FOUR Ia INVARIANT CHAIN FORMS DERIVE FROM A SINGLE GENE BY ALTERNATE SPLICING AND ALTERNATE INITIATION OF TRANSCRIPTION/TRANSLATION

By DEIRDRE M. O'SULLIVAN, DANIEL NOONAN, and VITO QUARANTA

From the Department of Immunology, Research Institute of Scripps Clinic, La Jolla, California 92037

The Ia molecules (HLA-DR, -DP, -DQ in man) are central elements of the immune system, restricting recognition of foreign antigens to histocompatible T lymphocytes (1). They are composed of two subunits, α and β, encoded by the class II genes of the MHC (2), and by a third subunit, the invariant or γ chain (3), encoded outside the MHC (4). Ia molecules are genetically polymorphic. Most of this structural polymorphism is contributed by the β chains, and to a minor extent by the α chains (2).

Although no allelic polymorphism was detected for the invariant chain, several lines of evidence suggested that alternative protein forms of this molecule are part of Ia oligomers. Two-dimensional (2-D)[1] analyses indicated that, in addition to spots corresponding to α chains (M_r 34,000), β chains (M_r 29,000) and to the principal invariant chain (M_r 31,000, hereafter termed γ1), other spots could be consistently detected (5, 6). Two of these spots were shown to be antigenically related to the γ1 chain by reactivity with an anti-γ1 monoclonal antibody, and were therefore termed γ2 and γ3 (6). Moreover, an M_r 41,000 protein (p41), also detected as Ia-associated on 2-D gel is also antigenically related to the γ1 chain because of its reactivity with antipeptide antisera produced against γ1 (O'Sullivan, unpublished observations). The γ2, γ3, and p41 proteins could not be accounted for by posttranslational modifications of the γ1 chain, because their precursors are apparently translated in vitro under conditions in which posttranslational processing is thought to be absent (6). On the other hand, isolation of genomic clones and genomic hybridization analyses indicated that γ1 chain cDNA probes hybridize to a single-copy gene (7, 8), making it unlikely that these proteins are encoded by a family of related genes. The possibility then remained that these antigenically related forms of invariant chains are encoded by a single gene through alternative processing of its transcript(s), as we previously suggested (7). We therefore investigated patterns of splicing, of transcrip-

This work was supported by NIH grants AI-10088 and GM-37191, and by Leukemia Society of America fellowship to D. M. O'Sullivan and a scholarship to V. Quaranta. Part of this work was presented in abstract form at the Sixth International Congress of Immunology, Toronto, Canada, July 1986. This is publication 4549IMM from the Research Institute of Scripps Clinic and Research Foundation, La Jolla, CA.

[1] *Abbreviation used in this paper:* 2-D, two-dimensional.

 J. Exp. Med. © The Rockefeller University Press · 0022-1007/87/08/0444/17 $2.00
Volume 166 August 1987 444–460

tional and translational initiation in transcripts from the invariant chain gene, which we recently cloned and characterized (7).

In this paper we demonstrate that invariant chain transcripts undergo alternate splicing and alternate initiation of translation. In addition, we mapped a minimum of three transcriptional start sites in the invariant chain gene, in contrast to previously reported single start sites (9, 10). We further show that these mechanisms are responsible for the expression of four closely related but structurally distinct invariant chains from a single gene.

Materials and Methods

Cell Lines. The human B lymphoblastoid cell lines Raji, 721-P, and LG-2, and the African green monkey kidney cell line COS-7 were grown in DMEM medium supplemented with 10% FCS, glutamine, and antibiotics. They can be obtained from the American Tissue Culture Collection, Rockville, MD. COS-7 cells support replication at high copy number of plasmids containing the SV40 origin of replication, since they express the SV40 large T antigen from an integrated gene (11).

Antisera. The anti–invariant chain murine monoclonal antibody VIC-Y1 has been described (6). Rabbit antisera against COOH-terminal peptides (C351) of the human γ1 chain have also been described elsewhere (12).

cDNA Libraries and Plasmids. mRNA was isolated from the cell line Raji by oligo-d(T) chromatography of total cytoplasmic RNA (13) obtained by phenol extraction and ethanol precipitation of detergent-solubilized cells, following removal of nuclei by low-speed centrifugation. cDNA libraries were constructed in the SV40-based expression vector, pcD, exactly as described by Okayama and Berg (14). The pcD vector contains the SV40 ori region, that permits replication in mammalian cells (14). Expression is directed by SV40 promoter, splicing, and polyadenylation sequences (14). The cDNA library was screened by colony hybridization (13), using a 5′ Pst I fragment of an invariant chain partial cDNA clone (15), ^{32}P-labeled by nick-translation (13). Positive clones were characterized by restriction mapping and nucleotide sequencing of their 5′ ends by the chemical method (16). Plasmids were purified by the alkaline lysis method (13), followed by two CsCl gradients (13). The plasmid P371 was constructed by subcloning a human genomic fragment comprising the entire invariant chain gene (7) in the plasmid pUC9.

Site-directed Mutagenesis. The plasmid pcDG-1 was mutagenized at a single base by the gapped plasmid method, as described (17). Briefly, the plasmid was either cut with the restriction endonuclease Cla I, or double cut with Sac II and Xmn I. Single and double cut plasmids were mixed in a 1:1 ratio, melted (100°C for 10 min) and then reannealed in the presence of a 18 bp oligonucleotide (see below) in which the G residue of the second ATG codon was replaced by a C. Incorporation of this oligonucleotide in the reannealed plasmid resulted in the substitution of an ATC codon for the ATG (isoleucine for methionine), and in the creation of a Cla I restriction site (ATCGAT). After repair with DNA polymerase III, the plasmid mixture was used to transform (13) *Escherichia coli* HB101. Mutant clones were identified by digestion of plasmid DNA minipreps with the enzyme Cla I. Mutants were found at a frequency of ~15%. Nucleotide sequencing (16) confirmed that the selected clone, pcDG-1M, contained the desired single-base change. The oligonucleotide AGCCAGTCATCGATGACC was synthesized on an Applied Biosystems (Foster City, CA) automatic synthesizer, according to the instruction manual, and was purified by polyacrylamide gel electrophoresis (13).

S1 Nuclease Protection Assays. Cytoplasmic poly(A)$^+$ RNA from B lymphoid cells, isolated as described above, was hybridized to a ^{32}P–end-labeled (13) Nar I–Bam HI human genomic fragment comprising the first exon of the invariant chain gene and 600 bp upstream. After digestion with S1 nuclease, the mixture was electrophoresed on a denaturing polyacrylamide gel (6% acrylamide, 0.8 M urea), in parallel with appropriate sequencing reactions of the same end-labeled fragment. The gel was dried and exposed on XRP-1 Kodak film. Densitometry of these autoradiograms was performed with an LKB soft-laser scanner (LKB Instruments, Bromma, Sweden).

