APPLICATION OF SYNTHETIC OLIGONUCLEOTIDES IN THE ISOLATION OF MURINE TRANSPLANTATION ANTIGEN cDNA CLONES

Thesis by

Antonio A. Reyes

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology
Pasadena, California

1982

(Submitted April 2, 1982)

Acknowledgments

The completion of this work would not have been possible without the help of professors, friends and loved ones. I am deeply indebted to:

Drs. R. Bruce Wallace and Herschel K. Mitchell for comments, advice and encouragement;

Merrie Jo Johnson and Monica Schöld for assistance in some of the experiments;

Caltech for funding;

City of Hope for the use of its facilities;

my family for support and understanding;

and Cynthia, most especially, for seeing me through.

Abstract

Cytoplasmic poly(A)$^+$RNA was extracted from the murine thymoma cell line EL4 (b haplotype) and used to construct a library of cloned cDNA in pBR322. Initially, 30,000 colonies were screened with a mixture of eight hexadecanucleotides representing all possible coding sequences for residues 51-56 of H-2K^b. The only clone isolated, pH2K01, contains the coding sequence for residues 50 through 91 of H-2K^b, followed by a Glu codon and a termination codon. It is speculated that the mRNA from which pH2K01 was derived and H-2K^b mRNA have a common precursor but differ in the manner of post-transcriptional splicing.

A 133-nucleotide probe containing most of the coding sequence of pH2K01 was constructed and used to screen the remainder of the library. Two clones were isolated. pH202 encodes H-2K^b from residue 66 through the carboxy-terminus and includes 386 nucleotides of 3'-untranslated sequence. pH203 codes for H-2D^b, from residue 82 through the carboxy-terminus, together with 476 nucleotides of 3'-untranslated sequence.

H-2K^b and H-2D^b share sequence homologies of 83% and 91% at the protein and nucleotide levels, respectively. The cytoplasmic region of the molecule proximal to the membrane is identical in both antigens. The next most conserved region is the third external domain.

The H-2K^b molecule is 10 amino acid residues longer than the H-2D^b molecule. Analysis of 3'-end coding sequences of pH202, pH203 and other H-2 clones reported in the literature suggests that the difference in length could be the result of different splicing patterns of the mRNAs.

Table of Contents

Chapter I

Introduction

The major histocompatibility complex (MHC) is a cluster of genes which mediates a variety of immune functions (Klein, 1975; Klein et al., 1981; Ploegh et al., 1981). The most thoroughly studied MHCs are those of the mouse (H-2) and man (HLA). The experiments reported in the following chapters concern the molecular biology of H-2. This introduction will give a summary of H-2 structure and function and also discuss briefly the methodology used to isolate the H-2 cDNA clone described in Chapter II.

Traditionally, H-2 has been defined to consist of the K, I, S and D regions, which span about 0.5 recombination units in chromosome 17. This collection of genes codes for at least three classes of molecules (Klein, 1975). Class I molecules, the major transplantation antigens, are encoded by K and D region genes. These molecules are involved in allograft rejection and T cell cytotoxicity restriction (Zinkernagel and Doherty, 1974). Class I molecules are transmembrane glycoproteins of about 45 K molecular weight. They are non-covalently associated with β 2-microglobulin and expressed on almost all cells and tissues. The transplantation antigen molecule can be divided into extracellular, transmembrane and cytoplasmic domains (Coligan et al., 1981; Nathenson et al., 1981). The extracellular domain can be further subdivided into three regions: the N-terminal and two Cys-containing portions.

Recent studies on the Tla and Qa antigens have shown that they are similar to Class I molecules in molecular weight, glycosylation pattern and the capacity to bind β 2-microglobulin (Stanton and Hood, 1980; Soloski et al., 1981; Vitteta et al., 1981; Michaelson et al., 1977). Hence the Tla region, which is outside the traditional boundaries of H-2, may actually be part of the murine MHC.

The I region codes for Class II molecules. These antigens, called Ia, consist of two chains, α (33K) and β (29K). Ia molecules are involved in the restriction of T_H cell specificity (Erb and Feldman, 1975; Sprent, 1978) and control of immune response and immune supression (Murphy et al., 1976). Class II molecules have a more limited distribution, being present mainly on B cells and macrophages. The

S region encodes Class III molecules which are components of the complement system (Shreffler, 1976).

It has been proposed that the true function of H-2-encoded molecules is to act as recognition markers in cell-cell interactions (Klein et al., 1981). The physiologically meaningful traits controlled by H-2, namely restriction of cytolytic and helper T-cell specificities and control of immune response and immune suppression, could all involve the recognition by T cells of antigens in the context of H-2 molecules (Klein et al., 1981). The mechanism for dual recognition of H-2 molecule and antigen by T-cells has to date not been elucidated.

In addition to the question of function, there is the related interest in the evolution of the major histocompatibility complex. Class I molecules are highly polymorphic, with differences generally scattered throughout the molecule. They do share gross structural similarities and considerable amino acid sequence homologies (Nathenson et al., 1981). While the K locus codes for only the K molecule, the D region codes for at least two molecules, D and L. Recent studies indicate that the Qa and Tla antigens are also Class I-like (see above). For these reasons, it has been speculated that Class I antigen genes arose by duplication of a primordial gene (Silver and Hood, 1976; Bodmer, 1973; Shreffler et al., 1971).

The experiments reported here concern the isolation of cDNA clones for the major transplantation antigens encoded in the K and D regions. The synthetic oligonucleotide hybridization approach was used to obtain an H-2K^b-like cDNA clone (Chapter II). Briefly, a region of the protein whose amino acid sequence is known and for which there is minimum codon degeneracy is chosen. A set of oligonucleotides containing all possible coding sequences (or their complementary sequences) for this region is synthesized. Synthesis is done in the solid phase and employs the triester approach (Miyoshi and Itakura, 1979). All sequences in a set are synthesized simultaneously (Wallace et al., 1981; Itakura et al., 1981). Previously it was shown in model studies that under stringent hybridization conditions a single base pair mismatch

is enough to prevent the formation of a stable duplex between the oligonucleotide probe and complementary DNA (Wallace et al., 1981). T_d, the temperature at which half of a perfectly matched oligonucleotide would dissociate from a duplex immobilized on nitrocellulose, was empirically defined as : T_d, °C = (number of AT base pairs X 2) + (number of GC base pairs X 4), in 1 M Na^+. At 2-5°C below T_d, mismatched duplexes are not formed. It was also shown that in mixed probe experiments, the presence of excess imperfectly matched oligonucleotides in the hybridization solution does not prevent the right sequence from hybridizing to target DNA (Wallace et al., 1981). These results indicated the potential use of this methodology for the isolation of specific cloned sequences. The mixed probe approach has been used successfully to isolate a human β 2-microglobulin cDNA clone (Suggs et al., 1981) and the H-$2K^b$-like clone pH2K01 described in Chapter II.

In Chapters III and IV the characterization of the cDNA clones for H-$2D^b$ and H-$2K^b$ are described. Comparisons between the two molecules are made at the amino acid and nucleotide levels. Analysis of the DNA sequences of these and other H-2 clones suggests that alternative splicing patterns in the 3'-end of the precursor RNA could be one mechanism to generate diversity in transplantation antigen molecules.

References

Bodmer, W. F.: A new genetic model for allelism at histocompatibility and other complex loci: Polymorphism for control of gene expression. Transplant Proc. 4: 1471-1475, 1973.

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J. and Nathenson, S. G.: Primary structure of a murine transplantation antigen. Nature 291: 35-39, 1981.

Erb, P. and Feldman, M.: The role of macrophages in the generation of T-helper cells. J. Exp. Med. 142: 460-472, 1975.

Itakura, K., Miyake, T., Kawashima, E. H., Ike, Y., Ito, H., Morin, C., Reyes, A. A., Johnson, M. J., Schold, M. and Wallace, R. B.: Chemical synthesis and application of oligonucleotides of mixed sequence. In: A. G. Walton (ed.): Case Western Third Cleveland Symposium, Elsevier-North Holland, Amsterdam, pp. 273-289, 1981.

Klein, J.: In: Biology of the Mouse Histocompatibility-2 Complex, Springer-Verlag, New York, 1975.

Klein, J., Juretic, A., Baxevanis, C. N. and Nagy, Z. A.: The traditional and a new version of the mouse H-2 complex. Nature 291: 455-460, 1981.

Michaelson, J., Flaherty, L., Vitetta, E. and Ponlik, M. D.: Molecular similarities between the Qa-2 alloantigen and other gene products of the 17th chromosome of the mouse. J. Exp. Med. 145: 1066-1070, 1977.

Miyoshi, K. and Itakura, K.: Solid phase synthesis of nonadecathymidylic acid by the phosphotriester approach. Tetrahedron Lett. 38: 3635-3638, 1979.

Murphy, D. B., Herzenberg, L. A., Okumura, K., Herzenberg, L. A. and McDevitt, H. O.: A new I subregion (I-J) marked by a locus (Ia-4) controlling surface determinants on suppressor T lymphocytes. J. Exp. Med. 144: 699-712, 1976.

Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J. and Coligan, J. E.: Primary structural analysis of the transplantation antigens of the murine H-2 major histocompatibility complex. Ann. Rev. Biochem. **50**: 1025-1052, 1981.

Ploegh, H. L., Orr, H. T. and Strominger, J. L.: Major histocompatibility antigens: The human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. Cell 24: 287-299, 1981.

Shreffler, D. C., David, C. S., Passmore, H. C. and Klein, J.: Genetic organization and evolution of the mouse H-2 region: A duplication model. Transplant. Proc. 3: 176-179, 1971.

Shreffler, D. C.: The S region of the mouse major histocompatibility complex (H-2): Genetic variation and functional role in complement system. Transplant. Rev. 32: 140-167, 1976.

Silver, J. and Hood, L.: Preliminary amino acid sequence of transplantation antigens: Genetic and evolutionary implications. Contemp. Top. Mol. Immunol. 5: 35-86, 1976.

Soloski, M. J., Uhr, J. W., Flaherty, L. and Vitetta, E. S.: Qa-2, H-2K and H-2D alloantigens evolved from a common ancestral gene. J. Exp. Med. 153: 1080-1093, 1981.

Sprent, J.: Restricted helper function of F_1 hybrid T cells positively selected to heterologous erythrocytes in irradiated parental strain mice. J. Exp. Med. 147: 1159-1174, 1978.

Stanton, T. H. and Hood, L.: Biochemical identification of the Qa-1 alloantigen. Immunogenetics 11: 309-314, 1980.

Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. and Itakura, K.: Use of synthetic oligonucleotides as hybridization probes. III. Isolation of cloned cDNA sequences for human β 2-microglobulin. Proc. Natl. Acad. Sci. USA 78: 6613-6617, 1981.

Vitetta, E. S., Poulik, M. D., Lein, J. and Uhr, J. W.: Beta 2-microglobulin is selectively associated with H-2 and TL alloantigens on murine lymphoid cells. J. Exp. Med. 144: 179-192, 1976.

Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T. and Itakura, K.: Hybridization of synthetic oligonucleotides to ØX174 DNA: The effect of single base pair mismatch. Nucleic Acids Res. 6: 3543-3557, 1979.

Wallace, R. B., Johnson, M. J., Hirose, J., Miyake, J., Kawashima, E. H. and Itakura, K.: The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequences to rabbit β-globin DNA. Nucleic Acids Res. 9: 879-894, 1981.

Zinkernagel, R. M. and Doherty, P. C.: Immunological surveillance against altered self components by sensitized T lymphocytes in lymphocytic choriomeningitis. Nature 251: 547-548, 1974.

Chapter II

Immunogenetics 14: 383–392, 1981

Identification of an H-2K^b-Related Molecule by Molecular Cloning

Antonio A. Reyes*, Merrie Jo Johnson, Monica Schöld, Hirataka Ito, Yoshimasa Ike, Christophe Morin†, Keiichi Itakura, and R. Bruce Wallace‡

Department of Molecular Genetics, Division of Biology, City of Hope Research Institute, 1450 East Duarte Road, Duarte, California 91010

Abstract. Based on the published amino-acid sequence of H-2K^b, we synthesized a mixture of eight 16-base long oligodeoxyribonucleotides representing all possible coding sequences for residues 51-56 (Trp-Met-Glu-Gln-Glu-Gly). The hexadecanucleotide mixture was used as a probe to screen recombinant DNA clones constructed from cytoplasmic PolyA$^+$ RNA isolated from the murine thymoma cell line EL4 (*b* haplotype). Of the 30 000 independent clones screened, one clone was found to hybridize with the probe. DNA sequence analysis showed that the cDNA clone was derived from a portion of an H-2K^b-related mRNA. The clone encodes a protein sequence identical with a region of H-2K^b in 42 consecutive residues (50 through 91). The sequence then diverges from the H-2K^b sequence and, after a single Glu codon, a termination codon is encountered. It is possible that this mRNA codes for a small 92 amino-acid protein with a sequence identical (except for a carboxy-terminal Glu residue) with the amino terminus of H-2K^b. It is further speculated that this mRNA is coded for by the *H-2K^b* gene and differs from the H-2K^b mRNA in the pattern of posttranscriptional splicing.

Introduction

The major histocompatibility complex (MHC) of the mouse (*H-2*) is a tightly linked collection of genes on chromosome 17 encoding several cell-surface glycoproteins (Klein 1975, 1979, Klein et al. 1981). One major class of glycoproteins encoded by the *H-2K*, *D*, and *L* loci are the classical transplantation antigens. These antigens,

* *Present address:* California Institute of Technology, Pasadena, California.
† *Present address:* Laboratoire de Chimie Organique, Institut Pasteur, Paris, France.
‡ To whom offprint requests should be sent.

aside from playing a major role in allograft rejection, are involved as restrictions elements in antigen recognition by cytotoxic T lymphocytes (Zinkernagel and Doherty 1974, 1979).

One of the most remarkable features of the transplantation antigen genes is their high degree of polymorphism. Biochemical studies on several H-2K and H-2D antigens have demonstrated that allelic antigens differ from each other by several amino-acid substitutions (Coligan et al. 1981). The K and D antigens show extensive homology. Indeed, allelic antigens appear to be no more homologous to each other than are antigens coded for by different loci. The polymorphic differences between different antigens appear to be distributed throughout the molecule (Ploegh et al. 1981).

The complete amino-acid sequence of H-2K^b has recently been determined (Coligan et al. 1981, Uehara et al. 1981). The protein is 346 amino acids in length, contains two carbohydrate moieties, and when inserted in the cell membrane is found to be noncovalently associated with a 99 amino-acid long protein β2 microglobulin (β2m). The transplantation antigen molecule can be divided into three regions, an extracellular region, a transmembrane region, and a cytoplasmic region. In addition, the extracellular region seems to be composed of three domain-like subregions of approximately 90 amino acids each (Ploegh et al. 1981, Strominger 1980). At least two of these domain-like subregions show striking homology to immunoglobulin constant region domains (Steinmetz et al. 1981).

Despite an immense amount of investigation, the genetic organization of the murine MHC is not well defined. The exact number of expressed genes is uncertain (Silver and Hood 1976, Bodmer 1973). In order to begin to understand the organization of the genes of the *H*-2 region at the molecular level, we have devised a method of isolating particular cloned DNAs coding for H-2-gene products using synthetic DNA probes (Wallace et al. 1981a). In this method, a mixture of oligodeoxyribonucleotides representing all possible coding sequences for a particular short amino-acid sequence is chemically synthesized. The mixture is then radiolabeled and used as a hybridization probe to screen recombinant clones for the desired DNA sequence. Using this method, we have recently reported the isolation of a cloned cDNA for human β2m (Suggs et al. 1981).

In this paper we report the isolation of a cDNA clone derived from a portion of an H-2K^b-related mRNA. The cloned cDNA codes for a protein whose amino-terminal amino-acid sequence is identical with a region of H-2K^b from residue 50 through 91 at which point the sequence diverges from that of H-2K^b, resulting in translation termination after the addition of a single Glu residue.

Materials and Methods

Oligonucleotide synthesis. All of the oligonucleotides (Fig. 1) were synthesized on a solid support by the triester approach (Miyoshi and Itakura 1979). The eight different sequences in H2K16Mix (Fig. 1) were synthesized simultaneously by the method outlined previously (Wallace et al. 1981a, Itakura et al. 1981).

