conserved in the human 6b exon and is also not part of the TgR consensus sequence. In contrast, all the other cysteine residues are conserved between the TgR and the 6b segment.

Discussion

The sequence and exon–intron structure of the murine Ii chain gene were determined. When compared to the human invariant chain gene, a high similarity in the overall structural organization, and particularly between the exons, was found. However, one significant difference was observed in the 5′ untranslated regions. The human In33 and In35 proteins result from alternative initiation at two in-phase AUG codons (Strubin *et al.*, 1986a). Only the second initiation site was found in the 5′ untranslated region of the mouse Ii mRNA (Figure 2). It thus ap-

pears that the 35-kd form in man might not have an essential function distinct from the 33-kd one.

The 41-kd form of invariant chain, in contrast, is found in all species which have been screened for its presence. This form is the result of alternative splicing (Yamamoto *et al.*, 1985b; Strubin *et al.*, 1986b). The exon used in this event is located between exons 6 and 7 and is therefore named 6b. Interestingly, when exons of the human and murine invariant chain gene were compared, exons 6b showed the highest homology, 88% homology was found between exons 6b, whereas 72–84% between the others (Table I).

The mechanism of alternative splicing is not yet understood. Examination of introns between the sixth and seventh exons did not reveal any obvious sequence motifs which might effect efficiency of splicing. Several other genes have been found to employ alternative splicing. These include the T3δ gene in human T cells (Tunnacliffe *et al.*, 1986) and the H-2 class I genes (Kress *et al.*, 1983; Transy *et al.*, 1984). In the T3δ gene a stretch of 44 bp of alternating GT was found to flank the alternatively spliced exon (Tunnacliffe *et al.*, 1986). Stretches of alternating GT are also found in the Ii gene between exons 6 and 7. Their significance for alternative splicing remains to be shown.

Exon–intron organization of the Ii gene is very similar to that found for the asialoglycoprotein receptor (ASGR) gene (Leung *et al.*, 1985). ASGR, like the Ii31 and Ii41 proteins, is a type II membrane protein. The cytoplasmic and membrane spanning segment of ASGR are each encoded by separate exons. Five exons encode the extracytoplasmic segment. The carbohydrate binding site in ASGR has been localized in the carboxy-terminal segment encoded by exons 7–9. The functional domain in Ii chains has not yet been identified.

Comparison of the biochemical properties of the Ii31 and Ii41 chain had revealed an extensive similarity. An additional cysteine-rich domain has been postulated for the Ii41 chain (Lipp and Dobberstein, 1986). Exon 6b codes for 64 amino acid residues, seven of which are cysteines. This segment in the Ii41 protein is located close to the carboxy terminus on the extracytoplasmic side of the membrane (Figure 3). Cysteine residues in secretory and membrane proteins are often found to organize structurally and functionally distinct domains. Best examples are protein domains of the class I and II histocompatibility antigens (Nathenson *et al.*, 1981) immunoglobulins (Sakano *et al.*, 1979) and the low density lipoprotein (LDL) receptor (Yamamoto *et al.*, 1984). A structural motif of repeated, cysteine-rich sequences was, for instance, demonstrated for the human EGF and LDL receptors (Rall *et al.*, 1985). When a protein data base was searched for sequences homologous to the segment encoded by exon 6b, a striking homology of 38% was found to a cysteine-rich segment in thyroglobulin (Tg). This segment is 10 times repeated in the amino-terminal half of Tg (Figure 4). A consensus sequence for the cysteine-rich TgR has been proposed. It centres around the sequence motif Cys-Trp-Cys-Val (Malthiery and Lissitzky, 1985). The entire TgR consensus sequence is found conserved in the sequence of mouse and human exon 6b (Figure 5).