Transfections, Cell Labeling and Lysis, Immunoprecipitation, and Electrophoresis. COS-7 cells were transfected by the calcium phosphate precipitation method (18). After 36 h at 37°C, the cells were incubated for 2 h in methionine-free medium containing 0.1 mCi/ml [^{35}S]methionine, then rinsed with saline and detergent lysed as described (6) in 2% of the nonionic detergent Renex-30 (Accurate Chemical and Scientific Corp., Westbury, NY). Radiolabeled Raji cell lysates were obtained by the same procedure. Immunoprecipitations with antibody–protein A–Sepharose immunosorbents were performed as previously reported (6). 2-D nonequilibrium pH gradient electrophoresis (19) and one-dimensional gel electrophoresis (20) were carried out as published. Gels were fluorographed (21), dried, and exposed to Kodak XAR-5 film for indicated times.

Results

We constructed cDNA libraries in the simian virus 40 (SV40)-based expression vector, pcD, described by Okayama and Berg (14), using mRNA isolated from the human Ia$^+$ lymphoblastoid cell line, Raji. The libraries were screened by colony hybridization with a γ1 chain partial cDNA (15) insert. ~150 positive colonies were isolated from a single library, with a frequency of ~1 in 100, in agreement with the estimated high-abundance of invariant chain–specific message (0.5–1%) (6, 15).

Positive clones were rescreened by hybridization to a labeled 5′ probe isolated from the γ1 chain partial cDNA insert. Clones hybridizing to this probe were characterized by restriction mapping, and those containing the longest inserts were further analyzed by sequencing their 5′ ends. From these analyses, several putative near-full-length cDNAs were identified. Two types of cDNA clones were subsequently studied. These differed by an apparent truncation of ~40 bp at the 5′ end. Inserts representative of the first (pcDG-1) and second (pcDG-4) type are illustrated in Fig. 1. The relevant feature that distinguished the two cDNA types was that the longer type contained two 5′ ATGs, both in-phase with the open reading frame, while the other type lacked the 5′-most of these ATGs. To determine whether either of these ATGs could be used as translational start sites of distinct proteins, the plasmids were used for transient expression assays by transfection in the African green monkey kidney cell line, COS-7. 36 h after transfection, the cells were metabolically radiolabeled, detergent lysed, and immunoprecipitated with either monoclonal antibodies or antipeptide antisera specific for the Ia γ1 chain (Fig. 2*A*). Upon transfection with pcDG-1, which contains both AUG codons (AUG/AUG), two bands were immunoprecipitated (Fig. 2*A*, lane *D*), with M_r 31,000 and 33,000 respectively, closely resembling authentic invariant-chain bands from the B lymphoid line Raji (Fig. 2*A*, lane *F*). Only the M_r 31,000 band, however, was immunoprecipitated from cells transfected with a cDNA clone, pcDG-4, which lacks the first AUG (—/AUG) (Fig. 2*A*, lane *C*). Immunoprecipitations with mock-transfected COS-7 cells were completely negative (Fig. 2*A*, lane *A*). Subsequent 2-D analyses indicated that the proteins expressed from pcDG-1 corresponded to γ1, γ2, and γ3 chains, while the ones expressed by pcDG-4 corresponded to γ1 chains, since they comigrated precisely with authentic invariant-chain spots from the B cell line Raji (Fig. 2*B*). These data suggested that, in these cDNA clones, both AUG codons can be used to initiate translation, giving rise to either γ2/γ3 or γ1 invariant-chain forms.

To directly prove that this was, in fact, the case, plasmids containing the first

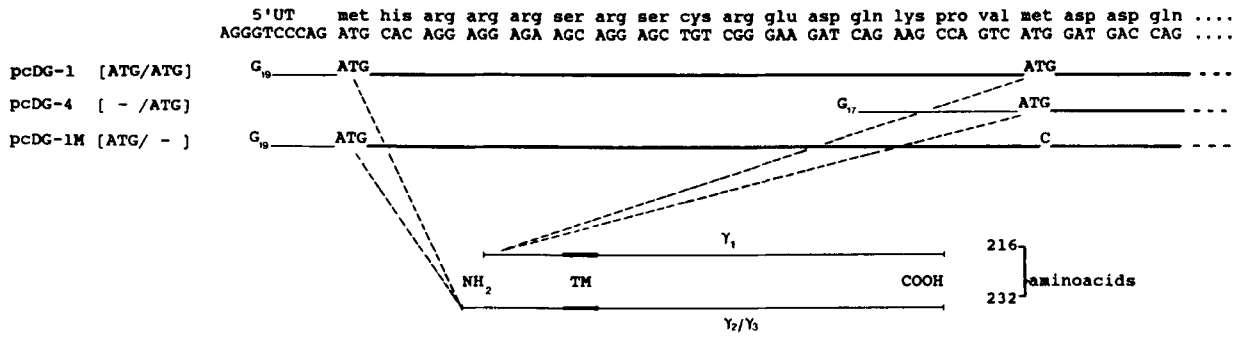


FIGURE 1. Diagram of the 5′ ends of invariant-chain cDNA inserts isolated from a Raji cell cDNA library in the expression vector pcD, by colony hybridization to an invariant chain partial cDNA probe. The inserts of the displayed plasmids, pcDG-1 and pcDG-4, are nearly full-length, as indicated by the nucleotide sequence of the invariant chain gene exon 1, shown by comparison (*5′UT*, 5′ untranslated region). The G tails (G_{19}, G_{17}) are part of the cloning vector. The two 5′ ATG in phase with the open reading frame are shown (see text for further explanation). Plasmid pcDG-1M was derived from pcDG-1 by site-directed mutagenesis of the second of these ATG to ATC. The bottom part of the figure schematically depicts that both of these ATGs are functional as translational start sites: two proteins of 216 and 232 amino acids, respectively, can be translated from pcDG-1 inserts, while pcDG-4 and pcDG-1M give rise to the 216 or the 232 protein, respectively. These proteins correspond to the Ia $\gamma 1$ and $\gamma 2/\gamma 3$ chains. (*TM*, transmembrane segment).