Growth of cells. The EL4 cell line (obtained from Dr. M. Cohn, Salk Institute, California) was grown in RPMI 1640 medium (Gibco) supplemented with 10% fetal calf serum (Gibco), glutamine (300 μg/ml),

H-2K^b-Like cDNA Clone

PROBE 1

		51	52	53	54	55	56	
Amino-Acid Sequence		Trp	Met	Glu	Gln	Glu	Gly	
mRNA Sequence	5′	UGG	AUG	GAG A	CAG A	GAG A	GGX	3′
H2K16Mix Probe Sequence	3′	CC	TAC	CTC T	GTC T	CTC T	CC	5′

PROBE 2

		58	59	60	61	
Amino-Acid Sequence		Glu	Tyr	Trp	Glu	
mRNA Sequences	5′	GAG A	UAC U	UGG	GAG A	3′
H2K11-I Probe Sequence	3′	CTT	ATA	ACC	CT	5′
H2K11-II Probe Sequence	3′	CTT	ATG	ACC	CT	5′
H2K11-III Probe Sequence	3′	CTC	ATA	ACC	CT	5′
H2K11-IV Probe Sequence	3′	CTC	ATG	ACC	CT	5′

Fig. 1. Sequences of oligonucleotide probes.

penicillin (100 units/ml), and streptomycin (100 μg/ml). Cells were harvested at a density of approximately 5×10^6 cells/ml.

RNA isolation. Cells (300 ml) were centrifuged at 1000 × g for 5 min, washed once with isotonic saline at 0 C, and resuspended in 0.01 M NaCl, 0.01 M tris-HCL, pH 7.4, 0.005 M $MgCl_2$ at 5×10^7 cells/ml, and brought to 0.5% NP40 (Shell Oil). After 5 min at 0 C, nuclei were removed by centrifugation at 1000 × g for 5 min. Cytoplasmic RNA was extracted and PolyA$^+$ RNA prepared as described by Singer and Penman (1973).

DNA isolation. Plasmid DNA was isolated from chloramphenicol amplified cultures by scaling up the procedure of Birnboim and Doly (1979). The DNA was further purified by isolating the covalently closed circular plasmid DNA after centrifugation in a CsCl-ethidium bromide density gradient (Clewell and Helinski 1969).

Construction of cDNA clones. Double-stranded cDNA was prepared from 20 μg of PolyA$^+$ RNA by the procedure of Goeddel and co-workers (1980). The cDNA was size selected by isolation of the excluded fraction after chromatography of the cDNA on a Sepharose CL-4B column (Sigma). The average size of the double-stranded cDNA was 700 base pairs (bp). Approximately 20 dCMP residues were added to the double-stranded cDNA with terminal transferase (gift of Dr. K. Fong) (Roychourdhury and Wu 1980). Similarly, approximately five to six dGMP residues were added to *Pst* I linearized pBR322 (Chang et al. 1978). After phenol-chlorophorm extraction, the C-tailed cDNA and the G-tailed pBR322 were combined in a mass ratio of 1:45, at 5.5 μg/ml in 0.3 M NaCl, 0.03 tris-HCl, pH 7.5, 0.003 M EDTA, heated to 65 C, and allowed to cool slowly to room temperature overnight. The plasmid/cDNA mixture was used to transform the *E. coli* K12 strain MC1061 (gift of Dr. J. Rossi) (Casadabhan and Cohen 1980) by the procedure of Kushner (1978). The transformed cells were plated onto L-broth agar plates containing 20 μg/ml tetracycline, such that there were 1000–2000 colonies per 100 mm petri dish. From 200 ng of cDNA, 120000 clones were obtained.

Screening cDNA clones. After overnight growth at 37 °C, 30000 unordered colonies from the transformation described above were lifted onto Whatman 540 filter paper, amplified on chloramphenicol-containing agar plates, and the filters prepared for hybridization as described by Gergen and co-workers (1979).

H2K16Mix was labeled by transfer of ^{32}P from $\gamma^{32}P$-ATP using T4 polynucleotide kinase (New England Nuclear) as described previously (Wallace et al. 1979), $\gamma^{32}P$-ATP was synthesized by the method of Walseth and Johnson (1979).

The colony filters were prehybridized in 6 × NET (NET = 0.15 M NaCl, 0.015 M tris-HCl, pH 7.5, 0.001 M EDTA), 0.5% NP40, 100 μg/ml sonicated, single-stranded *E. coli* DNA at 65 °C for 2 h. They were then hybridized with ^{32}P labeled H2K16Mix in 6 × NET, 0.5% NP40, 250 μg/ml yeast tRNA (Boehringer, Mannheim F.R.G.) at 2 ng/ml probe overnight at 46 °C. The filters were then washed at 0 °C for 20 min with four changes of 6 × SSC (SSC = 0.15 M NaCl, 0.015 M sodium citrate pH 7.2) and then at 37 °C and 46 °C for 5 min each in 6 × SSC, and exposed to Kodak XR-5 film between two intensifier screens (DuPont Lightning Plus) at −20 °C for 2–4 h.

Colonies from areas of the petri dishes showing positive hybridization to the probe were rescreened in a similar manner. Only one area yielded colonies which were positive on rescreening. The plasmid from this colony is designated pH2K01.

DNA sequence analysis. Restriction fragments generated by *Pst* I cleavage were labeled at their 3′ termini with α-^{32}P cordycepin triphosphate (New England Nuclear) as described (Tu and Cohen 1980). After digestion by *Eco*RI endonuclease, the restriction fragment containing the insert was separated by electrophoresis on a 5% acrylamide gel and isolated by electroelution. Fragments generated by *Hap* II digestion of the *Alu* I fragment containing the insert (see Fig. 3) were labeled at their 3′ termini by incorporating α-^{32}P dCTP (New England Nuclear) with the large fragment of DNA polymerase I (Boehringer-Mannheim) in the presence of the other three deoxynucleoside triphosphates unlabeled. End-labeled fragments were separated on a 5% acrylamide gel and isolated by electroelution. Under these conditions the *Alu* I ends do not become labeled. All end-labeled fragments were sequenced by the method of Maxam and Gilbert (1980).

Additional sequencing was performed by the chain termination method of Sanger and co-workers (1977) as described elsewhere (Wallace et al. 1981b) using the oligonucleotide TGGATGGA as a primer and the *Pst*I-*Eco*RI fragment containing the insert (Fig. 3) as the template.

Southern blot analysis. Plasmid DNA (0.25 μg) was digested in 5 μl with *Hin*fI endonuclease (New England Biolabs, 2.3 units) in 7 mM tris-HCl, pH 7.5, 7 mM NaCl, 7 mM $MgCl_2$, 7 mM DTT for 60 min at 37 °C. One volume of 20% glycerol, 18 mM tris-HCl, 2 mM boric acid, 0.5 mM EDTA, 0.2% bromphenol blue was added and the digested DNA subjected to electrophoresis in a 1.2% agarose gel, denatured in situ, and transferred to nitrocellulose by the method of Southern (1975) as described previously (Wallace et al. 1979). The nitrocellulose filters were baked in vacuo for 2 h at 65 °C and hybridized with a ^{32}P labeled probe. In the case of the H2K16Mix probe, hybridization was done in 6 × NET, 10 × Denhardt's solution, 0.5% NP40 with 12 ng/ml probe at 46 °C overnight. In the case of H2K11 I-IV, hybridization was done in the same solution with 3 ng/ml probe at 12 °C overnight. The filters were washed at 0 °C for 15 min with three changes of 6 × SSC and exposed to X-ray film as described above for colony screening.

Results

Isolation of cDNA clone using H-2K^b probe. We synthesized two sets of oligonucleotide probes for the identification of H-2K^b cDNA clones based on the published amino-acid sequence of the antigen (Coligan et al. 1981, Uehara et al. 1981). The first set (H2K16Mix) was a mixture of eight 16-base long oligonucleotides representing all possible coding sequences for amino acids 51–56 (Fig. 1). The second set of probes (H2K11 I-IV) were synthesized as four different 11-base long oligonucleotides representing all possible coding sequences for amino acids 58–61 (Fig. 1).

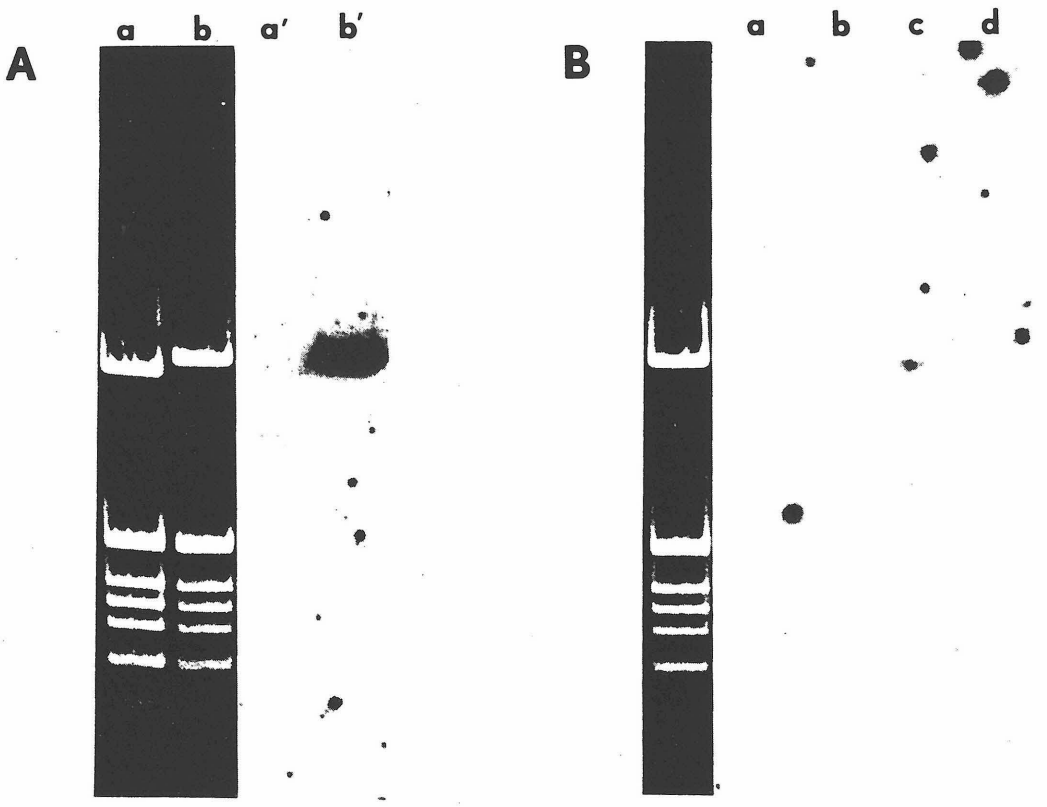

Fig. 2A and B. Southern blot analysis of pH2K01. pBR322 and pH2K01 DNA were digested with *Hin*fI endonuclease, electrophoresed on a 1.2% agarose gel, transferred to nitrocellulose filters and hybridized with various probes. (**A**) Lanes a and b show the ethidium-bromide stained gel, a′ and b′, an autoradiograph of the same lanes after blotting and hybridization with H2K16Mix, a,a′, pBR322, and b,b′, pH2K01. (**B**) The stained-gel pattern of pH2K01 is shown with an autoradiograph of the filters after blotting and hybridization with the H2K11 probes. Lane a, H2K11-I; b, H2K11-II; c, H2K11-III; and d, H2K11-IV.

Using the colony screening approach described in the *Methods* section, one clone, pH2K01, was found to hybridize to the H2K16Mix probe. Plasmid DNA was isolated from this clone, digested with *Hin*fI endonuclease, electrophoresed in an agarose gel and the restriction fragments transferred to nitrocellulose by the method of Southern (1975). These blots were then hybridized with ^{32}P labeled H2K16Mix, H2K11 I, II, III, and IV. The autoradiograph of the hybridized filters is shown in Figure 2. The H2K16Mix probe hybridizes to the largest restriction fragment of the pH2K01 DNA, which contains the cDNA cloned at the *Pst*I site (Sutcliffe 1978), while the probe did not hybridize to any pBR322 restriction fragments (Fig. 2A). Of the undecamers only H2K11-III was found to hybridize (Fig. 2B), indicating, first, that the clone contains sites complementary to both of the H-2K^b probes and, second, that the codon for Glu^{58} should be GAG and for Tyr^{59} UAU.

Characterization of pH2K01. Plasmid DNA from the clone was characterized by digestion with a number of restriction endonucleases. A number of sites were found

to be absent from the cloned DNA, including *Hin*fI, *Eco*RI, *Hin*dIII, *Cla*I, *Bam*HI, *Sal*I, and *Alu*I. Although the DNA was inserted at the *Pst*I site of pBR322 by the G-C tailing approach (Chang et al. 1978), only one *Pst*I site was found to be present. The inserted DNA was sequenced by the base-specific cleavage method of Maxam and Gilbert (1980) and by the chain termination method of Sanger and co-workers (1977) as described in the *Methods* section and in Figure 3B. The DNA sequence and the deduced amino-acid sequence are shown in Figure 3A. The insert is 163 bp in length. The deduced protein sequence from this clone is identical to H-2K^b from residue 50 through 91, at which point the clone sequence diverges from that of H-2K^b having a Glu residue rather than a Ser residue in the next position, followed by an in-frame termination codon (UGA).

A

```
 50   51   52   53   54   55   56   57   58   59   60   61   62   63
Arg  Trp  Met  Glu  Gln  Glu  Gly  Pro  Glu  Tyr  Trp  Glu  Arg  Glu
C G G T G G A T G G A G C A G G A G G G G C C C G A G T A T T G G G A G C G G G A G

 64   65   66   67   68   69   70   71   72   73   74   75   76   77
Thr  Gln  Lys  Ala  Lys  Gly  Asn  Glu  Gln  Ser  Phe  Arg  Val  Asp
A C A C A G A A A G C C A A G G G C A A T G A G C A G A G T T T C C G A G T G G A C

 78   79   80   81   82   83   84   85   86   87   88   89   90   91
Leu  Arg  Thr  Leu  Leu  Gly  Tyr  Tyr  Asn  Gln  Ser  Lys  Gly  Gly
C T G A G G A C C C T G C T C G G C T A C T A C A A C C A G A G C A A G G G C G G T

 92
Glu  Stop
G A G T G A C C C C G G G T C G G A G G T C A C G A C C C C T C C A C G T
```

B

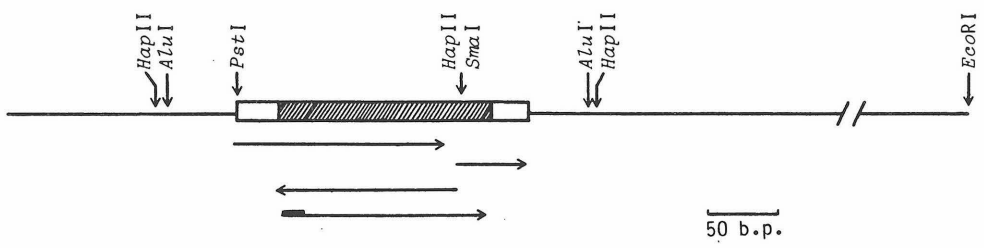

Fig. 3A and B. pH2K01 (**A**) Nucleotide sequence and deduced amino-acid sequence of pH2K01. (**B**) Partial restriction map of pH2K01. The *line* represents pBR322 sequences, *open boxes* are G-C tails, and *the hatched box* represents cDNA insert. The *arrows* indicate extent and direction of DNA sequencing by the Maxam and Gilbert (upper three *arrows*) and Sanger (lowest *arrow*) methods. The *bar* at the end of the lowest *arrow* indicates the location of the primer used in sequencing.

Discussion

The complete homology between the amino-acid sequence deduced from the cDNA clone and the published sequence of H2K^b for 42 consecutive residues argues strongly that we have isolated a cDNA clone coded for by the *H-2K* locus. Why then does the coding sequence diverge so dramatically after codon 91? While we have no direct evidence at present, it is suggested that the mRNA from which the pH2K01 clone was produced differs from the H-2K^b mRNA in the way in which the precursor RNA is posttranscriptionally spliced. The nucleotide sequence of the mRNA from which pH2K01 was produced in the region coding for amino acids 90 through 92 and including the termination codon is:

90	91	92	
Gly	Gly	Glu	Stop
GGC	G/GU	GAG	UGA

↑

The underlined sequence resembles an exon/intron splice junction (Breathnach et al. 1978, Lerner et al. 1980, Ohno 1980, Sharp 1981) except for the presence of the C residue (arrow) in the penultimate position of the exon. This position is most usually an A or U residue (Breathnach et al. 1978, Lerner et al. 1980, Ohno 1980, Sharp 1981). In subsequent experiments we have identified and partially sequenced two other H-2^b transplantation antigen cDNA clones (A.A. Reyes, M. Schöld and R. B. Wallace, manuscript in preparation). In at least one of these clones the codon for Gly90 is GGC, i.e., the nucleotide in the penultimate position of the 5′ exon is a C and is spliced. In an H-2^k transplantation antigen cDNA clone isolated by Steinmetz and co-workers (1981), the Gly90 codon is GGT, indicating that this splice boundary in other *H-2* genes might be UG/G. These observations suggest that the splice boundary CG/GUGAGU is sometimes spliced to produce an H-2K^b mRNA and sometimes is not, producing an mRNA which codes for a 92 amino-acid truncated protein. There are other examples where more one mRNA is produced from a single precursor RNA by alternate routes of splicing. Most notable in this regard are the splicing mechanisms involved in the generation of the membrane-bound and secreted forms of murine IgM (Singer et al. 1980, Rogers et al. 1980, Early et al. 1980) and the two mRNAs coding for murine IgD and IgM in some lymphocytes (Early et al. 1980, Maki et al. 1981, Moore et al. 1981).