What could be the structural and functional significance of such an extensive homology? Tg is an iodinated precursor protein for the production of thyroid hormone (Wollman, 1969). Bovine Tg is a glycosylated phosphorylated and sulfated protein of 2750 amino acid residues (for review see Herzog, 1984). It is a dimeric glycoprotein of 660 kd which is secreted by the thyrocytes and stored in the lumen of the thyroid follicle. Here the protein becomes iodinated at tyrosyl residues and at 3–4 of these residues thyroxine (T_4) and triiodothyronine (T_3) are formed. These are located at the extreme ends of Tg on amino acid residues 5, 2555, 2569 and 2748 (Figure 4 and Mercken *et al.*, 1985b). Active hormone is released after endocytosis or phagocytosis of thyroglobulin and its hydrolysis most likely in lysosomes. Some Tg, however, seems to escape lysosomal degradation and thus appears intact in the serum (Van Herle *et al.*, 1979; Herzog, 1984). It has been suggested that the large thyroglobulin structure has evolved for efficient and regulated iodination and coupling of the hormonogenic tyrosines.

The exon 6b of the Ii gene codes for 64 amino acid residues. Its amino acid sequence between residues 20 and 64 is as homologous to the second TgR as the 10 Tg repeats are among each other (Figures 4 and 5). It has been proposed that the TGRs have arisen by gene duplication of a primordial gene coding for a 60 amino acid long building block (Musti *et al.*, 1986). This suggestion is further supported by the location of exon–intron boundaries within the Tg gene. Most of the TgR units are encoded by separate exons (Musti *et al.*, 1986; R.Di Lauro, personal communication). Therefore, the exon 6b of the Ii gene might be derived from the same primordial building block as the 10 TgR units in the Tg gene. As the homology is high between the TgR segments in Ii41 and in Tg, these two segments might perform similar functions.

The function of the 10 TgR elements in Tg is not known. It has been suggested that the unusually large Tg protein structure supports the efficiency of iodination and the formation of the hormonogenic tyrosines. It is conceivable that TgR segments function in the formation of iodinated hormones. Hormone formation on the TgR molecule occurs outside the cell in the thyroid follicle. If Ii41 was a hormone precursor similar to Tg, one should find it on the cell surface. No clear evidence for a cell surface location has, however, been found for Ii41 protein. Clearly, more detailed studies are required to elucidate a possible hormone function of Ii41.

Tg undergoes extensive intracellular transport. It is a typical secretory protein which is secreted into the thyroid follicle, iodinated, endocytosed, degraded in the lysosomes and the hormone finally released in the circulation (Herzog, 1984; Vassart *et al.*, 1985). There must be some structural elements in Tg that direct this molecule to the different stations. The TgR element could function in the transport to the lysosomes or in the transport of T_4 and T_3 out of the lysosomes to the basal cell surface. A function involving a lysosomal or acidic compartment has also been suggested for invariant chains. This was largely based on the transient association of Ii31 and Ii41 with class II MHC antigens and their dissociation in an acid compartment (Machamer and Cresswell, 1984; Nowell and Quaranta, 1985). Class II MHC antigens are involved in the presentation of foreign antigens to T cells. Most antigens have to be processed before they can efficiently be presented. Processing appears to occur in an intracellular acidic compartment and involves in most cases proteolytic degradation (for review see Unanue, 1984). The two best-characterized acidic compartments in the cell are the endosomes and the lysosomes.

Digestive enzymes are well characterized in lysosomes. It has always been an enigma how processed antigen is retrieved from the processing compartments, associates with class II antigens and appears in association with class II antigens on the cell surface. Could Ii31 and Ii41 perform functions in the retrieval of antigen from different processing compartments? Each of these molecules might then serve a different route, one an endosome-like compartment (Ii31) and one the digestive lysosome compartment (Ii41).

It is not yet known in which form thyroid hormone reaches the cell surface and enters the circulation. Three to four hormonogenic peptides are released from one molecule of thyroglobulin. It is conceivable that the 10 TgRs are involved in the transport of hormone or hormonogenic peptides from the lysosomes to the cell surface where the hormone is released into the circulation. If this assumption is correct, then the TgR elements in Ii41 would be a carrier for processed antigen and in Tg for hormonogenic peptides or hormone between the lysosomes and an endosome-like compartment or the cell surface.