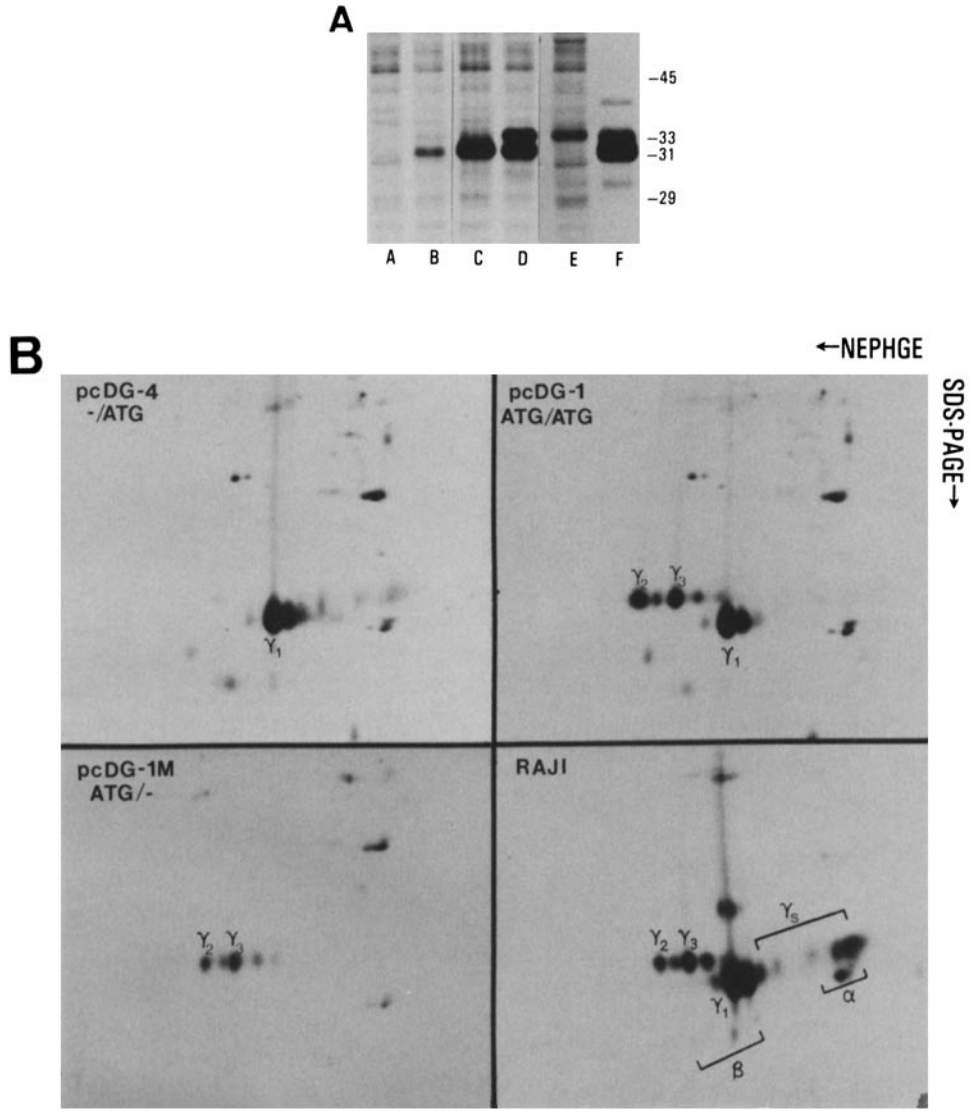


FIGURE 2. (*A*) One-dimensional gel electrophoresis analyses of anti-invariant chain monoclonal antibody VIC-Y1 immunoprecipitates with radiolabeled detergent lysates from either Raji cells (lane *F*), or COS-7 cells transfected with the following plasmids: pcDG-1M (ATG/—, lane *E*); pcDG-1 (ATG/ATG, lane *D*); pcDG-4 (—/ATG, lane *C*); p371, which contains the entire invariant-chain gene (lane *B*); and no-DNA control (lane *A*). Film was exposed for 5 d. (*B*) 2-D analyses of immunoprecipitates as in lanes *C–F* from Fig. 2*A* (clockwise from top left). Film exposure time was 2 wk.

but not the second AUG were generated by site-directed mutagenesis of the plasmid pcDG-1 (Fig. 1). The G residue of the downstream AUG was replaced with a C, thereby preempting initiation of translation from this codon. This mutation alters the encoded methionine to an isoleucine, a relatively neutral change for protein structure. Transfection with the mutagenized plasmid, pcDG-1M (AUG/—), resulted in expression of γ2 and γ3, but not γ1 chains, as

TABLE I
Relative Abundance of Invariant Chain Protein Forms

Cell line	Label uptake			Chain form ratios	
	γ1	γ2/γ3	p41	γ1:γ2/γ3	γ1:p41
	cpm				
Raji	1,881	416	188	4.5	10
	2,163	437	153	4.9	14
	567	122	68	4.6	8.3
COS*	214	134		1.6	
	239	189		1.3	
	269	204		1.3	

* COS-7 cells transfected with the plasmid pcDG-1 (AUG/AUG).

indicated by one-dimensional (Fig. 2*A*, lane *E*) and 2-D (Fig. 2*B*) analyses of anti–invariant chain immunoprecipitates. Together, these results demonstrated that the downstream AUG is the start site for translation of the principal M_r 31,000 γ1 chain, and that the first AUG is used to initiate translation of the γ2 and γ3 forms. Note that invariant chains lack a cleavable leader peptide. The heterogeneity of the M_r 33,000 proteins, generating the γ2 and γ3 spots, is probably due to posttranslational modifications (Fig. 2*B*; and see Discussion).

Although transfections with the plasmid pcDG-1 resulted in expression of both γ2/γ3 and γ1 chains identical to the ones observed in B cells, the relative ratios between these invariant chain forms were altered, as indicated by the relative intensity of γ2/γ3 and γ1 bands in Fig. 2*A* (lanes *C–F*), and 2-D spots in Fig. 2*B*. These differences were quantitated by excising bands and/or 2-D spots from gels, followed by determination of radioactivity (Table I). In pcDG-1–transfected cells, the γ2/γ3 to γ1 ratio was ~1.5:1, while in B cells it was ~5:1 (Table I). This higher ratio was not dependent on the cell type used for transfection, since transfections of COS-7 cells with a cloned human invariant chain chromosomal gene resulted in γ2/γ3:γ1 ratios similar to that of B cells (Fig. 2*A*, lane *B*) (Table I).

This high frequency of translational start at the second AUG in B cells raised therefore the possibility that mRNA species lacking the first AUG may exist in B cells. Thus, we examined the start site(s) for transcription in the human invariant chain gene by S1 nuclease protection experiments. The strategy for these assays is detailed in Materials and Methods, and is schematized in Fig. 3. Results of these experiments, performed with mRNA from three different B lymphoid lines, revealed three clusters of protected fragments, lying at approximately −5, −10, and −32 bp upstream of the first ATG codon of the open reading frame (Fig. 3). The 2–3 bp difference in length among fragments within a cluster may be due either to ragged ends generated by the S1 digestion, or may reflect microheterogeneity in the start sites for transcription. However, for simplicity, the initiation sites will be referred to as −5, −10, and −32. Even in long exposures of the film, no protected fragments initiating downstream from the first ATG were detected. These results indicated that, in B cells, all mature invariant chain mRNA species contain both the AUG codons identified as