If the proposed RNA splice site is correct for all transplantation antigens, then one would predict that amino acid 91 would always be Gly because the splicing event would reconstitute a Gly codon when a G is contributed from the 3′ exon, which is most often the case (Breathnach et al. 1978, Lerner et al. 1980, Ohno 1980, Sharp 1981). Residue 91 has been shown to be Gly in all mouse and human transplantation antigens sequenced to date (Coligan et al. 1981, Steinmetz et al. 1981).

Is a 92 amino-acid protein translated from the mRNA from which pH2K01 was produced and, if so, what is its function? Again, we do not have an answer to this question. Such a protein, to our knowledge, has never been detected with anti-H-2K^b alloantisera. There is precedent, however, for a role for small peptides, such

as β2m (Nakamuro et al. 1973, Peterson et al. 1974), βt (Ziegler and Milstein 1979), and Thy-1 antigen (Campbell et al. 1979), in the structure and function of membrane bound proteins. Strominger (1980, Ploegh et al. 1981) has suggested that the transplantation antigens can be divided into five structural regions, an N-terminal region (13 500 daltons) containing the glycosylated Asn, two 10 500 dalton regions containing internal disulfide bridges, a fourth transmembrane region (3 000 daltons) and a fifth cytoplasmic region (3 500 daltons). It seems reasonable to speculate that Gly91 is the boundary between the first and second regions and that pH2K01 encodes a portion of a protein consisting of only the N-terminal region. In addition, the results reported here suggest that, in the *H-2K*^{b} gene, the DNA coding for the first and second subregions are separated by an intervening sequence as is the DNA coding for the constant region domains in immunoglobulin genes (Sakano et al. 1979, Calame et al. 1980) and that expression of this gene may be controlled by alternative routes of posttranscriptional splicing.

In conclusion, we have identified a cloned cDNA derived from an H-2K^b-related mRNA. Since the clone was isolated by a method which relies only on the amino-acid sequence of the protein, we have no information about the protein or its function. The oligonucleotide screening method used has proven to be a very selective procedure for the isolation of an extremely low abundance cloned DNA. This method differs from that used by others for the isolation of human (Ploegh et al. 1980, Sood et al. 1981) and mouse (Kvist et al. 1981) transplantation antigen cDNA clones. In particular, the method described here allows mass screening of unordered colonies and is thus a rapid and efficient method for the identification of cloned DNAs coding for proteins with known amino-acid sequences.

Acknowledgements. We wish to thank Dr. Susumu Ohno for many stimulating discussions and Dr. Takeshi Matsunaga for critical reading of this manuscript. This work was supported by NIH grants GM26391 (RBW) and GM25658 (KI). C. Morin's stay was made possible in part by a grant from NATO.

References

Birnboim, H.C. and Doly, J.: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res. 7:* 1513–1523, 1979

Bodmer, W.F.: A new genetic model for allelism at histocompatibility and other complex loci: Polymorphism for control of gene expression. *Transplant. Proc. 4*: 1471–1475, 1973

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P.: Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA 75*: 4853–4857, 1978

Calame, K., Rogers, J., Early, P., Davis, M., Livant, D., Wall, R., and Hood, L.: Mouse Cμ heavy chain immunoglobulin gene segment contains three intervening sequences separating domains. *Nature 284*: 452–455, 1980

Campbell, D.G., Williams, A.F., Bayley, P.M., and Reid, K.B.M.: Structural similarities between Thy-1 antigen from rat brain and immunoglobulin. *Nature 282*: 341–342, 1979

Casadabhan, M. and Cohen, S.N.: Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli. J. Mol. Biol. 138*: 179–207, 1980

Chang, A.C.Y., Nunberg, J.H., Kaufman, R.J., Ehrlich, H.A., Schimke, R.T., and Cohen, S.N.: Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature 275:* 617–624, 1978

Clewell, D.B. and Helinski, D.R.: Supercoiled circular DNA-protein complex in *Escherichia coli*: Purification and induced conversion to an open circular DNA form. *Proc. Natl. Acad. Sci. USA 62*: 1159–1166, 1969

Coligan, J.E., Kindt, T.J., Uehara, H., Martinko, J., and Nathenson, S.G.: Primary structure of a murine transplantation antigen. *Nature 291*: 35–39, 1981

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L.: Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell 20*: 313–319, 1980

Gergen, J.P., Stern, R.H., and Wensink, P.C.: Filter replicas and permanent collections of recombinant DNA plasmids. *Nucleic Acids Res. 7*: 2115–2136, 1979

Goeddel, D.V., Shepard, H.M., Yelverton, E., Leung, D., and Crea, R.: Synthesis of human fibroblast interferon by *E. coli*. *Nucleic Acids Res. 8*: 4057–4074, 1980

Itakura, K., Miyake, T., Kawashima, E.H., Ike, Y., Ito, H., Morin, C., Reyes, A.A., Johnson, M.J., Schöld, M., and Wallace, R.B.: Chemical synthesis and application of oligonucleotides of mixed sequence. *In* A.G. Walton (ed.): *Case Western Third Cleveland Symposium*, Elsevier Amsterdam, in press, 1981

Klein, J.: *Biology of the Mouse Histocompatibility-2 Complex*, Springer-Verlag, New York, 1975

Klein, J.: The major histocompatibility complex of the mouse. *Science 203*: 516–521, 1979

Klein, J., Juretic, A., Baxevanis, C.N., and Nagy, Z.A.: The traditional and a new version of the mouse H-2 complex. *Nature 291*: 455–460, 1981

Kvist, S., Bregegere, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammar, D., Abastado, J.P., Gachelin, G., Peterson, P.A., Dobberstein, B., and Kourilsky, P.: cDNA clone coding for part of a mouse H-2^d major histocompatibility antigen. *Proc. Natl. Acad. Sci. USA 78*: 2772–2776, 1981

Kushner, S.R.: An improved method for transformation of *Escherichia coli* with ColE1 derived plasmids. *In* H.W. Boyer and S. Nicosia (eds): *Genetic Engineering*, pp. 17–23, Elsevier-North Holland, Amsterdam, 1978

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A.: Are snRNPs involved in splicing? *Nature 283*: 220–224, 1980

Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W., and Tonegawa, S.: The role of DNA rearrangement and alternative RNA processing in the expression of immunoglobulin delta genes. *Cell 24*: 353–365, 1981

Maxam, A. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol. 65*: 499–560, 1980

Miyoshi, K. and Itakura, K.: Solid phase synthesis of nonadecathymidylic acid by the phosphotriester approach. *Tetrahedron Lett. 38*: 3635–3638, 1979

Moore, K.W., Rogers, J., Hunkapiller, T., Early, P., Nottenburg, C., Weissman, I., Bazin, H., Wall, R., and Hood, L.: Expression of IgD may use both DNA rearrangement and RNA splicing mechanisms. *Proc. Natl. Acad. Sci. USA 78*: 1800–1804, 1981

Nakamuro, K., Tanigaki, N., and Pressman, D.: Multiple common properties of human β2-microglobulin and the common portion fragment derived from HL-A antigen molecules. *Proc. Natl. Acad. Sci. USA 70*: 2863–2865, 1973

Ohno, S.: Origin of intervening sequences within mammalian genes and the universal signal for their removal. *Differentiation 17*: 1–15, 1980

Peterson, P.A., Rask, L., and Lindblom, J.B.: Highly purified papain-solubilized HL-A antigens contain β_2-microglobulin. *Proc. Natl. Acad. Sci. USA 71*: 35–39, 1974

Ploegh, H.L., Orr, H.T., and Strominger, J.L.: Molecular cloning of a human histocompatibility antigen cDNA fragment. *Proc. Natl. Acad. Sci. USA 77*: 6081–6085, 1980

Ploegh, H.L., Orr, H.T., and Strominger, J.L.: Major histocompatibility antigens: The human (HLA-A, -B, -C) and murine (H-2K, H-2D) class I molecules. *Cell 24*: 287–299, 1981

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R.: Two mRNAs with different 3′ ends encode membrane bound and secreted forms of immunoglobulin μ chain. *Cell 20*: 303–312, 1980

Roychourdhury, R. and Wu, R.: Terminal transferase-catalyzed addition of nucleotides to the 3′ termini of DNA. *Methods Enzymol. 65*: 43–62, 1980

Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S.: Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. *Nature 277*: 627–633, 1979

Sanger, F., Nicklen, S., and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA 74*: 5463–5467, 1977

Sharp, P.A.: Speculations on RNA splicing. *Cell 23*: 643–646, 1981

Silver, J. and Hood, L.: Preliminary amino acid sequence of transplantation antigens: Genetic and evolutionary implications. *Contemp. Top. Mol. Immunol. 5*: 35–86, 1976

Singer, P.A., Singer, H.H., and Williamson, A.R.: Different species of messenger RNA encode receptor and secretory IgM μ chains differing at their carboxy termini. *Nature 285*: 294–300, 1980

Singer, R.H. and Penman, S.: Messenger RNA in HeLa cells: Kinetics of formation and decay. *J. Mol. Biol. 78*: 321–334, 1973

Sood, A.K., Pereira, D., and Weissman, S.M.: Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligodeoxynucleotide primer. *Proc. Natl. Acad. Sci. USA 78*: 616–620, 1981

Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol. 98*: 503–517, 1975

Steinmetz, M., Frelinger, J.G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S.M., Uehara, H., Nathenson, S., and Hood, L.: Three cDNA clones encoding mouse transplantation antigens: Homology to immunoglobulin genes. *Cell 24*: 125–134, 1981

Strominger, J.L.: Structure of products of the major histocompatibility complex in man and mouse. *In* M. Fougereau and J. Dausset (eds.): *Progress in Immunology*, Vol. 4, pp. 541–554, Academic Press, New York, 1980

Sutcliffe, J.G.: pBR322 restriction map derived from the DNA sequence: Accurate DNA size makers up to 4361 nucleotide pairs long. *Nucleic Acids Res. 5*: 2721–2728, 1978

Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H., and Itakura, K.: Use of synthetic oligonucleotides as hybridization probes. III. Isolation of cloned cDNA sequences for human β2-microglobulin. *Proc. Natl. Acad. Sci. USA*, in press, 1981

Tu, C.-P.D. and Cohen, S.N.: 3'-End labeling of DNA with [α-^{32}P] cordycepin-5'-triphosphate. *Gene 10*: 177–183, 1980

Uehara, H., Coligan, J.E., and Nathenson, S.G.: Amino acid sequence of the COOH-terminal hydrophilic region of the H-2K^b MHC alloantigen: Completion of the entire primary structure of the H-2K^b molecule. *Biochemistry*, in press, 1981

Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K.: Hybridization of synthetic oligodeoxynucleotides to ϕX174 DNA: The effect of single base pair mismatch. *Nucleic Acids Res. 6*: 3543–3557, 1979

Wallace, R. B., Johnson, M. J., Hirose, J., Miyake, J., Kawashima, E.H., and Itakura, K.: The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequences to rabbit β-globin DNA. *Nucleic Acids Res. 9*: 879–894, 1981a

Wallace, R.B., Johnson, M.J., Suggs, S.V., Miyoshi, K., Bhatt, R., and Itakura, K.: A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmid vector pBR322. *Gene*, in press, 1981b

Walseth, T.F. and Johnson, R.A.: The enzymatic preparation of [α-^{32}P] nucleoside triphosphates, cyclic [^{32}P]AMP, and cyclic[^{32}P]GMP. *Biochim. Biophys. Acta 562*: 11–31, 1979

Ziegler, A. and Milstein, C.: A small polypeptide different from β2-microglobulin associated with a human cell surface antigen. *Nature 279*: 243–244, 1979

Zinkernagel, R.M. and Doherty, P.C.: Immunological surveillance against altered self components by sensitized T lymphocytes in lymphocytic choriomeningitis. *Nature 251*: 547–548, 1974

Zinkernagel, R.M. and Doherty, P.C.: MHC-restricted cytotoxic T cells: Studies on the biological role of polymorphic transplantation antigens determining T-cell restriction-specificity, function, and responsiveness, *Adv. Immunol. 27*: 51–177, 1979

Chapter III

Isolation of a cDNA Clone for the Murine Transplantation Antigen H-2K^b

Antonio A. Reyes, Monica Schöld*, Keiichi Itakura*
and R. Bruce Wallace*

*Department of Molecular Genetics, City of Hope Research Institute, Duarte, CA 91010.

Abstract

A library of cloned cDNA constructed from the poly(A)$^+$ RNA of the murine thymoma cell line EL4 (b-haplotype) was screened with a probe encoding a short region of the H-2K^b transplantation antigen. One of the clones isolated, pH202, contains a region which can code for a transplantation antigen with an amino acid sequence 98% homologous to that previously published for H-2K^b. Based on this high degree of homology pH202 appears to encode the H-2K^b antigen, from amino acid residue 66 through the carboxy-terminus including 386 nucleotides of 3'-untranslated sequence. The amino acid sequence deduced from pH202 suggests that the H-2K^b antigen is actually 2 amino acids longer than previously reported (a total of 348 residues). Four other differences in amino acid assignments are seen. Analysis of the DNA sequences of pH202 and other H-2 clones previously described in the literature suggests that alternative routes of splicing at the 3'-end of the coding region are involved in the production of different transplantation antigen mRNAs.

Introduction

The major histocompatibility complex (MHC) of the mouse (H-2) encodes several proteins which are involved in the immune function (1, 2). The classical transplantation antigens, or Class I antigens, are integral membrane glycoproteins encoded in the H-2K, D and L loci. These molecules play a major role in allograft rejection as well as in the recognition of altered cells by cytotoxic T lymphocytes (3, 4). Class I antigens are highly polymorphic. Analysis of limited protein sequence data from several antigens has shown that the differences are distributed throughout the molecule (5). Antigens from different alleles are no more homologous to each other than those from different loci.

In order to understand the molecular basis of this polymorphism and its correlation with function, as well as to obtain tools to probe the genetic organization of the

MHC, it has become necessary to clone MHC genes. cDNA clones homologous to short segments of several MHC antigens have been isolated from mouse (6-8) and man (9-11). The characterization of a mouse H-2 pseudogene clone has also been reported (12). Previously we described the identification of an H-2K^b-related clone, pH2K01, which encodes the N-terminal domain of H-2K^b (8). We have produced a radiolabeled probe from pH2K01 and used it to identify a mouse cDNA clone which contains the coding region for most of the H-2K^b sequence, starting from amino acid 66 through the carboxy-terminus, together with a portion of the 3'-untranslated region.

Materials and Methods

Enzymes and Reagents

Restriction enzymes were purchased from Bethesda Research Laboratories, New England Biolabs and Boehringer-Mannheim and used as suggested by the suppliers. DNA polymerase I (Klenow fragment) was from Boehringer-Mannheim. $\alpha[^{32}P]$dNTPs, terminal deoxynucleotidyl transferase and T4 polynucleotide kinase were from New England Nuclear. $\gamma[^{32}P]$ATP was synthesized by the method of Walseth and Johnson (13).