Materials and methods

Cosmid and plasmids

The genomic Ii chain clone cos 10.7 containing the entire gene for Ii chain was selected from a cosmid library made from AKR mouse DNA. It was obtained from M.Steinmetz, Basel (Yamamoto *et al.*, 1985b). Plasmid pIi-5 containing most of the coding sequence of Ii chain and the 3′ non-coding region has been described previously (Singer *et al.*, 1984). It lacks the sequences coding for the cytoplasmic segment of Ii and part of the membrane-spanning region (Singer *et al.*, 1984). Plasmid pγ2 was obtained from P.A.Peterson, Sweden. It encodes the entire human invariant chain (Claesson *et al.*, 1983). Plasmid, Ii-5 was used to locate the 3′ end of the Ii gene and the 5′ 320-bp *Pst*I fragment was used to identify its 5′ end.

DNA mapping, subcloning and sequencing

A restriction map of the 40-kb insert in clone cos 10.7 was established by using the methods of Rackwitz *et al.* (1985) and Zehetner and Lehrach (1986). By Southern blotting, using murine and human cDNAs as probes, two adjacent *Eco*RI fragments of 2.9 and 10 kb were identified to contain the Ii gene (Southern, 1975) (Figure 1). The 2.9-kb *Eco*RI fragment was further deletion subcloned by the method of Frischauf *et al.* (1980). Size-selected subclones were sequenced either by the Sanger dideoxy-chain termination method (Sanger *et al.*, 1977) or by the method of Labeit using α-phosphorothioates (Labeit *et al.*, 1986). In the latter method the α-thiotriphosphate analogs of deoxynucleoside triphosphates are used to incorporate exonuclease III-resistant residues into DNA (Labeit *et al.*, 1987). The 10-kb *Eco*RI fragment was digested with *Hin*dIII and the resulting fragments subcloned into pUC8 or pBr322 (see Figure 1 for details).

The 3′ end of the 2.3-kb *Eco*RI–*Hin*dIII fragment was sequenced after cloning into pBR 322. The 0.75-, 3.7- and 0.8-kb fragments were cloned into pUC8 and their 5′ and 3′ ends sequenced. The 3.7-kb fragment was deletion subcloned (Frischauf *et al.*, 1980) and selected fragments with deletions in their 5′ end, sequenced. The localization of the fragments is shown in Figure 1B.

Acknowledgements

We thank M.Steinmetz for the Ii genomic clone cos 10.7, P.Argos for search of the protein data base, M.Burmeister, S.Labeit, A.Nordheim, W.Röwekamp and K.Seedorf for many helpful discussions, M.-T.Haeuptle, V.Herzog, W.Huttner, S.Kvist, and K.Simons for critical comments on the manuscript, A.Steiner for expert typing of the manuscript and P.Riedinger for the drawings. This work was supported by grants from the Deutsche Forschungsgemeinschaft, Ko 810/2-2 and Do 199/4-3.