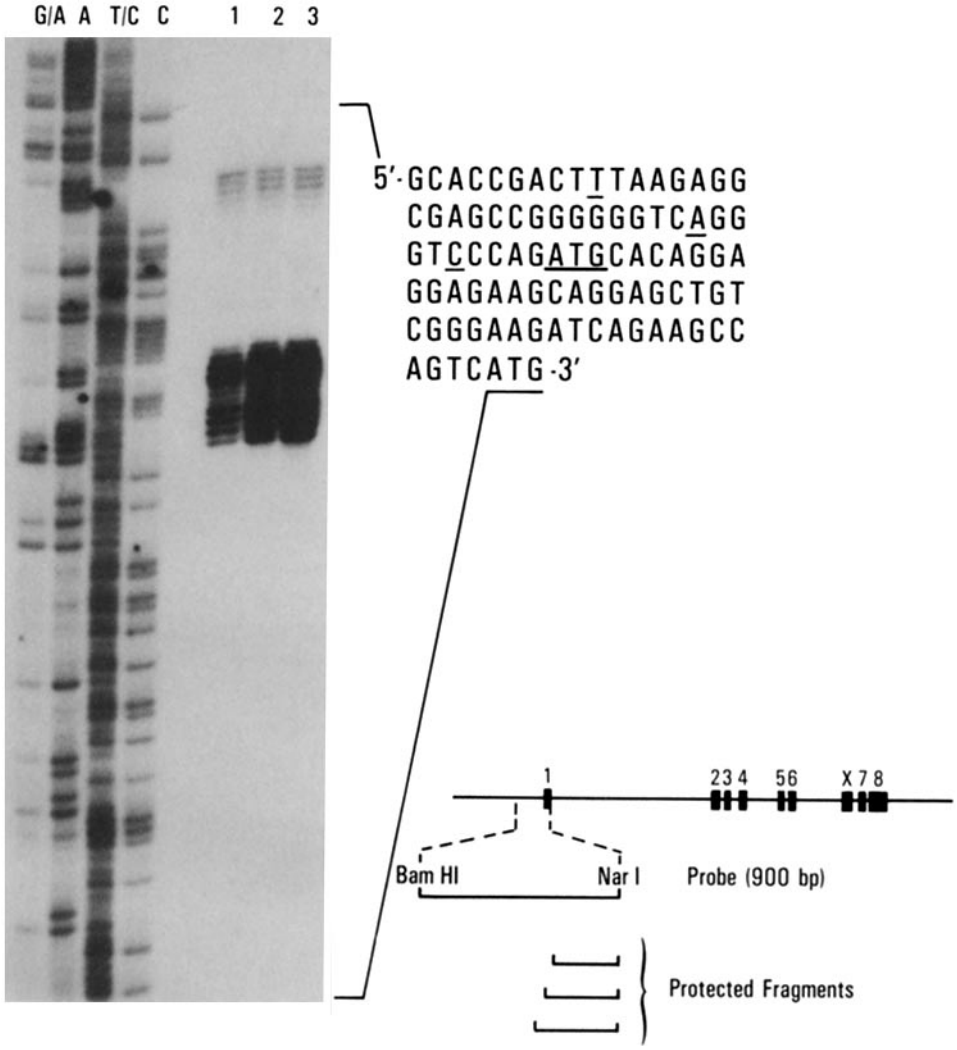
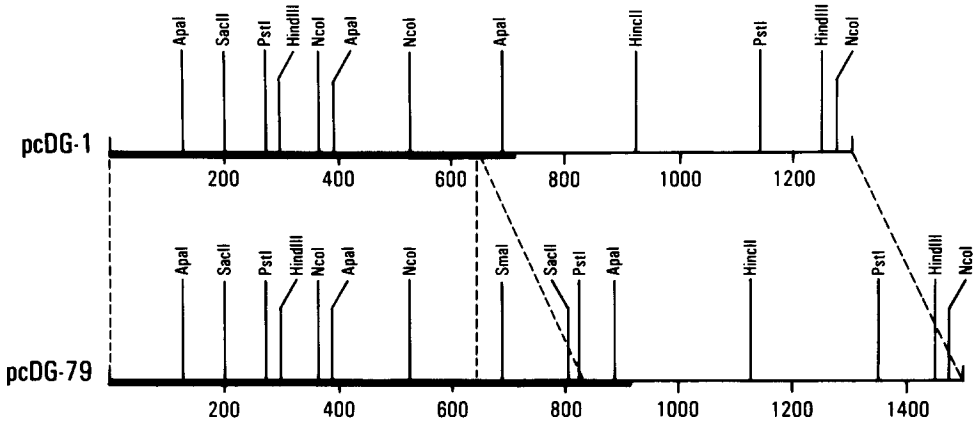


FIGURE 3. S1 nuclease protection experiments carried out on a Bam HI–Nar I human genomic fragment comprising the 5′ end of the invariant chain gene (*lower right*), after hybridization to mRNA from the B cell lines Raji (*1*), LG-2 (*2*), and 721-P (*3*). The ^{32}P end-labeled protected fragment were resolved by electrophoresis in denaturing polyacrylamide gels and visualized by autoradiography. Chemical nucleotide sequencing reactions on the same labeled fragment were run in parallel lanes. The nucleotide sequence, reversed and complemented, is displayed (*top right*). Approximate start points for transcription are underlined, as the most intense band within each of the three clusters of protected fragments. The two in-frame ATG (see text) are also underlined. In the lower right part of this figure, the results are schematized.

potential translation start sites. However, heterogeneity of the 5′ termini in these mRNAs is caused by transcriptional start at three major sites.

To explore the relationship between initiation of transcription at these alternate sites and expression of $\gamma 2/\gamma 3$ or $\gamma 1$ proteins, the relative abundance of these mRNA species was determined by densitometric scans of autoradiograms




```
        cca ccg aaa gTA CTG ACC AAG TGC CAG GAA GAG GTC AGC CAC ATC CCT GCT GTC CAC CCG
p41  190 Pro Pro Lys Val Leu Thr Lys Cys Gln Glu Glu Val Ser His Ile Pro Ala Val His Pro
γ1   190 Pro Pro Lys G

        GGT TCA TTC AGG CCC AAG TGC GAC GAG AAC GGC AAC TAT CTG CCA CTC CAG TGC TAT GGG
p41  210 Gly Ser Phe Arg Pro Lys Cys Asp Glu Asn Gly Asn Tyr Leu Pro Leu Gln Cys Tyr Gly

                                                    *   *   *
        AGC ATC GGC TAC TGC TGG TGT GTC TTC CCC AAC GGC ACG GAG GTC CCC AAC ACC AGA AGC
p41  230 Ser Ile Gly Tyr Cys Trp Cys Val Phe Pro Asn Gly Thr Glu Val Pro Asn Thr Arg Ser

                        *   *   *
        CGC GGG CAC CAT AAC TGC AGT Gag tca ctg gaa
p41  250 Arg Gly His His Asn Cys Ser Glu Ser Leu Glu ----------------------280
γ1   194                          lu Ser Leu Glu ----------------------216
```

FIGURE 4. Restriction endonuclease map of cDNA inserts from plasmids pcDG-1 and pcDG-79. The thicker lines indicate protein-coding regions. The nucleotide sequence of an extra coding region found in the pcDG-79 insert is shown in the bottom part of the figure. The position of this extra region relative to the pcDG-1 insert is marked by the broken lines. The location of the extra amino acid stretch within the p41 protein, and its relative position compared to the γ1 protein are also shown. Asterisks mark potential *N*-linked glycosylation sites.

of the S1 protected fragments, and compared to the relative ratio of γ1 vs. γ2/γ3 protein (Table I). Integration of the peaks corresponding to the three clusters of fragments allowed us to estimate that the −5 and the −10 positions are used ~40 and ~50% of the time, respectively, to initiate transcription, while the −32 position is used at a lower frequency, ~10%. Therefore, no simple correlation was obvious between mRNA ratios (4:6:1) and γ1 vs. γ2/γ3 ratio (5:1) (Table I).