Synthesis of Probe

pH2K01 DNA (8) was used as a template to synthesize a 133-nucleotide probe, using the oligonucleotide 5'-TGGATGGA-3' (which codes for Trp51-Glu53) as a primer (Fig. 1). pH2K01 DNA was digested with AluI endonuclease and the fragment containing the cDNA insert isolated by electrophoresis on a 5% acrylamide gel followed by electroelution. The AluI fragment (100 ng) was digested with SmaI endonuclease in 7 mM NaCl, 7 mM Tris-HCl pH .5, 7 mM DTT, 7 mM $Mg(OAc)_2$ at 37°C for one hour. After adjusting the NaCl concentration to 50 mM and adding 40 ng of TGGATGGA, the mixture was heated to 100°C for three minutes and immediately cooled. All

four $\alpha[^{32}P]$dNTPs were added to a final concentration of 8 µM. After the addition of DNA polymerase I (Klenow fragment) the reaction was incubated for 15 minutes at room temperature, unlabeled dNTPs were added to 0.1 mM, and the reaction allowed to proceed for another five minutes. The DNA was precipitated with two volumes of ethanol after the addition of yeast tRNA as carrier, dissolved in formamide loading buffer (14), boiled for one minute and electrophoresed in a 5% acrylamide/8 M urea gel. The band corresponding to the 133-nucleotide probe was cut from the gel and the DNA isolated by electroelution. Specific activity obtained was approximately 2×10^8 cpm/µg.

Prior to colony screening, the labeled DNA (20 ng) was mixed with 1.4 µg of nicked pBR322 DNA as cold competitor to reduce nonspecific hybridization. The mixture was adjusted to 60 mM NaCl, 50 mM Tris-HCl pH 7.5, 7 mM DTT, 7 mM $Mg(OAc)_2$, heated to 100°C for three minutes and incubated at 30°C for 60 minutes.

Colony Screening

Double stranded cDNA constructed from the poly(A)$^+$ RNA of the murine thymoma cell line EL4 (obtained from Dr. M. Cohn of Salk Institute) was inserted into the PstI site of pBR322 by the G-C tailing method as described previously (8). The recombinant plasmid was used to transform E. coli K12 strain MC1061 (gift of Dr. J. Rossi) (15). After overnight growth at 37°C on L-broth agar plates containing 20 µg/ml tetracycline, plasmid DNA was prepared from 86,000 colonies by the procedure of Birnboim (16). This DNA represents a "library" of cDNA clones. A fraction of the DNA was used to retransform MC1061 cells. Approximately 100,000 colonies were plated onto nitrocellulose filters placed on ten 100 mm plates. Replica filters were made on nitrocellulose and the filters prepared for screening as described (17). The filters were prehybridized in 6 X NET (NET = 0.15 M NaCl, 0.015 M Tris-HCl pH 7.5, 0.001 M EDTA), 0.5% NP40, 100 µg/ml sonicated, single stranded E. coli DNA at 55°C for two hours. They were then hybridized with ^{32}P-labeled probe (above)

in 6 X NET, 0.5% NP40, 250 μg/ml yeast tRNA at 1-2 ng/ml probe overnight at 60°C. The filters were then washed at room temperature for 15 minutes with three changes of 6 X SSC (SSC = 0.15 M NaCl, 0.015 M sodium citrate pH 7.2), 0.1% SDS and autoradiographed. Colonies giving positive hybridization were replated at a lower density and rescreened similarly. For a final identification, plasmid DNAs were prepared from colonies which were positive on the second screening. DNA samples were digested with HinfI endonuclease, electrophoresed on a 1% agarose gel, transferred to nitrocellulose filter by the method of Southern (18), and hybridized with the same probe. Two different recombinant clones were thus identified, and one of these is designated pH202. Plasmid DNA was purified as described previously (8).

DNA Sequence Analysis

The cDNA insert of pH202 was mapped with several restriction enzymes. Restriction fragments were 3'-end labeled with $\alpha[^{32}P]$dNTPs using the large fragment of DNA polymerase I (Klenow fragment) or with $\alpha[^{32}P]$cordycepin triphosphate using terminal deoxynucleotidyl transferase (19). Other fragments were 5'-end labeled with $\gamma[^{32}P]$ATP using T4 polynucleotide kinase (14). Fragments thus labeled at both ends were then either cleaved with a second restriction enzyme or strand separated (14). All uniquely end-labeled fragments were sequenced by the method of Maxam and Gilbert (14).

Results and Discussion

Earlier we reported the isolation of the mouse cDNA clone pH2K01 (8) which codes for a protein sequence identical to 42 consecutive amino acid residues of $H\text{-}2K^b$ (50 through 91). A 133-nucleotide probe was synthesized using pH2K01 DNA as template and the oligonucleotide 5'-TGGATGGA-3' (Trp51-Glu53) as a primer (Fig. 1). The primer was annealed to the AluI-SmaI fragment of pH2K01 and extended in

the 3' direction by incorporating $\alpha[^{32}P]$dNTPs using the large fragment of DNA polymerase I. The resulting 133-nucleotide probe contains all but the initial three bases of the cDNA coding region and extends seven bases into the 3'-noncoding region of pH2K01. A cloned cDNA library derived from mRNA of the EL4 cell line (b-haplotype) was screened with this probe. Two clones were isolated. One of these, pH202, was mapped with various restriction enzymes (Fig. 1) and sequenced completely (Fig. 2). The cDNA insert of pH202 is 1235 bp long and contains most of the coding region for H-2K^b (residues 66-348) together with the 3'-untranslated region. pH202 and pH2K01 have identical sequences where they overlap, from positions 66 through 90 (a splice site is proposed at position 91, see below).

The deduced amino acid sequence of pH202 is 98% homologous with the published sequence of H-2K^b (5, Fig. 3). There are nine differences in amino acid assignments, namely in positions 193, 263, 268, 275, 276, 313, 343, 347 and 348. Three of these, His263, Glu275 and Pro276, have been shown to be correct for H-2K^b by additional protein sequencing studies by Coligan and coworkers (J. Coligan, personal communication). Two of the remaining differences involve the assignment of Asp(D). The radiochemical sequencing method used to sequence H-2K^b is least reliable for determining Asp since this residue is assigned indirectly if a given cycle does not contain any of the other 19 amino acids in the carboxy-terminus, Leu and Ala. There is no evidence to date that these two residues are present in the mature protein, although the inability to detect them could be due to the difficulty in sequencing the C-terminus. Protein sequence analysis of this region of H-2K^b has not ruled out the possibility that the molecule is longer than 346 residues (20). Because of the reasons cited above we believe that pH202 encodes H-2K^b. However, it is possible that pH202 in fact codes for a previously unidentified protein very closely related to H-2K^b.

Steinmetz et al. (12) recently isolated and sequenced the transplantation antigen pseudogene 27.1 which was speculated to be split into eight exons. Splice sites were

proposed at residues 91, 183, 275, 315, 326 and 339. Analysis of the DNA sequence of pH202 in these regions shows the presence of exon donor/acceptor sites which are identical to those proposed in pseudogene 27.1 (Fig. 2). However, if we assume that pseudogene 27.1 is representative of all transplantation antigen genes, then pH202 appears to be spliced differently at its 3'-end. In the model proposed by Steinmetz et al. (12), the splice between exons 7 and 8 straddles the codon for amino acid 339, which is immediately followed by a termination codon and the 3'-untranslated region. The intron-exon boundaries in pseudogene 27.1 were delineated by comparison with the structure of the H-2^d cDNA clone pH-2I isolated by Hood and coworkers (6, Fig. 4). pH-2I codes for 27 amino acids homologous to positions 313 to 339 of H-$2K^b$, followed by a termination codon. Since pH202 codes for nine additional amino acids (340-348) in the carboxy-terminus which are not found in pH-2I, and since the 3'-noncoding regions of pH202 and pH-2I are highly homologous, we propose that for H-$2K^b$ exon 8 starts further upstream (5') than was proposed by Steinmetz et al. (12). Exon 8 could contain two alternative acceptor sites (Fig. 4). Splicing pattern a gives rise to an H-$2K^b$-like message in which exons 7 and 8a are spliced, while b generates a pH-2I-like mRNA in which exon 7 is spliced to exon 8b. In either case the same termination codon is encountered, followed by the same 3'-noncoding region. Indeed, a splice site can be found in the last intron of pseudogene 27.1, 27 bases upstream of the proposed start of exon 8. If the gene were spliced at this site, the intron spliced out would still obey the GT-AG rule proposed by Breathnach and Chambon (21). Furthermore, the 27-base sequence which follows this putative splice site is 85% homologous with the region coding for amino acid residues 399-347 in pH202.

Another H-2^d cDNA clone, pH-2^d, isolated by Kvist et al. (7) has a sequence identical to that of pH-2I in the coding region shown in Fig. 4. The partial sequence of 3'-untranslated DNA immediately following the stop codon in pH-2^d-1 is also highly homologous to that in pH-2I. Thus it is possible that pH-2I and pH2^d-1 are

generated by the same splicing route. Hood and coworkers (5) have isolated another H-2^d cDNA clone, pH-2II, in which exon 8 appears not to be spliced at all. The coding sequence of pH-2II reads through the end of exon 7 into the intron wherein a termination codon is encountered after 13 codons. Although these observations suggest that different transplantation antigen mRNAs may be generated by alternative patterns of 3'-end splicing of their transcripts, this prediction may only be verified with the isolation and characterization of the corresponding genes and analysis of the primary transcripts and mature mRNAs.

Alternative processing of a precursor RNA has been observed in other systems, particularly in the generation of membrane-bound and secreted forms of murine IgM (22-24) and IgD (25, 26). Transplantation antigens are integral membrane molecules with carboxy-termini located inside the cell. Alternative splicing routes at the 3'-end of the transcript could play a role in generating antigens with different cytoplasmic domains and possibly, functions.

The 3'-untranslated region of pH202 does not contain the sequence AATAAA which is thought to be the recognition signal for poly(A) addition (27). The 3'-untranslated regions of pH202 and pseudogene 27.1 are highly homologous, however, and if the two sequences are aligned, the double poly(A) addition signal found in the pseudogene sequence is about 15 nucleotides 3' to where the noncoding region of pH202 ends. Hence, pH202 probably contains most of the 3'-end of the H-$2K^b$ mRNA.

Acknowledgements

We wish to thank Drs. Susumo Ohno and T. Matsunaga for many stimulating discussions and Dr. J. Coligan for providing unpublished information. This work was supported by USPHS grant GM26391. R.B.W. and K.I. are members of the Cancer Research Center (CA16434) at the City of Hope Research Institute. We gratefully acknowledge the use of the PDP 11/34 computer endowed by Daisy E. and Isidor M. Gross.

References

1. Klein, J.: Biology of the Mouse Histocompatibility-2 Complex, Springer-Verlag, New York, NY, 1975.
2. Klein, J., Juretic, A., Baxevanis, C. N., and Nagy, Z. A.: Nature 291: 455-460, 1981.
3. Zinkernagel, R. M., and Doherty, P. C.: Nature 251: 547-548, 1974.
4. Zinkernagel, R. M., and Doherty, P. C.: Adv. Immunology 27: 51-177, 1979.
5. Coligan, T. E., Kindt, T. J., Uehara, H., Martin, K. J., and Nathenson, S. G.: Nature 29: 35-39, 1981.
6. Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S., Uehara, H., Nathenson, S., and Hood, L.: Cell 24: 125-134, 1981.
7. Kvist, S., Bregegere, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammer, D., Abastado, J. P., Gachelin, C., Peterson, P. A., Dobberstein, B., and Kourilsky, P.: Proc. Natl. Acad. Sci. USA 78: 2772-2776, 1981.
8. Reyes, A. A., Johnson, M. J., Schöld, M., Ito, H., Ike, Y., Morin, C., Itakura, K., and Wallace, R. B.: Immunogenetics 14: 383-392, 1981.
9. Jordan, B. R., Bregegere, F., and Kourilsky, P.: Nature 290: 521-523, 1981.
10. Sood, A. K., Perevia, P., and Weissman, S. M.: Proc. Natl. Acad. Sci. USA 78: 616-620, 1981.
11. Pleogh, H. L., Orr, H. T., and Strominger, J. L.: Proc. Natl. Acad. Sci. USA 77: 6081-6085, 1980.
12. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A., and Hood, L.: Cell 25: 683-692, 1981.
13. Walseth, T. F., and Johnson, R. A.: Biochim. Biophys. Acta 562: 11-31, 1979.
14. Maxam, A., and Gilbert, W.: Methods in Enzymology 65: 499-560, 1980.
15. Casadabhan, M., and Cohen, S. N.: J. Mol. Biol. 138: 179-207, 1980.
16. Birnboim, H. C., and Doly, J.: Nucl. Acids Res. 7: 1513-1523, 1979.
17. Hanahan, D., and Meselson, M.: Gene 10: 63-65, 1980.

18. Southern, E. M.: J. Mol. Biol. 98: 503-517, 1975.

19. Tu, C.-P. D., and Cohen, S. N.: Gene 10: 177-183, 1980.

20. Uehara, H., Coligan, J. E., and Nathenson, S. G.: Biochemistry 20: 5940-5945, 1981.

21. Breathnach, R., and Chambon, P.: Ann. Rev. Biochem. 50: 349-383, 1981.

22. Singer, P. A., Singer, H. H., and Williamson, A. R.: Nature 285: 294-300, 1980.

23. Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R.: Cell 20: 303-312, 1980.

24. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L.: Cell 20: 313-319, 1980.

25. Moore, K. W., Rogers, J., Hunkapiller, T., Early, P., Nottenburg, C., Weissman, I., Basin, H., Wall, R., and Hood, L.: Proc. Natl. Acad. Sci. USA 78: 1800-1804, 1981.

26. Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W., and Tonegawa, S.: Cell 24: 353-365, 1981.

27. Proudfoot, N. J., and Brownlee, G. G.: Nature 263: 211-214, 1976.

Fig. 1. Partial restriction maps of pH2K01 (8) and pH202. The horizontal lines represent pBR322 sequences; open boxes, coding regions; hatched boxes, 3'-untranslated regions; block boxes, G-C tails. The vertical lines indicate overlapping regions of pH2K01 and pH202. The bar below the pH2K01 map shows the location of the 133-nucleotide probe. Arrows indicate direction of sequencing by the Maxam-Gilbert method.

Fig. 2. Nucleotide sequence and deduced amino acid sequence of pH202. Arrowheads show locations of exon-exon boundaries as determined from Steinmetz et al. (12).

Fig. 3. The complete amino acid sequence of H-2K^b. Italicized residues were taken from Coligan et al. (5). Residues 50 through 348 were deduced from the nucleotide sequences of pH2K01 (8) and pH202. Residues which are underlined differ from the previously published H-2K^b sequence (5). Although residues 263, 275 and 276 also differ from the published sequence, recent reevaluation of the protein sequencing data has demonstrated that these residues are actually His, Glu and Pro, respectively (Coligan, personal communication).

Fig. 4. Possible alternative splicing patterns at the 3'-end of transplantation antigen RNAs.

A. The 3'-end DNA sequences of pH202 and pH-2I (6) are aligned with that of pseudogene 27.1 (12). Arrows indicate proposed splice sites. The termination codon is boxed.

B. Splice pattern a generates pH202 (H-2K^b)-like message, while splice b results in pH-2I-like message. Note that both splice patterns obey the GT-AG rule of Breathnach and Chambon (21). The open boxes denote coding regions; hatched boxes, 3'-untranslated region; lines, introns; double vertical lines, the termination codon.