References

Angel,P., Pöting,A., Mallick,U., Rahmsdorf,H.J., Schorpp,M. and Herrlich,P. (1986) *Mol. Cell Biol.*, **6**, 1760–1766.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Charron,D.J., Aellen-Schulz,M.-F., St Geme,J.,III, Erlich,H.A. and McDewitt,H.O. (1983) *Mol. Immunol.*, **20**, 21–32.
Claesson,L. and Peterson,P.A. (1983) *Biochemistry*, **22**, 3206–3213.
Claesson,L., Larhammar,D., Rask,L. and Peterson,P.A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7395–7399.
Claesson-Welsh,L., Barker,P.E., Larhammar,D., Rask,L., Ruddle,F.H. and Peterson,P.A. (1984) *Immunogenetics*, **20**, 89–93.
Collins,T., Korman,A.J., Wake,C.T., Boss,J.M., Kappes,D.J., Fiers,W., Ault,K.A., Gimbrone,M.A., Strominger,J.L. and Pober,J.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4917–4921.
Cresswell,P. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8188–8192.
Frischauf,A.-M., Garoff,H. and Lehrach,H. (1980) *Nucleic Acids Res.*, **8**, 5541–5549.
Gidoni,D., Dynan,W.S. and Tjian,R. (1984) *Nature*, **312**, 409–413.
Herzog,V. (1984) *Int. Rev. Cytol.*, **91**, 107–139.
Jones,P.P., Murphy,D.B., Hewgill,D. and McDewitt,H.O. (1978) *Immunochemistry*, **16**, 51–60.
Koch,N. and Harris,A.W. (1984) *J. Immunol.*, **132**, 12–15.
Koch,N., Wong,G.H.W. and Schrader,J.W. (1984) *J. Immunol.*, **132**, 1361–1369.
Koch,N. and Hämmerling,G.J. (1986) *J. Biol. Chem.*, **261**, 3434–3440.
Kress,M., Glaros,D., Khoury,G. and Jay,G. (1983) *Nature*, **306**, 602–604.
Kudo,J., Chao,L.-Y., Narni,F. and Saunders,G.F. (1985) *Nucleic Acids Res.*, **13**, 8827–8841.
Kvist,S., Wiman,K., Claesson,L., Peterson,P.A. and Dobberstein,B. (1982) *Cell*, **29**, 61–69.
Labeit,S., Lehrach,H. and Goody,R.S. (1986) *DNA*, **5**, 173–177.
Labeit,S., Lehrach,H. and Goody,R.S. (1987) *Anal. Biochem.*, in press.
Leung,J.O., Holland,E.C. and Drickamer,K. (1985) *J. Biol. Chem.*, **260**, 12523–12527.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Lipp,J. and Dobberstein,B. (1986) *J. Cell Biol.*, **102**, 2169–2175.
Long,E.D. (1985) *Survey Imunol. Res.*, **4**, 27–34.
Machamer,C.E. and Cresswell,P. (1982) *J. Immunol.*, **129**, 2564–2569.
Machamer,C.E. and Cresswell,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1287–1291.
Malthiery,Y. and Lissitzky,S. (1985) *Eur. J. Biochem.*, **147**, 53–58.
Mercken,L., Simons,M.-J., De Martynoff,G., Swillens,S. and Vassart,G. (1985a) *Eur. J. Biochem.*, **147**, 59–64.
Mercken,L., Simons,M.-J., Swillens,S., Massaer,M. and Vassart,G. (1985b) *Nature*, **316**, 647–651.
Miller,J. and Germain,R.N. (1986) *J. Exp. Med.*, **164**, 1478–1489.
Momburg,F., Koch,N., Möller,P., Moldenhauer,G., Butcher,G.W. and Hämmerling,G.J. (1986) *J. Immunol.*, **136**, 940–948.
Musti,A.M., Avvedimento,E.V., Polistina,C., Ursini,V.M., Obici,S., Nitsch,L., Cocozza,S. and Di Laura,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 323–327.
Nathenson,S.G., Uehara,H., Ewenstein,B.M., Kindt,T.J. and Coligan,J.E. (1981) *Annu. Rev. Biochem.*, **50**, 1025–1052.
Nowell,J. and Quaranta,V. (1985) *J. Exp. Med.*, **162**, 1371–1376.
Omary,M.B. and Trowbridge,I.S. (1981) *J. Biol. Chem.*, **256**, 12888–12892.