Restriction mapping of our invariant chain cDNA clones revealed that 7 inserts out of 66 contained a common extra Sma I–Pst I restriction fragment, not seen in pcDG-1, a prototype near-full-length γ1 chain cDNA (Fig. 4). Out of these seven cDNAs, the longest one (pcDG-79) was completely sequenced, confirming that it contained an additional 192 bp of coding sequence precisely interjected between the last basepairs of exons 6 and the first nucleotides of exon 7 (Fig. 4). Comparison of the restriction map of clone pcDG-79 with that of a previously characterized (7) phage clone containing the invariant chain chromosomal gene,

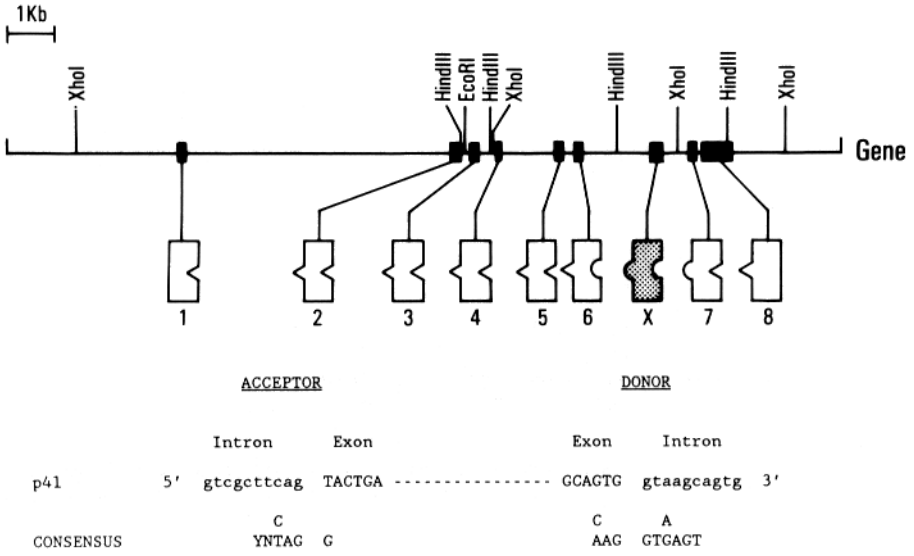


FIGURE 5. Location of exon X in the invariant chain gene, as determined by restriction mapping and nucleotide sequencing. Nucleotide sequences of the 5′ and 3′ border of this exon are shown to indicate their compliance to the splicing acceptor/donor consensus sequences.

allowed us to identify a similar Sma I–Pst I restriction fragment, located as expected in the intron separating exons 6 and 7. This fragment was sequenced (16) together with its flanking regions, revealing it to contain a 192 bp coding sequence identical to the extra fragment of the pcDG-79 cDNA insert. Sequences conforming to the canonical AG/GT splice consensus sequences were found both at the 5′ and the 3′ boundaries of this exon (Fig. 5).

To correlate alternative splicing of this extra exon (exon X) with the generation of an invariant chain–like protein in B cells, the plasmid pcDG-79 was transfected in COS-7 cells. Immunoprecipitation with anti–invariant chain antisera followed by 2-D electrophoresis revealed a protein with electrophoretic mobility identical to the previously observed, Ia-associated p41 component from a B lymphoid cell, Raji (Fig. 6). Two additional spots, whose relative mobility to the p41 was similar to the relative mobility of $\gamma 2/\gamma 3$ to the $\gamma 1$ spot, were also observed in these immunoprecipitations (Fig. 6, vertical arrows). They are probably accounted for by the fact that the pcDG-79 clone is nearly full length, contains both 5′ AUGs, and translation initiates at either of them. These results demonstrate that the Ia-associated p41 protein is derived from the invariant-chain gene by a mechanism of alternative exon splicing, and therefore bears a close structural relationship to the Ia $\gamma 1$ chain.

The extra exon (exon X) of the p41 encodes a stretch of 64 amino acids (shown in Figs. 4 and 7). This sequence includes six cysteines, and does not resemble any other part of the $\gamma 1$ chain. It contains two additional sites, Asn-Gly-Thr and Asn-Cys-Ser, that could be used for NH_2-linked glycosylation (Fig. 4). A computer search of the Genbank database with the program DFASTP (22) identified amino acid homology, with a repeating unit found in bovine and