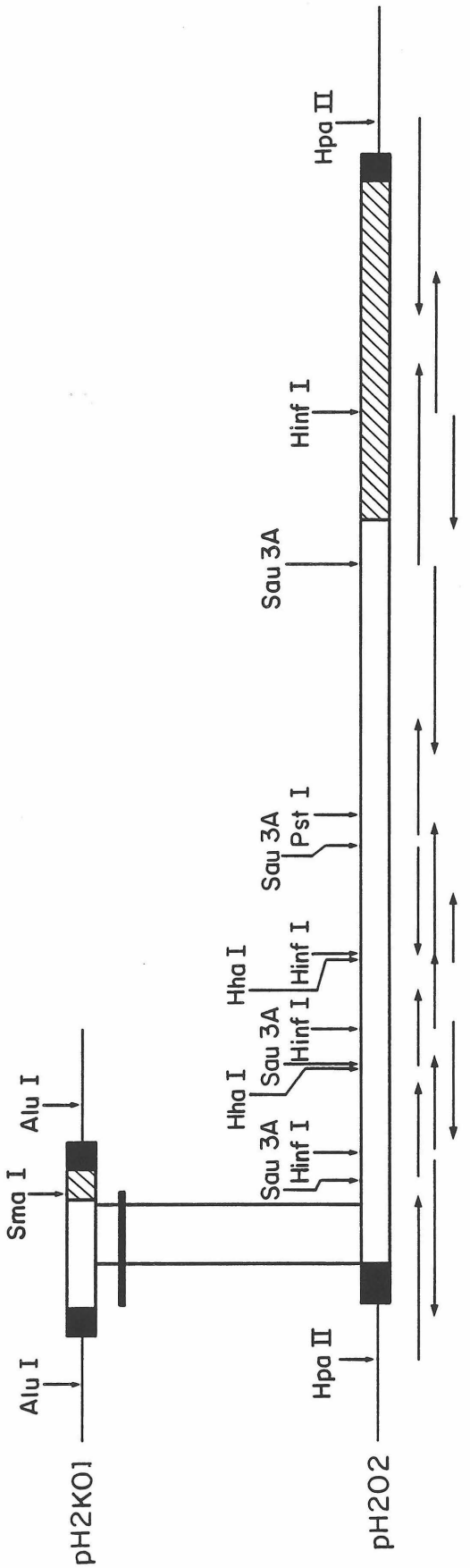
pH2KO1
pH2O2
Alu I
Sma I
Alu I
Hpa II
Sau 3A
Hinf I
Hha I
Sau 3A
Hinf I
Hha I
Sau 3A
Hinf I
Sau 3A
Pst I
Hinf I
Sau 3A
Hpa II
100 bp

Fig. 1

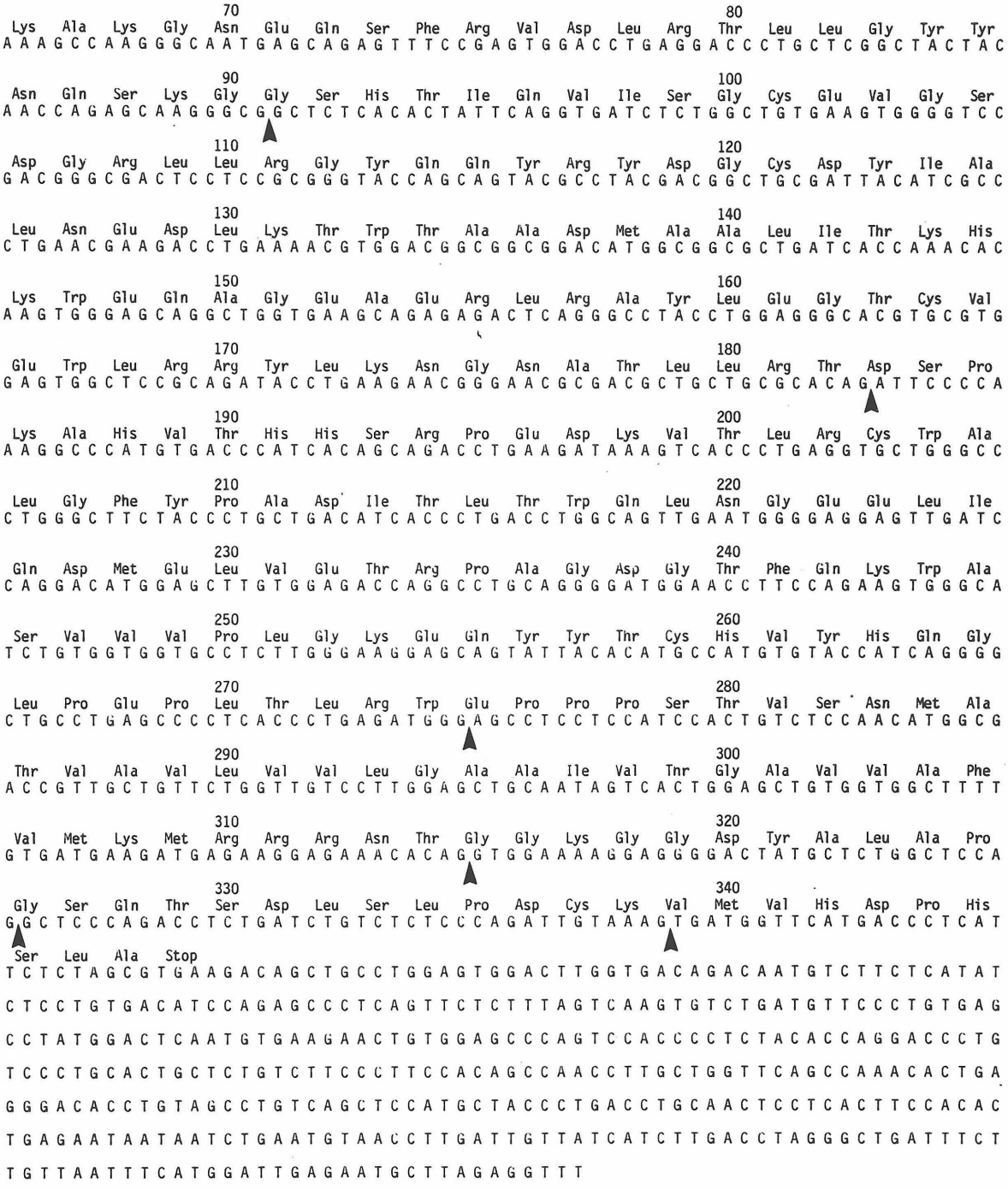
70 80
Lys Ala Lys Gly Asn Glu Gln Ser Phe Arg Val Asp Leu Arg Thr Leu Leu Gly Tyr Tyr
AAAGCCAAGGGCAATGAGCAGAGTTTCCGAGTGGACCTGAGGACCCTGCTCGGCTACTAC
90 100
Asn Gln Ser Lys Gly Gly Ser His Thr Ile Gln Val Ile Ser Gly Cys Glu Val Gly Ser
AACCAGAGCAAGGGCGGCTCTCACACTATTCAGGTGATCTCTGGCTGTGAAGTGGGGTCC
110 120
Asp Gly Arg Leu Leu Arg Gly Tyr Gln Gln Tyr Arg Tyr Asp Gly Cys Asp Tyr Ile Ala
GACGGGCGACTCCTCCGCGGGTACCAGCAGTACGCCTACGACGGCTGCGATTACATCGCC
130 140
Leu Asn Glu Asp Leu Lys Thr Trp Thr Ala Ala Asp Met Ala Ala Leu Ile Thr Lys His
CTGAACGAAGACCTGAAAACGTGGACGGCGGCGGACATGGCGGCGCTGATCACCAAACAC
150 160
Lys Trp Glu Gln Ala Gly Glu Ala Glu Arg Leu Arg Ala Tyr Leu Glu Gly Thr Cys Val
AAGTGGGAGCAGGCTGGTGAAGCAGAGAGACTCAGGGCCTACCTGGAGGGCACGTGCGTG
170 180
Glu Trp Leu Arg Arg Tyr Leu Lys Asn Gly Asn Ala Thr Leu Leu Arg Thr Asp Ser Pro
GAGTGGCTCCGCAGATACCTGAAGAACGGGAACGCGACGCTGCTGCGCACAGATTCCCCA
190 200
Lys Ala His Val Thr His His Ser Arg Pro Glu Asp Lys Val Thr Leu Arg Cys Trp Ala
AAGGCCCATGTGACCCATCACAGCAGACCTGAAGATAAAGTCACCCTGAGGTGCTGGGCC
210 220
Leu Gly Phe Tyr Pro Ala Asp Ile Thr Leu Thr Trp Gln Leu Asn Gly Glu Glu Leu Ile
CTGGGCTTCTACCCTGCTGACATCACCCTGACCTGGCAGTTGAATGGGGAGGAGTTGATC
230 240
Gln Asp Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly Thr Phe Gln Lys Trp Ala
CAGGACATGGAGCTTGTGGAGACCAGGCCTGCAGGGGATGGAACCTTCCAGAAGTGGGCA
250 260
Ser Val Val Val Pro Leu Gly Lys Glu Gln Tyr Tyr Thr Cys His Val Tyr His Gln Gly
TCTGTGGTGGTGCCTCTTGGGAAGGAGCAGTATTACACATGCCATGTGTACCATCAGGGG
270 280
Leu Pro Glu Pro Leu Thr Leu Arg Trp Glu Pro Pro Pro Ser Thr Val Ser Asn Met Ala
CTGCCTGAGCCCCTCACCCTGAGATGGGAGCCTCCTCCATCCACTGTCTCCAACATGGCG
290 300
Thr Val Ala Val Leu Val Val Leu Gly Ala Ala Ile Val Thr Gly Ala Val Val Ala Phe
ACCGTTGCTGTTCTGGTTGTCCTTGGAGCTGCAATAGTCACTGGAGCTGTGGTGGCTTTT
310 320
Val Met Lys Met Arg Arg Arg Asn Thr Gly Gly Lys Gly Gly Asp Tyr Ala Leu Ala Pro
GTGATGAAGATGAGAAGGAGAAACACAGGTGGAAAAGGAGGGGACTATGCTCTGGCTCCA
330 340
Gly Ser Gln Thr Ser Asp Leu Ser Leu Pro Asp Cys Lys Val Met Val His Asp Pro His
GGCTCCCAGACCTCTGATCTGTCTCTCCCAGATTGTAAAGTGATGGTTCATGACCCTCAT
Ser Leu Ala Stop
TCTCTAGCGTGAAGACAGCTGCCTGGAGTGGACTTGGTGACAGACAATGTCTTCTCATAT
CTCCTGTGACATCCAGAGCCCTCAGTTCTCTTTAGTCAAGTGTCTGATGTTCCCTGTGAG
CCTATGGACTCAATGTGAAGAACTGTGGAGCCCAGTCCACCCCTCTACACCAGGACCCTG
TCCCTGCACTGCTCTGTCTTCCCTTCCACAGCCAACCTTGCTGGTTCAGCCAAACACTGA
GGGACACCTGTAGCCTGTCAGCTCCATGCTACCCTGACCTGCAACTCCTCACTTCCACAC
TGAGAATAATAATCTGAATGTAACCTTGATTGTTATCATCTTGACCTAGGGCTGATTTCT
TGTTAATTTCATGGATTGAGAATGCTTAGAGGTTT

Fig. 2

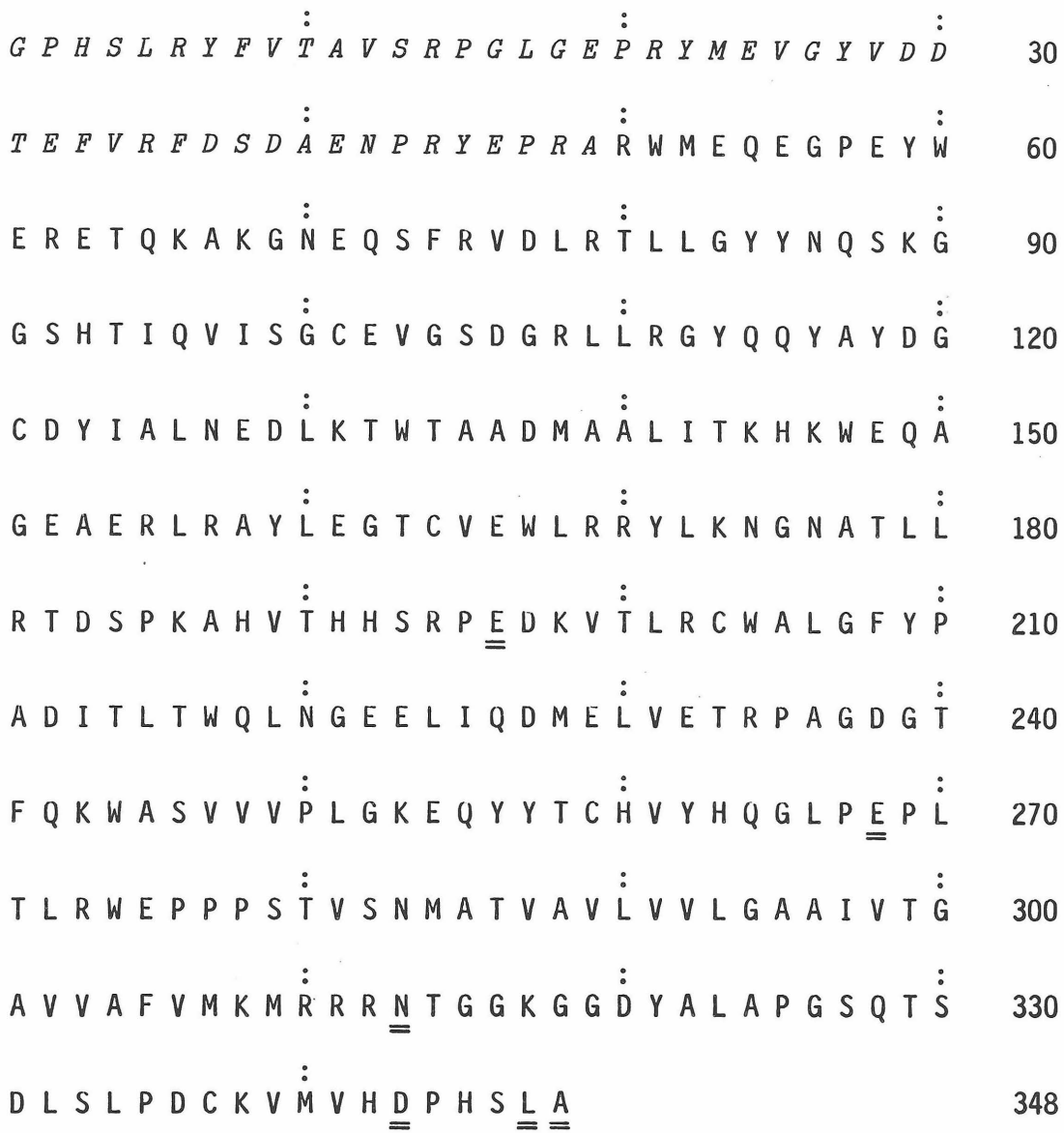
G P H S L R Y F V T A V S R P G L G E P R Y M E V G Y V D D 30
T E F V R F D S D A E N P R Y E P R A R W M E Q E G P E Y W 60
E R E T Q K A K G N E Q S F R V D L R T L L G Y Y N Q S K G 90
G S H T I Q V I S G C E V G S D G R L L R G Y Q Q Y A Y D G 120
C D Y I A L N E D L K T W T A A D M A A L I T K H K W E Q A 150
G E A E R L R A Y L E G T C V E W L R R Y L K N G N A T L L 180
R T D S P K A H V T H H S R P E D K V T L R C W A L G F Y P 210
A D I T L T W Q L N G E E L I Q D M E L V E T R P A G D G T 240
F Q K W A S V V V P L G K E Q Y Y T C H V Y H Q G L P E P L 270
T L R W E P P P S T V S N M A T V A V L V V L G A A I V T G 300
A V V A F V M K M R R R N T G G K G G D Y A L A P G S Q T S 330
D L S L P D C K V M V H D P H S L A 348