O'Sullivan,D., Larhammar,D., Wilson,M.C., Peterson,P.A. and Quaranta,V. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4484–4488.
Polla,B.S., Poljak,A., Geier,S.G., Nathenson,S.G., Ohara,J., Paul,W.E. and Glimcher,L.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4878–4882.
Quaranta,V., Majdic,O., Stingl,G., Liszka,K., Honigsmann,H. and Knapp,W. (1984) *J. Immunol.*, **132**, 1900–1905.
Quill,H. and Schwartz,B.D. (1983) *Mol. Immunol.*, **12**, 1333–1345.
Rall,L.B., Scott,J., Bell,G.J., Crawford,R.J., Penschow,J.D., Niall,H.O. and Coghlan,J.P. (1985) *Nature*, **313**, 228–231.
Rackwitz,H.R., Zehetner,G., Murialdo,H., Delius,H., Chai,J.-H., Poustka,A., Frischauf,A. and Lehrach,H. (1985) *Gene*, **40**, 259–266.
Rahmsdorf,H.J., Koch,N., Mallick,U. and Herrlich,P. (1983) *EMBO J.*, **2**, 811–816.
Rudd,C.E., Bodmer,J.G., Bodmer,W.F. and Crumpton,M.J. (1985) *J. Biol. Chem.*, **260**, 1927–1936.
Sakano,H., Rogers,J.H., Hüppi,K., Brack,C., Traunecker,A., Maki,R., Wall,R. and Tonegawa,S. (1979) *Nature*, **277**, 627–633.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schneider,C., Owen,M.J., Banville,D. and Williams,J.G. (1984) *Nature*, **311**, 675–678.
Sekaly,R.P., Torelle,C., Strubin,M., Mach,B. and Long,E.O. (1986) *J. Exp. Med.*, **164**, 1490–1504.
Singer,P.A., Lauer,W., Dembic,Z., Mayer,W.E., Lipp,J., Koch,N., Hämmerling,G.J., Klein,J. and Dobberstein,B. (1984) *EMBO J.*, **3**, 873–877.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Strubin,M., Long,E.O. and Mach,B. (1986a) *Cell*, **47**, 619–625.
Strubin,M., Berte,C. and Mach,B. (1986b) *EMBO J.*, **5**, 3483–3488.
Sung,E., Duncan,W.R., Streilein,J.W. and Jones,P.P. (1982) *Immunogenetics*, **16**, 425–433.
Swillens,S., Ludgate,M., Mercken,L., Dumont,J.E. and Vassart,G. (1986) *Biochem. Biophys. Res. Commun.*, **137**, 142–148.
Transy,C., Lalanne,J.-L. and Kourilsky,P. (1984) *EMBO J.*, **3**, 2383–2386.
Tunnacliffe,A., Sims,J.E. and Rabbitts,T.H. (1986) *EMBO J.*, **5**, 1245–1252.
Unanue,E.R. (1984) *Annu. Rev. Immunol.*, **2**, 395–428.
Van Herle,A.J., Vassart,G. and Dumont,J.E. (1979) *New Engl. J. Med.*, **301**, 239–249.
Vassart,G., Bacolla,A., Brocas,H., Christophe,D., Martynoff,G.de, Lerich,A., Mercke,L., Parma,J., Pohl,V., Targovnik,H. and van Henverswyn,B. (1985) *Mol. Cell Endocrinol.*, **40**, 89–97.
Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55–62.

Wollman,S.H. (1969) In Dingle,J.T. and Fell,H.B. (eds), *Lysosomes in Biology and Pathology*. North-Holland, Amsterdam, pp. 483–512.
Yamamoto,K., Floyd-Smith,G., Francke,U., Koch,N., Lauer,W., Dobberstein,B., Schäfer,R. and Hämmerling,G.J. (1985a) *Immunogenetics*, **21**, 83–90.
Yamamoto,K., Koch,N., Steinmetz,M. and Hämmerling,G.J. (1985b) *J. Immunol.*, **134**, 3461–3467.
Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L., Goldstein,J.L. and Russell,D.W. (1984) *Cell*, **39**, 27–38.
Zecher,R., Ballhausen,W., Reske,K., Linder,D., Schlüter,M. and Stirm,S. (1984) *Eur. J. Immunol.*, **14**, 511–517.
Zehetner,G. and Lehrach,H. (1986) *Nucleic Acids Res.*, **14**, 335–349.

Received on February 19, 1987; revised on April 2, 1987