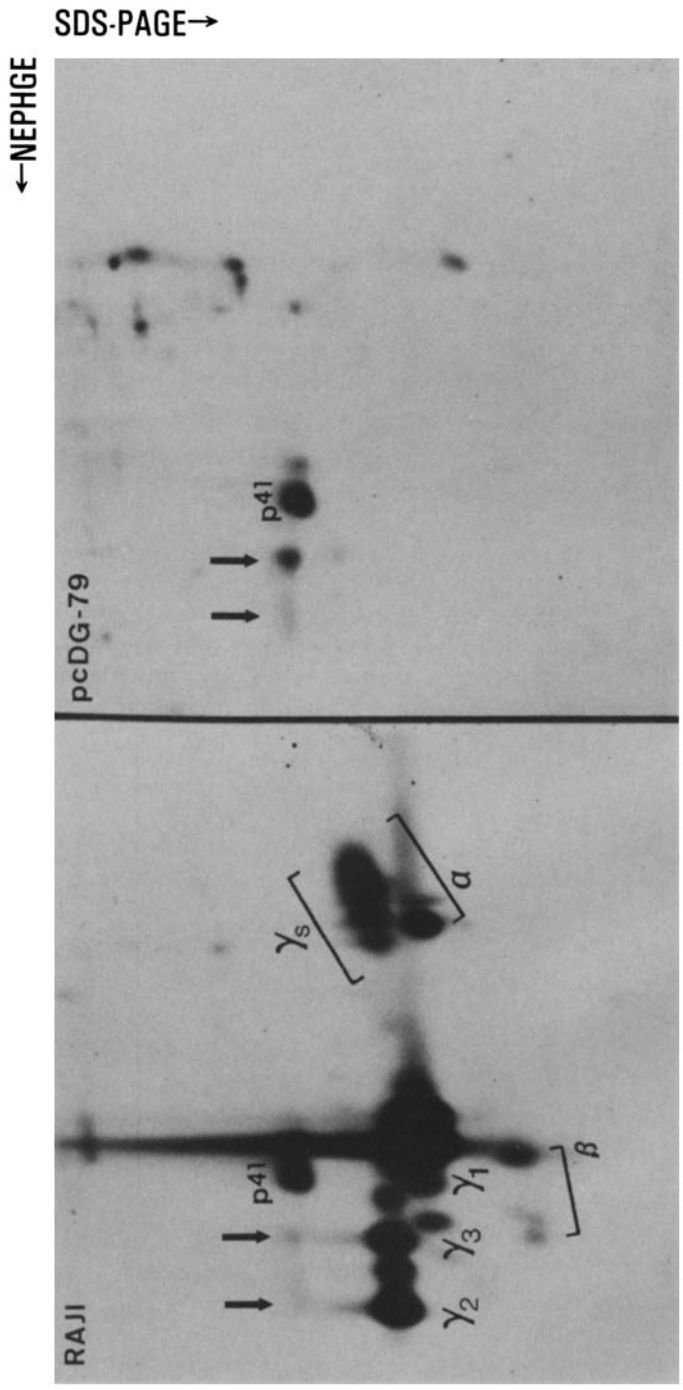


FIGURE 6. 2-D analysis of radiolabeled detergent lysates from either Raji cells or COS-7 cells transfected with plasmid pcDG-79, immunoprecipitated with anti-invariant chain peptide antiserum C351. In addition to the main p41 protein, two p43 proteins are expressed by transfection with pcDG-79, which can also be seen in Raji immunoprecipitates (*vertical arrows*). 2-wk exposures.

```
Human
Thyroglobulin 299 TKC..EVERFTATSFGHPYVPSCRRNGDYQAVQC.QTEGPCWCVDAQGKE 345
                  |||  ||    |   |    | |  || |   ||    | ||||   | |
p41             3 TKCHEEVSHIPAVHPGSFR.PKCDENGNYLPLQCYGSIGYCWCVFPNGTE  51


Bovine
Thyroglobulin 670 SLLGSQPAGSSLFVPACTSKGNFLPVQCFNS..ECYCVDTEGQPIPGTR 716
                  |       ||  | | |   || || ||  |   | ||   |   | ||
p41            10 SHIPAVHPGS..FRPKCDENGNYLPLQCYGSIGYCWCVFPNGTEVPNTR  56


                    S          Y          Y
Consensus       CPTXCE 12-27 FXPQC 3 GXF 2 VQC 4-6 CWCVD 2 GXE 2 GTR 6-8
                  | |          | | |   | |    ||      ||||    | |    ||
p41             VLTKCH  14   FRPKC 3 GNY 2 LQC  6  CWCVF 2 GTE 2 NTR  6
```

FIGURE 7. Homology of the p41 amino acid sequence encoded by exon X with regions of human (37) and bovine (38) thyroglobulin. The homology regions correspond to a sequence imperfectly repeated 10 times in thyroglobulin. The repeats that scored highest in the computer search are shown. In addition, a consensus sequence derived from the repeats of human thyroglobulin precursor is shown at the bottom line. Cysteine residues are in italics. Numbering indicates position of the sequences shown in the respective proteins.

human thyroglobulin (Fig. 7). The statistical significance of this homology was confirmed by random shuffling using the program RDF (22). The calculated Z values of the scores were 16.1 and 11.9, respectively, for human and bovine thyroglobulin ($Z > 3$, possibly significant; $Z > 6$, probably significant; $Z > 10$, significant). Similar results were obtained with a different algorithmn, WORDSEARCH (University of Wisconsin Genetics Computer Group, Madison, WI).

Discussion

To contribute to the understanding of the molecular mechanisms controlling Ia-restricted recognition of antigen, we investigated some structural properties of a nonpolymorphic human Ia subunit, the invariant chain. The experiments presented here show the existence of a complex pattern of transcription and transcript processing in its corresponding gene, resulting in the expression of four invariant chain forms ($\gamma 1$, $\gamma 2/\gamma 3$, p41, and p43), all associated with Ia oligomers and potentially of distinct functional significance. First, transcription of the gene initiates a minimum of three sites. Second, the primary transcripts are alternatively spliced to include or exclude one of the protein-coding exons. Third, two AUGs are alternatively used to initiate translation in the mature mRNAs.

Transcriptional start at more than one site is consistent with the absence in the invariant chain gene of a conventional TATA box (7), which is believed to accurately position the initiation site (23, 24). Other sequences, less efficient than TATA boxes, may act as promoters for multiple start sites in the invariant chain gene, as proposed for Xenopus histone (24) and adenovirus (25) genes. Another formal possibility is that the 5′ terminus of a single transcript is differentially cleaved posttranscriptionally, but this is very unlikely since no enzymes capable of performing this function are known (24, 25). The heterogeneity of 5′ termini in the invariant chain mRNAs is of critical importance to interpret another mechanism observed, alternate initiation of translation (see below). Our results contrast with previous reports of one single site for initiation of transcription in

the same gene (9, 10). The reasons for such discrepancy are unclear, since actual data of S1 protection experiments were not included in those reports.

Besides this heterogeneity of 5′ termini, distinct mRNAs are also generated from the invariant chain gene by a mechanism of alternate splicing of an internal coding exon. This exon (termed exon X) is 192 bp long, and becomes the longest of the coding exons in the invariant chain gene (7). Although its 5′ and 3′ boundaries conform to the AG/GT rule (26) for acceptor and donor splice sites, they are used ~10% of the time as determined by frequency of exon X–containing cDNAs in two libraries we screened. Differential frequency of usage for apparently functional splice sites is a well-known phenomenon in eukaryotes, but the factors governing it are not understood (27). There are also several precedents in eukaryotes for alternate splicing of primary transcripts, giving rise to messages encoding two or more proteins from a single gene. In several of these cases, the alternate splicing is regulated in either a tissue-specific (28, 29) or developmental (29, 30) fashion. It will be of interest to determine whether this is true for the invariant chain gene. The relative abundance of γ1 and p41 approximated a ratio of 10:1 (Table I), but showed some variability that could reflect different turnover rates for these two proteins (V. Quaranta, unpublished observations).

Both types of alternatively spliced invariant chain mRNAs contain two functional start sites for translation, placed in-frame and 48 bp apart, that can be alternatively used to synthesize two proteins differing by a 16-amino-acid NH_2-terminal segment of M_r 33,000 and 31,000, respectively. Exon X–containing mRNAs give rise, instead, to proteins of M_r 43,000 and 41,000. Reactivity with anti–invariant chain antisera and 2-D mobility revealed that these proteins were identical to Ia-associated proteins, namely γ1 (M_r 31,000), γ2/γ3 (M_r 33,000) and p41 (M_r 41,000), isolated from B cells. Spots corresponding to the M_r 43,000 proteins, previously unnoticed in Ia immunoprecipitates, can also be seen in overexposed autoradiograms of Ia immunoprecipitates from B cells (Fig. 6, vertical arrows).