Fig. 3

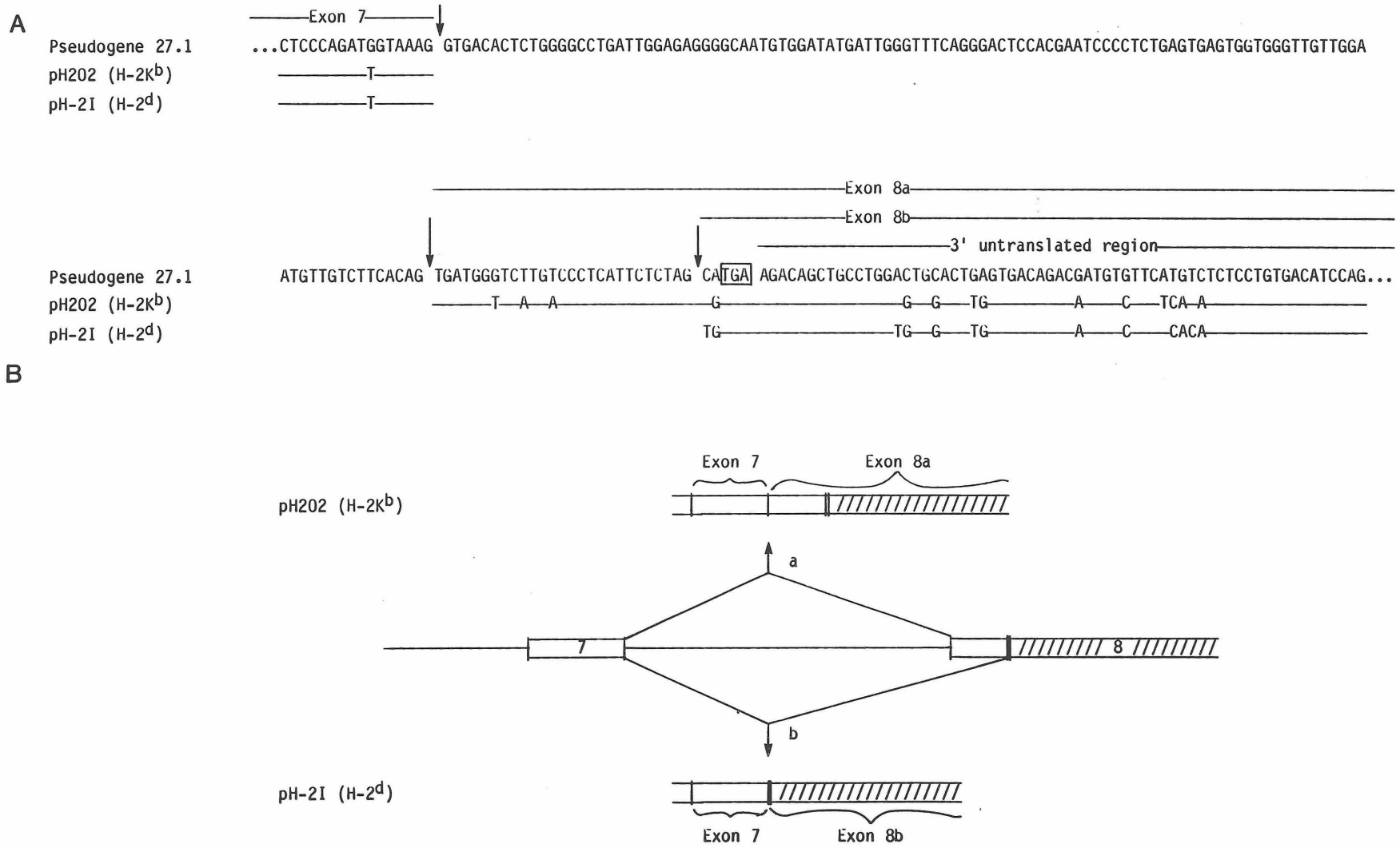
A
Exon 7
Pseudogene 27.1
...CTCCCAGATGGTAAAG GTGACACTCTGGGGCCTGATTGGAGAGGGGCAATGTGGATATGATTGGGTTTCAGGGACTCCACGAATCCCCTCTGAGTGAGTGGTGGGTTGTTGGA
pH202 (H-2K^b)
T
pH-2I (H-2^d)
T
Exon 8a
Exon 8b
3' untranslated region
Pseudogene 27.1
ATGTTGTCTTCACAG TGATGGGTCTTGTCCCTCATTCTCTAG CATGA AGACAGCTGCCTGGACTGCACTGAGTGACAGACGATGTGTTCATGTCTCTCCTGTGACATCCAG...
pH202 (H-2K^b)
T A A G G TG A C TCA A
pH-2I (H-2^d)
TG TG G TG A C CACA
B
Exon 7
Exon 8a
pH202 (H-2K^b)
a
7
8
b
pH-2I (H-2^d)
Exon 7
Exon 8b

Fig. 4

Chapter IV

The Complete Amino Acid Sequence of the Murine Transplantation Antigen H-2D^{b} as Deduced by Molecular Cloning

Antonio A. Reyes, Monica Schöld* and R. Bruce Wallace*

*Department of Molecular Genetics, City of Hope Research Institute, Duarte, CA 91010.

Abstract

A mouse cDNA library derived from the EL4 cell line (b-haplotype) was screened with a probe containing a small part of the H-2K^b coding region. One of the clones isolated, pH203, encodes a protein whose deduced amino acid sequence is identical with the known sequence of H-2D^b in 141 out of 141 positions available for comparison. The clone, therefore, is believed to code for the H-2D^b transplantation antigen. The cDNA insert of pH203 contains the coding region for residues 82 through the carboxy-terminus of H-2D^b, and includes 476 nucleotides of the 3'-untranslated sequence. Comparison between the H-2D^b cDNA clone and a previously isolated H-2K^b cDNA clone shows homologies of 83% and 91% at the amino acid and nucleotide levels, respectively. Analysis of DNA sequences at the 3'-coding and untranslated regions suggests that the mRNAs of H-2K^b and H-2D^b are spliced differently at their 3'-coding ends.

Introduction

The major histocompatibility complex of the mouse (H-2) is a tightly linked cluster of genes on chromosome 17 which controls several immune functions (Klein, 1975; Klein et al., 1981). The H-2 complex is traditionally divided into four regions, K, I, S and D, although there is evidence that the Tla, Qa2,3 and Qa1 regions which map to the right of D and have gene products similar to those of K and D are actually part of the complex (Stanton and Hood, 1980; Soloski et al., 1981; Vitteta et al., 1976; Michaelson et al., 1977). The K and D loci are about 0.5 centimorgan apart and encode Class I molecules—classical transplantation antigens involved in graft rejection and restriction of T-cell specificity. The K region encodes at least one antigen, K, while the D region codes for at least two, D and L. (Levy and Hansen, 1980; Demant and Ivanyi, 1981; Hansen et al., 1981). Class I molecules are integral membrane glycoproteins with a molecular weight of 45,000 and are noncovalently

associated with β 2-microglobulin. Because of the structural similarities between molecules coded for by different loci, it has been speculated that Class I antigen genes arose by duplication of a primordial gene (Bodmer, 1973; Silver and Hood, 1976).

Protein sequencing studies have shown a high degree of polymorphism in murine transplantation antigens. No characteristic "K-ness" or "D-ness" has been found, i.e., antigens coded for by alleles are about as homologous to each other as products of different loci.

In this paper we report the identification of a clone encoding the cDNA for H-2D^b. The deduced amino acid sequence of this clone completes the previously known partial sequence of H-2D^b (Maloy et al., 1981; Maloy and Coligan, 1982). Together with the previously reported H-2K^b DNA and amino acid sequences (Reyes et al., 1982; Coligan et al., 1981), we can now begin to compare products of the K and D regions at a more fundamental level. Furthermore, these two clones should be useful in probing the organization and evolutionary relatedness of the major transplantation antigen genes.

Materials and Methods

A library of cDNA clones was constructed from the murine thymoma cell line EL4 (b-hyplotype) as described previously (Reyes et al., 1981). A total of 100,000 colonies was screened with a 133-nucleotide probe made from the H-2K^b-related cDNA clone, pH2K01 (Reyes et al., 1981, 1982). One of the clones isolated, pH203, was further characterized by mapping with various restriction enzymes (Bethesda Research Laboratories, Boehringer-Mannheim, New England Biolabs). For sequence analysis, uniquely end-labeled fragments were generated by labeling appropriate restriction fragments at the 3'-ends with α[^{32}P]dNTPs or α[^{32}P]cordycepin triphosphate (both from New England Nuclear) or at the 5'-ends with γ[^{32}P]ATP followed by strand separation or digestion with a second restriction enzyme as described (Reyes et al., 1982). DNA was sequenced by the method of Maxam and Gilbert (1980).

Results and Discussion

A 133-nucleotide probe constructed from the H-2K^b-related cDNA clone pH2K01 (Reyes et al., 1981, 1982) was used to screen a library of cloned cDNA derived from the murine thymoma cell line EL4 (b-haplotype). Of the 100,000 colonies screened, two were found to hybridize with the probe. One of these, pH202, was found to encode H-2K^b and has been completely characterized (Reyes et al., 1982). The other clone isolated, pH203, was sequenced completely by the method of Maxam and Gilbert (1980). The cDNA insert of pH203 is 1242 bp long (Fig. 1). The first 773 bp encode the translated portion of the H-2D^b cDNA from residue 82 through the carboxy-terminus. The deduced amino acid sequence of pH203 (Fig. 2) is identical with known positions in H-2D^b in all of the 141 residues available for comparison (Maloy et al., 1981; Maloy and Coligan, 1982). pH203 is therefore believed to code for the H-2D^b transplantation antigen. The termination codon is followed by 476 bp of untranslated sequence.

Together with the published partial sequence of H-2D^b (Maloy et al., 1981; Maloy and Coligan, 1082) pH203 allows the assignment of the complete amino acid sequence of H-2D^b (Fig. 3). The complete protein sequence of H-2K^b has been determined previously (Coligan et al., 1981). We have also reported the isolation of the cDNA clone pH202 which encodes H-2K^b (Reyes et al., 1982). Hence it is now possible to compare H-2D^b and H-2K^b at both amino acid sequence and DNA sequence levels (Fig. 4, Table I).

The H-2D^b molecule is 338 amino acids long. To obtain maximum alignment with H-2K^b, a gap must be inserted in the H-2D^b sequence at position 309. The overall amino acid sequence homology between the two molecules where they overlap (H-2K^b is 10 amino acids longer) is 83%. Differences are distributed throughout the molecule, although regions of high variability are also observed.

The transplantation antigen genes are believed to be split into 8 exons that

correlate with protein domains (Steinmetz et al., 1981a). The antigen molecule can thus be divided into seven regions as defined by these exons (the first exon codes for a hydrophobic leader peptide). Region 1 (exon 2) - first external domain; region 2 (exon 3) - second external domain; region 3 (exon 4) - third external domain; region 4 (exon 5) - transmembrane domain; regions 5-7 (exons 6-8) - cytoplasmic domain. Region 7 is absent in H-2D^b, probably due to a different splicing pattern at the last intron-exon junction (Reyes et al., 1982). The DNA sequences of H-2K^b and H-2D^b can be compared in regions 2 through 6 and a small portion of region 1. The DNA sequence homology in regions 2-6, a total of 744 bases, is 91%.

Several interesting observations can be made from these comparisons. Region 5, the first cytoplasmic exon, consists of 33 absolutely conserved bases in H-2D^b and H-2K^b. An identical sequence is found in corresponding regions of two H-2^d cDNA clones, pH-2I and pH-2II isolated by Steinmetz et al. (1981b). This highly conserved region might have a significant function. Being located just inside the cell membrane, region 5 could play a critical role in communication between the extracellular domains of the antigen and the interior of the cell, although this remains to be demonstrated.

Region 3, the third external domain, is the next most conserved region. Amino acid differences are clustered at either end of this domain. For example, in residues 193-198, five out of six positions are substituted. The middle part, from positions 199 to 255, is identical in both molecules except for one residue. At the DNA level, this corresponds to two changes in 171 bases (98.8% homology).

The third external domain of transplantation antigens has been shown to share limited amino acid sequence homology with immunoglobulin constant region domains (Orr et al., 1979; Strominger et al., 1980) and β 2-microglobulin (Orr et al., 1979; Nathenson et al., 1981). The biological significance of the immunoglobulin-like character of this highly conserved domain is as yet unknown, but might play a role in the non-covalent binding of β 2-microblobulin to this domain (S. Nathenson, personal communication).

Region 4, the transmembrane domain, is not highly conserved. However, a stretch of about 25 amino acids (residues 282-307) which is thought to span the membrane, consists entirely of uncharged and hydrophobic residues in both H-2K^b and H-2D^b. The four amino acid changes in residues 296-299 correspond to 8 base changes, 7 of them consecutive, in the DNA. The high degree of divergence for this region of the antigen could reflect the pressure to maintain neither specific DNA nor amino acid sequences, but rather the highly hydrophobic character of the transmembrane domain.

The H-2K^b and H-2D^b molecules differ in the number of attached carbohydrate moieties. In both molecules, glycosyl units are attached to Asn 86 and Asn 176 (Uehara et al., 1980a, 1980b; Kimball et al., 1980). However, H-2D^b contains a third carbohydrate group which has been localized in the third external domain of the molecule (Kimball et al., 1981; Maloy and Coligan, 1982). The linkage of the third carbohydrate unit of H-2D^b has not been determined. If it is N-asparagine-linked like the first two units, then a possible glycosylation site can be found in the third external domain of H-2D^b. The sequence Asn 256-Tyr 257-Thr 258, a recognition site for glycosyl transferase (Marshall, 1972) is determined by the DNA sequence of pH203. Maloy and Coligan (1982) have also presented several lines of evidence which suggest that residue 256 is a glycosylated Asn.

About 70% of the amino acid differences between H-2K^b and H-2D^b in regions 2-6 are due to single base changes. There are also 7 instances where a single base change in the third position of the codon does not result in the translation of a different amino acid (a "silent" substitution, Table I). It is interesting to note that there are about the same number of transitions as transversions, and that in region 2, the second external domain, the latter predominate.

The H-2D^b molecule is 10 amino acid residues shorter than the H-2K^b molecule, such that the C-terminal residue of H-2D^b aligns with residue 339 of H-2K^b. The

splice between exons 7 and 8 in the transplantation antigen gene is proposed to occur at this position (Steinmetz et al., 1981a). In a previous paper (Reyes et al., 1982) we speculated that exon 8 has two alternative acceptor sites. In H-2K^b exon 7 is spliced to the upstream acceptor site of exon 8, while in the H-2^d cDNA clone pH-2I isolated by Steinmetz et al. (1981b) the downstream acceptor site is used. In either case, the same termination codon is encountered, followed by the same stretch of 3'-noncoding DNA. The DNA sequence of pH203 about the exon 7-exon 8 splice site is very similar to that of pH-2I. Hence, both pH203 (H-2D^b) and pH-2I are probably spliced via the same pathway, while H-2K^b is generated by a different processing route. Differential splicing patterns at the 3'-end of the transplantation antigen gene could result in the production of antigens with distinct cytoplasmic domains and functions.

A potential poly(A) addition site (Proudfoot and Brownlee, 1976) is found in the 3'-untranslated region of pH203 (Fig. 2). The 3'-untranslated DNA sequences of pH202 and pH203 are 89% homologous for about 300 bases following the termination codon. Beyond this region the sequences diverge sharply. It is interesting to note that the 3'-noncoding regions of the two H-2^d cDNA clones, pH-2I and pH-2II, as well as the pseudogene 27.1 isolated by Steinmetz et al. (1981a), share this 300 base homology with pH202 and pH203. The remaining 3'-noncoding DNA sequence of pH202 (H-2K^b) is similar to that of pseudogene 27.1, and that of pH203 (H-2D^b) to pH-2II. Whether this divergence in the 3'-noncoding DNA is due to alternative splicing patterns or to different DNA sequences in the gene and what its significance might be remain to be seen.

The D region in some haplotypes is known to encode at least two gene products, D and L. For example, H-2L has been detected in the d and the q haplotypes (Hansen and Sachs, 1980; Hansen et al., 1981), but not in the b haplotype (Levy and Hansen, 1980). It is not known whether this variability in the number of D region products

among strains is due to a difference in the number of genes or in the level of expression of the genes.

Nairn and coworkers (1980, 1981) have sequenced 179 residues of H-2D^b. The homologies among D^b, D^d and L^d can be calculated over the same 179 positions available for D^d, using protein sequence data reported here and in the accompanying paper (Maloy and Coligan, 1982 and references therein). In comparing these 179 residues H-2D^b is more homologous to H-2L^d (90%) than to H-2D^d (84%). As is suggested by Maloy and Coligan (1982), H-2D^b may be allelic to H-2L^d rather than to H-2D^d. However, this may only be verified by the isolation of appropriate recombinants and the unambiguous immunochemical definition of "D-ness" and "L-ness".

Acknowledgements

We wish to thank Drs. Susumo Ohno and T. Matsunaga for many stimulating discussions and Dr. J. Coligan for providing unpublished information.

This work was supported by USPHS grant GM26391. R.B.W. is a member of the Cancer Research Center (CA16434) at the City of Hope Research Institute. We gratefully acknowledge the use of the PDP 11/34 computer endowed by Daisy E. and Isidor M. Gross.

References

Bodmer, W. F.: A new genetic model for allelism at histocompatibility and other complex loci: polymorphism for control of gene expression. Transplantation Proceedings 4: 1471-1475, 1973.

Coligan, J. E., Kindt, T. J., Nairn, R., Nathenson, S. G., Sachs, D. H., and Hansen, T. H.: Primary structural studies of an H-2L molecule confirms that it is a unique gene product with homology to H-2K and H-2D antigens. Immunology 77: 1134-1138, 1980.

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J., and Nathenson, S. G.: Primary structure of a murine transplantation antigen. Nature 291: 35-39, 1981.

Demant, P., and Ivanyi, D.: Further molecular complexities of H-2 K- and D-region antigens. Nature 290: 146-149, 1981.

Hansen, T. H., and Sachs, D. H.: Isolation and antigenic characterization of the product of a third polymorphic H-2 locus, H-2L. J. Immunology 121: 1469-1472, 1978.

Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. J., and Sachs, D. H.: Immunochemical evidence in two haplotypes for at least three D region-encoded molecules, D, L, and R. J. Immunology 126: 1713-1716, 1981.

Kimball, E. S., Maloy, W. L., Martinko, J. M., Nathenson, S. G., and Coligan, J. E.: Structural studies on H-2D^b molecules indicate that murine major histocompatibility b and d haplotype alloantigens share structural features. Molec. Immunology 17: 1283-1291, 1980.