The high γ1:γ2/γ3 ratio in B lymphoid cells (5:1, Table I) may imply downregulation of the first AUG as a start site. Although the factors regulating initiation of translation are not completely understood, mRNA nucleotide sequences near the 5′ terminus are thought to play an important role (31). A comparison of sequence contexts of known translational start sites revealed a consensus sequence (CCG/ACCAUGG) surrounding them (32). The occurrence of this consensus around the second but not the first of the invariant chain AUGs (Fig. 1) may explain the lower usage of the latter. Another possibility is that the different lengths of 5′ untranslated leaders resulting from the heterogeneous transcriptional start sites may affect usage of the first ATG. Untranslated leaders are thought to contain sequences that provide binding sites for ribosomes and initiation factors (33). Because the structural requirements of these binding sites in eukaryotes are not known, it is difficult to make predictions from the sequences of the invariant chain mRNA 5′ termini. In vitro translation of appropriate RNA constructs might be used to distinguish among these possibilities. It should also be noted that differential stability of the two mRNA products, or differences in

the turnover rates of the encoded proteins might also contribute to the observed γ1:γ2/γ3 ratios.

It will be interesting to determine whether, in any of these mRNA species distinguished by their 5′ termini, translation initiates alternatively at either AUG, so that the same message gives rise to two proteins of nonidentical primary structure. Examples of this kind were previously described in certain eukaryotic viruses, which generate as many as three proteins from a single message (34). In spite of several studies, such strategy could never be demonstrated for constitutive cell messages (35), and the invariant chain mRNA could be the first such example. Strubin et al. addressed this issue in a recent study (9), but unfortunately did not recognize the heterogeneity of invariant-chain 5′ termini and did not therefore test critical mRNA constructs. Artificial examples of translational initiation at internal AUGs were recently prepared by inserting multiple ATGs in front of mammalian genes cloned in SV40-based expression vectors (35, 36). Indeed, the plasmid pcDG-1, described in this study, could be classified with this group. However, the fact that, in cells transfected with pcDG-1 the relative usage of the first and second in-frame AUGs is much different than in B cells cautions against uncritically extending data from such cloned cDNAs to constitutive mRNAs. In this context, the human invariant chain mRNA could be an interesting model for regulation of translational initiation at internal AUGs in constitutive cell messages.

Cumulatively, these results established the molecular basis for the presence of structurally distinct, antigenically related invariant chains in Ia oligomers. The structural features that distinguish the four protein forms of invariant chain can be summarized as follows, taking the predominant, M_r 31,000 γ1 chain form as the prototype invariant chain.

The γ2/γ3 forms, translated from the first in-frame AUG codon of exon X–lacking mRNAs, are structurally identical to the γ1 form, except for a 16-amino-acid NH_2-terminal segment, of very basic overall character. Because the invariant chains lack a cleavable signal sequence (15, 37), and because their transmembrane orientation is with NH_2 terminus facing the cytoplasm (15), this results in the γ2/γ3 forms having a longer (46 as opposed to 30 residues) and more basic cytoplasmic tail than the γ1 chain, which may mediate different types of functional interactions. In this regard, we recently determined that the γ1, but not the γ2/γ3 proteins is phosphorylated at their cytoplasmic tail in several cell types (R. Spiro and V. Quaranta, unpublished observations). Similar relationships could be valid for the p43 and p41 forms. Transfections with the plasmid pcDG-1M, which contained a single ATG translational start site, clearly indicated that the γ2 and γ3 proteins are translated with identical amino acid sequences. Their differential 2-D mobility, i.e., two spots with identical relative molecular mass and distinct pI each accompanied by slightly more acidic satellite spots (Fig. 2*B*), strongly suggests that they are distinguished by a posttranslational modification introducing charges.

The p41 form, generated by the splicing-in of exon X, contains 64 amino acids more than the γ1 form, and is otherwise identical to it. This stretch of amino acids, located near the COOH terminus, contains six cysteines, while the remainder of the protein has only one. Such unusual concentration of cysteines

suggests a tight globular structure of this extra domain, and a possible involvement in the formation of disulfide-bridged dimers or multimers. An overall homology at the amino acid level of >40% was found between this segment and a region of similar length that is repeated 10 times in human (38) and bovine thyroglobulin (39). The similarity is made more significant by the virtually absolute conservation of the cysteine spacing patterns, as observed for other protein families (40). Interestingly, the repeat units of thyroglobulin are likely encoded by discrete exons (39). Therefore, the evolutionary mechanism that generated this homology may be one of intergenic exon shuffling (41). The biological significance and possible functional implications of this finding are unknown.

All of the invariant chain forms described here are part of Ia oligomers, since they are found in Ia molecules isolated with anti-β chain antisera. However, it remains to be seen at what stage the various invariant chain forms associate to Ia oligomers, and whether they display preference of association with α or β chains of the DR, DP, or DQ series. The stoichiometric ratios of these interactions also need to be determined. Variations of this kind in the composition of Ia molecules may well have functional implications. If so, it may be possible to understand the strategic significance of encoding four different proteins in a single gene rather than duplicated genes.

Summary

We determined the structural basis for the presence of electrophoretically-distinct, antigenically-related forms of invariant chains in Ia oligomers, and established the mechanisms by which they can be expressed from a single gene. S1 nuclease protection assays indicated that, in B cells, transcription of this gene initiates at a minimum of three sites. Thus, unlike previously thought, invariant chain mRNAs have heterogeneous 5′ untranslated segments that may differentially affect initiation of translation. Further, restriction mapping and nucleotide sequencing of cDNAs revealed two kinds of invariant chain mRNAs differing by an internal coding segment of 192 bp. This segment represents an alternatively spliced exon, as demonstrated by nucleotide sequencing of corresponding genomic regions. The exon (exon X) encodes a cysteine-rich stretch of 64 amino acids near the COOH terminus that displays a striking and surprising homology to an internal amino acid repeat of thyroglobulin, suggesting an evolutionary mechanism of exon shuffling. Transient expression of cDNAs indicated that both types of alternatively spliced mRNAs contain two in-frame AUGs functioning as alternate start sites for translation. Thus, transfections with exon X–lacking cDNAs resulted in the expression of M_r 33,000 and 31,000 proteins, detected by immunoprecipitation with anti–invariant chain antisera, and identical by two-dimensional gel (2-D) analyses to the B cell invariant-chain forms $\gamma 1$ (M_r 31,000), $\gamma 2$, and $\gamma 3$ (M_r 33,000). Similarly, exon X–containing cDNAs expressed M_r 43,000 and 41,000 proteins, also identical by 2-D migration to Ia-associated proteins. Thus, human Ia molecules contain four forms of invariant chain of closely related but nonidentical primary structure that are generated from a single gene by a complex pattern of alternate transcriptional start, exon splicing, and translational start.