Kimball, E. S., Maloy, W. L., and Coligan, J. E.: Evidence for three carbohydrate prosthetic groups on mouse histocompatibility antigens H-2K^d and H-2D^b. (Molec.) Immunology 18: 677-689, 1981.

Klein, J.: Biology of the Mouse Histocompatibility-2 Complex, Springer-Verlag, New York, 1975.

Klein, J., Juretic, A., Baxevanis, C. N., and Nagy, Z. A.: The traditional and a new version of the mouse H-2 complex. Nature 291: 455-460, 1981.

Levy, R. B., and Hansen, T. H.: Functional studies of the products of the H-2L locus. Immunogenetics 10: 7-17, 1980.

Maloy, W. L., and Coligan, J. E.: Primary structure of the H-2D^{b} alloantigen: Additional amino acid sequence information, localization of a third site of glycosylation and evidence for K and D region specific sequences. Immunogenetics, in press, 1982.

Maloy, W. L., Nathenson, S. G., and Coligan, J. E.: Primary structure of murine major histocompatibility complex alloantigens. J. Biol. Chem. 256: 2863-2872, 1981.

Marshall, R. D.: Glycoproteins. Ann. Rev. Biochemistry 41: 673-702, 1972.

Maxam, A., and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. In Methods in Enzymology 65: 499-560, 1980.

Michaelson, J., Flaherty, L., Vitteta, E., and Poulik, M. D.: Molecular similarities between the Qa-2 alloantigen and other gene products of the 17th chromosome of the mouse. J. Exp. Medicine 145: 1066-1070, 1977.

Nairn, R., Nathenson, S. G., and Coligan, J. E.: Isolation, characterization and amino acid sequence studies of the cyanogen bromide fragments of the H-2D^{d} glycoprotein. Eur. J. Immunology 10: 495-503, 1980.

Nairn, R., Nathenson, S. G., and Coligan, J. E.: Amino acid sequence of cyanogen bromide fragment CN-C (residues 29-48) of the mouse histocompatibility antigen H-2D^{d}: A comparison of the amino-terminal 100 residues of H-2D^{d}, D^{b}, K^{d} and K^{b} reveals discrete areas of diversity. Biochemistry 20: 4739-4745, 1981.

Nathenson, S. G., Uehara, H., Ewenstein, B. M., Kindt, T. J., and Coligan, J. E.: Primary structural analysis of the transplantation antigens of the murine H-2 major histocompatibility complex. Ann. Rev. Biochemistry 50: 1025-1052, 1981.

Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A., and Strominger, J. L.: The heavy chain of human histocompatibility antigen HLA-B7 contains an immunoglobulin-like region. Nature 282: 266-270, 1979.

Proudfoot, N. J., and Brownlee, G. G.: 3' Noncoding region sequences in eukaryotic messenger RNA. Nature 263: 211-214, 1976.

Reyes, A. A., Johnson, M. J., Schöld, M., Ito, H., Ike, Y., Morin, C., Itakura, K., and Wallace, R. B.: Identification of an H-2K^b-related molecule by molecular cloning. Immunogenetics 14: 383-392, 1981.

Reyes, A. A., Schöld, M., Itakura, K., and Wallace, R. B.: Isolation of a cDNA clone for the murine transplantation antigen H-2K^b. Proc. Natl. Acad. Sci. USA, in press, 1982.

Silver, J., and Hood, L.: Preliminary amino acid sequences of transplantation antigens: Genetic and evolutionary implications. Contemporary Topics in Molecular Immunology 5: 35-86, 1976.

Soloski, M. J., Uhr, J. W., Flaherty, L., and Vitetta, E. S.: Qa-2, H-2K, and H-2D alloantigens evolved from a common ancestral gene. J. Exp. Med. 153: 1080-1093, 1981.

Stanton, T. H., and Hood, L.: Biochemical identification of the Qa-1 alloantigen. Immunogenetics 11: 309-314, 1980.

Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A., and Hood, L.: A pseudogene homologous to mouse transplantation antigens: Transplantation antigens are encoded by eight exons that correlate with protein domains. Cell 25: 683-692, 1981a.

Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S., and Hood, L.: Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. Cell 24: 125-134, 1981b.

Strominger, J. L., Orr, H. T., Parham, P., Ploegh, H. L., Mann, D. L., Bilofsky, H., Saroff, H. A., Wu, T. T., and Kabat, E. A.: An evaluation of the significance of amino acid sequence homologies in human histocompatibility antigens (HLA-A and HLA-B) with immunoglobulins and other proteins, using relatively short sequences. Scand. J. Immunol. 11: 573-592, 1980.

Uehara, H., Ewenstein, B. M., Martinko, J. M., Nathenson, S. G., Coligan, J. E., and Kindt, T. J.: Primary structure of murine major histocompatibility complex alloantigens: amino acid sequence of the amino-terminal one hundred and seventy-three residues of the H-2K^{b} glycoprotein. Biochemistry 19: 306-315, 1980a.

Uehara, H., Ewenstein, B. M., Martinko, J. M., Nathenson, S. G., Kindt, T. J., and Coligan, J. E.: Primary structure of murine major histocompatibility alloantigens: amino acid sequence of the cyanogen bromide fragment Ia (positions 139-228) from the H-2K^{b} molecule. Biochemistry 19: 6182-6188, 1980b.

Vitteta, E. S., Poulik, M. D., Klein, J., and Uhr, J. W.: Beta 2-microglobulin is selectively associated with H-2 alloantigens on murine lymphoid cells. J. Exp. Medicine 144: 179-192, 1976.

Table I

COMPARISON OF AMINO ACID AND DNA SEQUENCES OF H-2K^b AND H-2D^b

REGION[a]	No. Amino Acids in Region	% Amino Acid Sequence Homology	% DNA Sequence Homology	No. Amino Acid Differences			No. Amino Acids with "Silent" Substitution	Number of Base Changes		
				1 Base Change	2 Base Changes	3 Base Changes		Transitions	Transversions	Deletions
1	90	86[b]	93[c]	0	1	0	0	1	1	0
2	92	78	89	12	7	1	5	13	21	0
3	92	90	96	8	1	0	1	7	4	0
4	40	75	85	7	0	3	2	11	4	3
5	11	100	100	0	0	0	0	0	0	0
6	13	69	90	4	0	0	0	0	4	0
Total Homology		83	91[d]							

[a] Regions are defined by exon-exon boundaries proposed by Steinmetz, *et al.* (1981b).
[b] N-terminal amino acid sequence data for H-2K^b and H-2D^b were obtained from Coligan, *et al.* (1981) and Maloy, *et al.* (1981), respectively.
[c] The comparison was based on an overlap consisting of only 30 bases and this figure need not be representative of the whole region.
[d] Regions 2 - 6.

Fig. 1. Partial restriction map of pH203. The horizontal lines represent pBR322 sequence; open box, coding region; hatched box, 3'-untranslated region, black boxes, G-C tails. Arrows indicate direction of sequencing by the Maxam-Gilbert method.

Fig. 2. Nucleotide sequence and deduced amino acid sequence of pH203. Arrowheads show location of exon-exon boundaries as determined by Steinmetz et al. (1981a). A potential poly(A) addition site is underlined.

Fig. 3. The complete amino acid sequence of H-2D^{b}. Italicized residues were taken from Maloy et al. (1981).

Fig. 4. A comparison of the amino acid sequences of H-2K^{b} (Coligan et al., 1981; Reyes et al., 1982) and H-2D^{b}. Numbers above the sequences correspond to regions of the transplantation antigen molecule (see text) while arrows indicate exon-exon boundaries which define these regions (Steinmetz et al., 1981a).

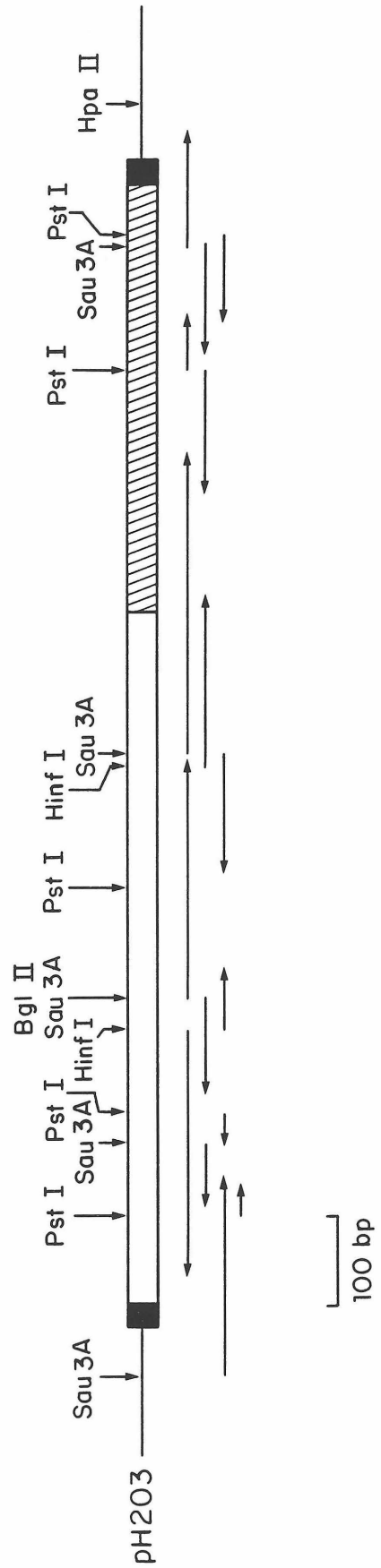
Hpa II
Pst I
Sau 3A
Pst I
Sau 3A
Hinf I
Pst I
Bgl II
Sau 3A
Hinf I
Pst I
Sau 3A
Pst I
Sau 3A
pH203
100 bp

Fig. 1

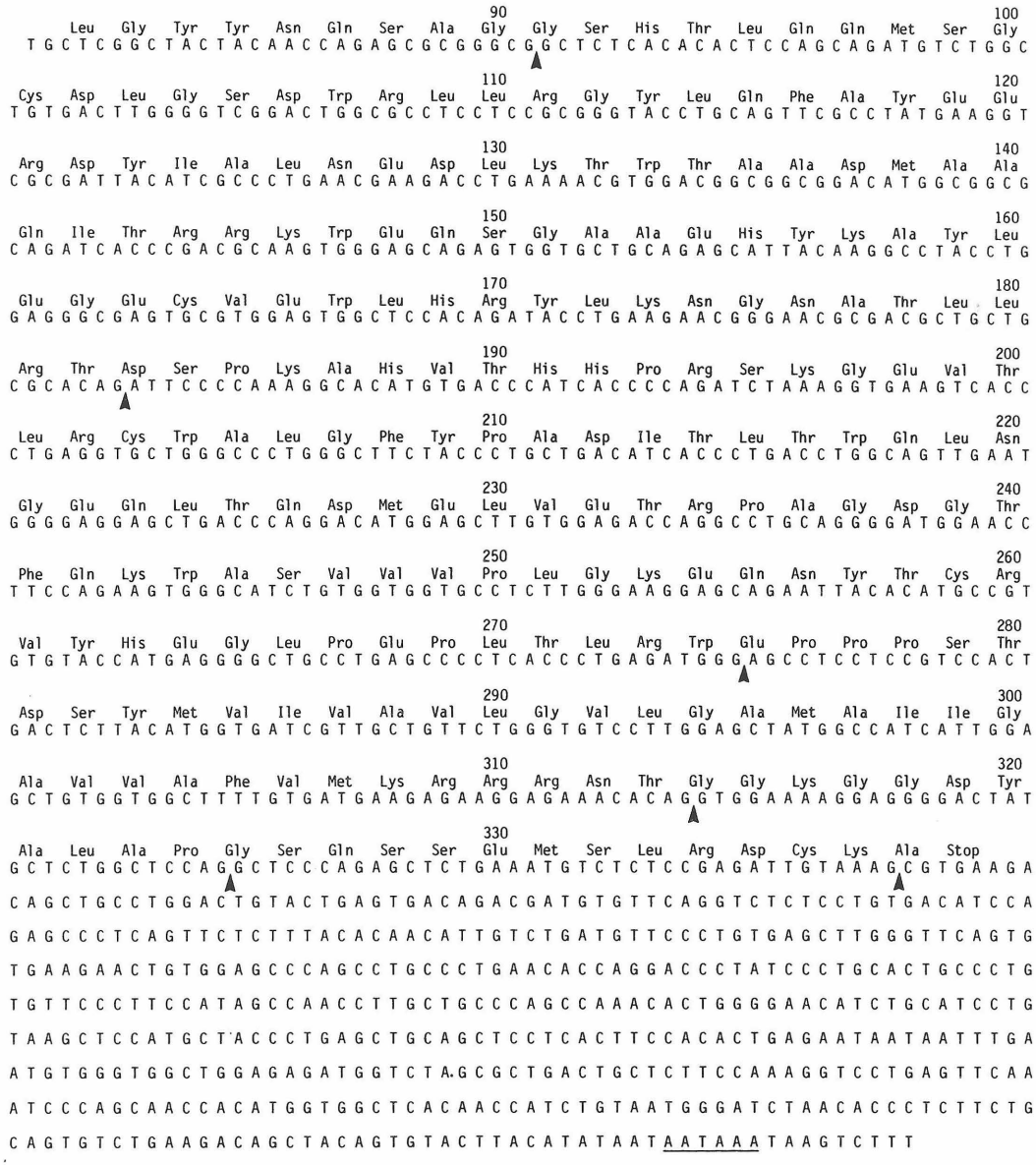
90 100
Leu Gly Tyr Tyr Asn Gln Ser Ala Gly Gly Ser His Thr Leu Gln Gln Met Ser Gly
TGCTCGGCTACTACAACCAGAGCGCGGGCGGCTCTCACACACTCCAGCAGATGTCTGGC
110 120
Cys Asp Leu Gly Ser Asp Trp Arg Leu Leu Arg Gly Tyr Leu Gln Phe Ala Tyr Glu Glu
TGTGACTTGGGGTCGGACTGGCGCCTCCTCCGCGGGTACCTGCAGTTCGCCTATGAAGGT
130 140
Arg Asp Tyr Ile Ala Leu Asn Glu Asp Leu Lys Thr Trp Thr Ala Ala Asp Met Ala Ala
CGCGATTACATCGCCCTGAACGAAGACCTGAAAACGTGGACGGCGGCGGACATGGCGGCG
150 160
Gln Ile Thr Arg Arg Lys Trp Glu Gln Ser Gly Ala Ala Glu His Tyr Lys Ala Tyr Leu
CAGATCACCCGACGCAAGTGGGAGCAGAGTGGTGCTGCAGAGCATTACAAGGCCTACCTG
170 180
Glu Gly Glu Cys Val Glu Trp Leu His Arg Tyr Leu Lys Asn Gly Asn Ala Thr Leu Leu
GAGGGCGAGTGCGTGGAGTGGCTCCACAGATACCTGAAGAACGGGAACGCGACGCTGCTG
190 200
Arg Thr Asp Ser Pro Lys Ala His Val Thr His His Pro Arg Ser Lys Gly Glu Val Thr
CGCACAGATTCCCCAAAGGCACATGTGACCCATCACCCCAGATCTAAAGGTGAAGTCACC
210 220
Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Asp Ile Thr Leu Thr Trp Gln Leu Asn
CTGAGGTGCTGGGCCCTGGGCTTCTACCCTGCTGACATCACCCTGACCTGGCAGTTGAAT
230 240
Gly Glu Gln Leu Thr Gln Asp Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly Thr
GGGGAGGAGCTGACCCAGGACATGGAGCTTGTGGAGACCAGGCCTGCAGGGGATGGAACC
250 260
Phe Gln Lys Trp Ala Ser Val Val Val Pro Leu Gly Lys Glu Gln Asn Tyr Thr Cys Arg
TTCCAGAAGTGGGCATCTGTGGTGGTGCCTCTTGGGAAGGAGCAGAATTACACATGCCGT
270 280
Val Tyr His Glu Gly Leu Pro Glu Pro Leu Thr Leu Arg Trp Glu Pro Pro Pro Ser Thr
GTGTACCATGAGGGGCTGCCTGAGCCCCTCACCCTGAGATGGGAGCCTCCTCCGTCCACT
290 300
Asp Ser Tyr Met Val Ile Val Ala Val Leu Gly Val Leu Gly Ala Met Ala Ile Ile Gly
GACTCTTACATGGTGATCGTTGCTGTTCTGGGTGTCCTTGGAGCTATGGCCATCATTGGA
310 320
Ala Val Val Ala Phe Val Met Lys Arg Arg Arg Asn Thr Gly Gly Lys Gly Gly Asp Tyr
GCTGTGGTGGCTTTTGTGATGAAGAGAAGGAGAAACACAGGTGGAAAAGGAGGGGACTAT
330
Ala Leu Ala Pro Gly Ser Gln Ser Ser Glu Met Ser Leu Arg Asp Cys Lys Ala Stop
GCTCTGGCTCCAGGCTCCCAGAGCTCTGAAATGTCTCTCCGAGATTGTAAAGCGTGAAGA
CAGCTGCCTGGACTGTACTGAGTGACAGACGATGTGTTCAGGTCTCTCCTGTGACATCCA
GAGCCCTCAGTTCTCTTTACACAACATTGTCTGATGTTCCCTGTGAGCTTGGGTTCAGTG
TGAAGAACTGTGGAGCCCAGCCTGCCCTGAACACCAGGACCCTATCCCTGCACTGCCCTG
TGTTCCCTTCCATAGCCAACCTTGCTGCCCAGCCAAACACTGGGGAACATCTGCATCCTG
TAAGCTCCATGCTACCCTGAGCTGCAGCTCCTCACTTCCACACTGAGAATAATAATTTGA
ATGTGGGTGGCTGGAGAGATGGTCTA.GCGCTGACTGCTCTTCCAAAGGTCCTGAGTTCAA
ATCCCAGCAACCACATGGTGGCTCACAACCATCTGTAATGGGATCTAACACCCTCTTCTG
CAGTGTCTGAAGACAGCTACAGTGTACTTACATATAATAATAAATAAGTCTTT