We thank Jane Nowell and Anita Holmsen for excellent technical assistance, Alice Bruce for outstanding secretarial help, and Per Peterson, Ron Ogata, and Michael Wilson for invaluable advice.

Received for publication 9 February 1987 and in revised form 16 April 1987.

References

1. Giles, R. C., and J. D. Capra. 1985. Structure, function and genetics of human class II molecules. *Adv. Immunol.* 37:1.
2. Rask, L., K. Gustafsson, D. Larhammar, H. Ronne, and P. A. Peterson. 1985. Generation of class II antigen polymorphism. *Immunol. Rev.* 84:123.
3. Jones, P. P., D. B. Murphy, D. Hewgill, and H. O. McDevitt. 1979. Detection of a common polypeptide chain in I-A and I-E sub-region immunoprecipitates. *Mol. Immunol.* 16:51.
4. Claesson-Welsh, L., P. E. Barker, D. Larhammar, L. Rask, F. H. Ruddle, and P. A. Peterson. 1984. The gene encoding the human class II antigen-associated gamma chain is located on chromosome 5. *Immunogenetics.* 20:89.
5. deKretser, T. A., M. J. Crumpton, J. G. Bodmer, and W. F. Bodmer. 1982. Two-dimensional gel analysis of the polypeptides precipitated by a polymorphic HLA-DR1, w6 monoclonal antibody: evidence for a third locus. *Eur. J. Immunol.* 12:600.
6. Quaranta, V., O. Majdic, G. Stingl, K. Liszka, H. Honigsmann, and W. Knapp. 1984. A human Ia cytoplasmic determinant located on multiple forms of invariant chain (gamma1, gamma2, gamma3). *J. Immunol.* 132:1900.
7. O'Sullivan, D. M., D. Larhammar, M. C. Wilson, P. A. Peterson, and V. Quaranta. 1986. Structure of the human Ia-associated invariant (gamma)-chain gene: Identification of 5′ sequences shared with major histocompatibility complex class II genes. *Proc. Natl. Acad. Sci. USA.* 83:4484.
8. Kudo, J., L. Y. Chao, F. Narni, and G. F. Saunders. 1985. Structure of the human gene encoding the invariant gamma-chain of class II histocompatibility antigens *Nucleic Acids Res.* 13:8827.
9. Strubin, M., E. O. Long, and B. Mach. 1986. Two forms of the Ia antigen-associated invariant chain result from alternative initiation at two in-phase AUGs. *Cell.* 47:619.
10. Strubin, M., B. Mach, and E. O. Long. 1984. The complete sequence of the mRNA for the HLA-DR-associated invariant chain reveals a polypeptide with an unusual transmembrane polarity. *EMBO. (Eur. Mol. Biol. Organ.) J.* 3:869.
11. Gluzman, Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell.* 23:175.
12. Giacoletto, K. S., A. J. Sant, C. Bono, J. Gorka, D. M. O'Sullivan, V. Quaranta, and B. D. Schwartz. 1986. The human invariant chain is the core protein of the human class II-associated proteoglycan. *J. Exp. Med.* 164:1422.
13. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
14. Okayama, H., and P. Berg. 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol. Cell. Biol.* 3:280.
15. Claesson, L., D. Larhammar, L. Rask, and P. A. Peterson. 1983. cDNA clone for the human invariant (gamma) chain of Class II histocompatibility antigens and its implications for the protein structure. *Proc. Natl. Acad. Sci. USA.* 80:7395.
16. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499.
17. Vlasuk, G. P., and S. Inouye. 1983. Experimental manipulation of gene expression. M. Inouye, editor. Academic Press, New York. 291.

18. Graham, F. L., and A. J. Van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology.* 52:456.
19. O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell.* 12:1133.
20. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 277:680.
21. Bonner, W. M., and R. A. Laskey. 1974. A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. *Eur. J. Biochem.* 46:83.
22. Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC).* 227:1435.
23. Wasylijk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, A. R. Teoule, and P. Chambon. 1980. Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. *Proc. Natl. Acad. Sci. USA.* 77:7024.
24. Grosschedl, R., and M. L. Birnstiel. 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. *Proc. Natl. Acad. Sci. USA.* 77:1432.
25. Baker, C. C., J. Herisse, G. Courtois, F. Galibert, and E. Ziff. 1979. Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5′ end. *Cell.* 18:569.
26. Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349.
27. Mount, S., and J. Steitz. 1983. Lessons from mutant globins. *Nature (Lond.).* 303:380.
28. Nawa, H., H. Kotani, and S. Nakanishi. 1984. Tissue-specific generation of two preprotachykinin mRNAs from one gene by alternative RNA splicing. *Nature (Lond.)* 312:729.
29. Breitbart, R. E., H. T. Nguyen, R. M. Medford, A. T. Destree, V. Mahdavi, and B. Nadal-Ginard. 1985. Intricate combinatorial patterns of exon splicing generate multiple regulated troponin T isoforms from a single gene. *Cell.* 41:67.
30. Falkenthal, S., V. P. Parker, and N. Davidson. 1985. Developmental variations in the splicing patterns of transcripts from the Drosophila gene encoding myosin alkali light chain result in different carboxy-terminal amino acid sequences. *Proc. Natl. Acad. Sci. USA.* 82:449.
31. Kozak, M. 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature (Lond.)* 308:241.
32. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857.
33. Moldave, K. 1985. Eukaryotic protein synthesis. *Annu. Rev. Biochem.* 54:1109.
34. Giorgi, C., B. M. Blumberg, and D. Kolakofsky. 1983. Sendai virus contains overlapping genes expressed from a single mRNA. *Cell.* 35:829.
35. Peabody, D. S., S. Subramani, and P. Berg. 1986. Effect of upstream reading frames on translation efficiency in simian virus 40 recombinants. *Mol. Cell. Biol.* 6:2704.
36. Peabody, D. S., and P. Berg. 1986. Termination-reinitiation occurs in the translation of mammalian cell mRNAs. *Mol. Cell. Biol.* 6:2695.
37. Kvist, S., K. Wiman, L. Claesson, P. A. Peterson, and B. Dobberstein. 1982. Membrane insertion and oligomeric assembly of HLA-DR histocompatibility antigens. *Cell.* 29:61.
38. Malthiery, Y., and S. Lissitzky. 1985. Sequence of the 5′-end quarter of the human-thyroglobulin messenger ribonucleic acid and of its deduced amino-acid sequence. *Eur. J. Biochem.* 147:53.
39. Mercken, L., M.-J. Simons, S. Swillens, M. Massaer, and G. Vassart. 1985. Primary

structure of bovine thyroglobulin deduced from the sequence of its 8,431-base complementary DNA. *Nature (Lond.).* 316:647.
40. Doolittle, R. F. 1985. The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* 10:233.
41. Gilbert, W. 1978. Why genes in pieces? *Nature (Lond.)* 271:501.