Fig. 2


G P H S M R Y F E T A V S R P G L E E P R Y I S V G Y V D N 30
K E F V R F D S D A E N P R Y E P R A P W M E Q E G P E Y W 60
E R E T Q K A K G Q E Q W F R V S L R N L L G Y Y N Q S A G 90
G S H T L Q Q M S G C D L G S D W R L L R G Y L Q F A Y E G 120
R D Y I A L N E D L K T W T A A D M A A Q I T R R K W E Q S 150
G A A E H Y K A Y L E G E C V E W L H R Y L K N G N A T L L 180
R T D S P K A H V T H H P R S K G E V T L R C W A L G F Y P 210
A D I T L T W Q L N G E E L T Q D M E L V E T R P A G D G T 240
F Q K W A S V V V P L G K E Q N Y T C R V Y H E G L P E P L 270
T L R W E P P P S T D S Y M V I V A V L V G L G A M A I I G 300
A V V A F V M K R R R N T G G K G G D Y A L A P G S Q S S E 330
M S L R D C K A 338

Fig. 3

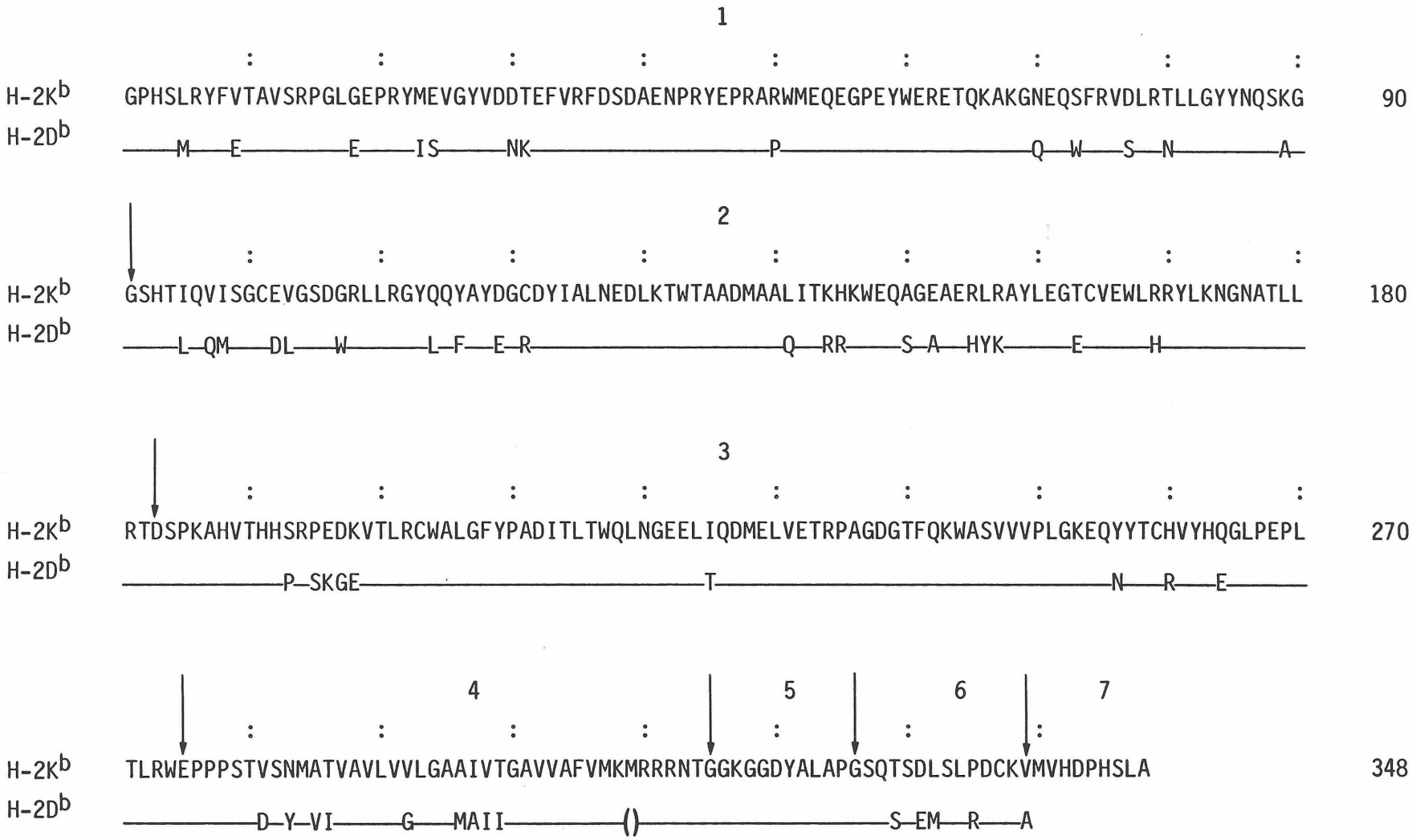
1
H-2K^b GPHSLRYFVTAVSRPGLGEPRYMEVGYVDDTEFVRFDSDAENPRYEPRARWMEQEGPEYWERETQKAKGNEQSFRVDLRTLLGYYNQSKG 90
H-2D^b ——M—E———E——IS——NK————————P————————Q—W—S—N———A—
2
H-2K^b GSHTIQVISGCEVGSDGRLLRGYQQYAYDGCDYIALNEDLKTWTAADMAALITKHKWEQAGEAERLRAYLEGTCVEWLRRYLKNGNATLL 180
H-2D^b ——L—QM—DL—W———L—F—E—R————————Q—RR——S—A—HYK——E——H———
3
H-2K^b RTDSPKAHVTHHSRPEDKVTLRCWALGFYPADITLTWQLNGEELIQDMELVETRPAGDGTFQKWASVVVPLGKEQYYTCHVYHQGLPEPL 270
H-2D^b ————P—SKGE————————T————————N—R—E———
4 5 6 7
H-2K^b TLRWEPPPSTVSNMATVAVLVVLGAAIVTGAVVAFVMKMRRRNTGGKGGDYALAPGSQTSDLSLPDCKVMVHDPHSLA 348
H-2D^b ————D—Y—VI——G—MAII———()————S—EM—R—A

Fig. 4

Chapter V

Conclusion

We have demonstrated the general applicability of the mixed oligonucleotide probe approach in the isolation of specific cloned DNA sequences. A mixture of hexadecanucleotides, representing all possible coding sequences for a portion of the protein sequence of H-2K^b, was used to obtain the H-2K^b-related cDNA clone, pH2K01 (Chapter II). This was the only clone isolated from a total of 30,000 colonies screened. Although pH2K01 does not contain the entire coding sequence of H-2K^b, its availability was crucial to the isolation of the two longer clones, pH202 (H-2K^b) and pH203 (H-2D^b).

In the mixed-probe approach described in this report, the oligonucleotide set was radiolabeled and used directly as hybridization probe to detect H-2K^b sequences. The cDNA clone coding for human β 2-microglobulin was isolated similarly (Suggs et al., 1981). Other investigators have taken a slightly different approach. A unique oligonucleotide sequence or a mixture of oligonucleotides is used to prime the synthesis of cDNA from the desired mRNA. The radiolabeled cDNA is then used to screen for the desired clones. The clones coding for human fibroblast and leukocyte interferons (Goeddel et al., 1980a,b) and the HLA-B antigen (Sood et al., 1981) were isolated in this manner.

For the detection of clones containing rare mRNA sequences, it could be more advantageous to use synthetic oligonucleotides directly as probes rather than as primers for cDNA synthesis. The simplicity of the direct approach minimizes the number of experimental manipulations necessary to isolate the specific clone. Furthermore, obtaining sufficient cDNA from low abundance mRNAs necessitates the isolation of a large amount of message.

pH2K01 is apparently derived from the H-2K^b precursor RNA but differs from the mRNA represented in pH202 due to the absence of splicing at the exon 2-intron 3 junction (amino acid residue 91) (Steinmetz et al., 1981a). The remaining 3'-end untranslated sequence of pH2K01 beyond the point where it diverges from pH202 is highly homologous to the 5'-end sequences of intron 3 in the H-2 genes isolated

by Steinmetz et al. (1981a) and Moore et al. (1982), respectively. Lack of splicing has been observed in the H-2^d cDNA clone, pH-2II (Steinmetz et al., 1981b). The sequence of pH-2II reads through the end of exon 7 into the following intron.

pH2K01 and pH2-II could have been derived from intermediate splicing products. However, it is also possible that these two clones represent mature messages which are spliced (or not spliced) in patterns different from those observed in other H-2 cDNA clones.

That different splicing patterns are employed in the production of H-2 mRNAs is supported by our observation on pH202 (H-$2K^b$) and pH203 (H-$2D^b$). Comparison of the cDNA sequence with H-2 gene sequences (Steinmetz et al., 1981a, Moore et al., 1982) suggests that there are at least two ways by which splicing at the 3'-end of the gene can occur, resulting in molecules with shorter (H-$2D^b$) or longer (H-$2K^b$) cytoplasmic domains. Whether this structural difference gives rise to distinct cytoplasmic functions remains to be seen.

The H-2D region in mouse codes for at least two gene products, D and L. In the d and q haplotypes, both D and L have been isolated (Hansen and Sachs, 1980; Hansen et al., 1981). In the b haplotype, however, no L antigen has been detected (Levy and Hansen, 1980). The precise boundaries of the D and L loci have not been determined because of the lack of appropriate recombinant strains.

The complete primary structure of the D^b molecule is known from sequencing studies of the protein (Maloy et al., 1981; Maloy and Coligan, 1982) and the cDNA clone pH203 (Chapter IV). About 75% of the DNA coding region and 476 base pairs of the 3'-untranslated region of D^b are contained in pH203.

The amino acid sequence of L^d has been determined completely except for a few residues in the C-terminus (see Maloy and Coligan, 1982). Moore et al. (1982) recently reported the DNA sequence of a genomic clone 27.5 which encodes an L^d polypeptide. The identity of the gene was established by gene transfer studies (Goodenow

et al., 1982) and comparison of the translated sequence with the known partial L^d protein sequence.

The transplantation antigen molecule can be divided into regions as defined by exons in the gene (Steinmetz et al., 1981a). These regions correlate with protein domains. Comparison of D^b and L^d shows homologies at the protein and DNA levels of 94% and 97%, respectively (see Table I). In fact, the two molecules are identical from regions 3 through 6. Antigenic differences therefore reside in the first two external regions.

A seventh region has been shown to exist in K^b but not in D^b (Chapter IV). However, there is no direct evidence for its presence or absence in L^d. The sequence of gene 27.5 shows that the gene can be spliced in either of two ways at the 3'-end and can give rise to either a K^b-like or D^b-like molecule (with respect to the presence of the seventh region). The characterization of the 3'-end of the L^d mRNA awaits the completion of the C-terminal protein sequencing, or the isolation of a cDNA clone.

Is it possible that D^b is in fact allelic to L^d and not to D^d? Only a few residues of D^d have been sequenced to date (Nairn et al., 1980, 1981). The overall homology between D^b and D^d is 85% over 179 residues available for comparison. Region I of D^b, which has been completely sequenced, is 82% homologous to D^b (compared to 92% for L^d vs. D^b).

Because of the incomplete sequence data for D^d no firm conclusions can be made. However, Coligan and co-workers have shown that D^b and L^d share common structural features. Both D^b and L^d have three carbohydrate moieties compared to only two in D^d (Nairn et al., 1980), and D^b and L^d bind β-2 microglobulin with less affinity than D^d (Maloy et al., 1980; Coligan et al., 1980). Together with the 97% DNA sequence homology between D^b and L^d, all these observations suggest that D^b and L^d are products of genes that are either allelic, or very recently duplicated.

The 3'-untranslated sequences of several H-2 clones are now available for comparison. The genomic clone pseudogene 27.1 (Steinmetz et al., 1981a) has been mapped to the Qa-2,3 region of the Tla complex. Gene 27.5 (Moore et-al., 1982) encodes an L^d polypeptide. The cDNA clones pH202 and pH203 encode K^b and D^b, respectively. For about 300 bases following the termination codon, all four sequences are highly homologous. Beyond this region, the sequences diverge. The remaining sequences of pseudogene 27.1 and pH202 are similar on the one hand, while that of gene 27.5 and pH203 are alike on the other. The region of pseudogene 27.1 which is homologous to the sequences of gene 27.5 and pH203 beyond the 300 base pair breakpoint has been transposed into the third intron of the pseudogene (Steinmetz et al., 1981a). It would be interesting to find out if the same transposition has occurred in the K^b gene.

It appears that the 3'-untranslated sequence of H-$2K^b$ is more closely related to that of a pseudogene mapping to the Qa-2,3 region, which is to the right of the D region, than to those of D-region genes. On the other hand, in other regions of the molecule, D^b, L^d and 27.1 are more homologous to each other than they are to K^b (Maloy and Coligan, 1982). For example, the first three lack a methionine in position 309. The sequence of K^b for residues 141-145 is LITKH, while for D^b, L^d and 27.1, it is QITRR. The simplest explanation for these observations is that the Qa-2,3-region pseudogene arose from a recombinational event between K and D region genes. That such an event could have occurred is not surprising in view of the 91% nucleotide sequence homology between K^b and D^b (Chapter IV).

TABLE I. % Homology of regions of D^b to L^d (Moore et al., 1982; Maloy and Coligan, 1982) and D^d (Nairn et al., 1980, 1981).

		L^d		D^d
	Region	Amino Acid	DNA	Amino Acid
	1[a]	92		82
External	2	85	91	
	3	100	100	
Transmembrane	4	100	100	
	5	100	100	
Cytoplasmic				
	6	100	100	
	3'-untranslated		98	

[a]N-terminal amino acid sequence of D^b is taken from Maloy, et al. (1981). The DNA sequence of D^b is known for only 29 out of 270 positions in Region 1.

References

Coligan, J. E., Kindt, T. J., Nairn, R., Nathenson, S. G., Sachs, D. H., and Hansen, T. H.: Primary structural studies of an H-2L molecule confirm that it is a unique gene product with homology to H-2K and H-2D antigens. Proc. Natl. Acad. Sci. USA 77: 1134-1138, 1980.

Goeddel, D. V., Shepard, H. M., Yelverton, E., Leung, D., Crea, R., Sloma, A., and Pestka, S.: Synthesis of human fibroblast interferon by E. coli. Nucleic Acids Res. 8: 4057-4074, 1980a.

Goeddel, D. V., Yelverton, E., Ullrich, A., Heyneker, H. L., Miozzari, G., Holmes, W., Seeburg, P. H., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandliss, R., Sloma, A., Tabor, J. M., Gross, M., Familletti, P. C. and Pestka, S.: Human leukocyte interferon produced by E. coli is biologically active. Nature 287: 411-416, 1980b.

Goodenow, R. S., McMillan, M., Orn, A., Nicholson, M., Davidson, N., Frelinger, J. A., and Hood, L.: Identification of a BALB/c H-2L^{d} gene by DNA-mediated gene transfer. Science 215: 677-679, 1982.

Hansen, T. H., and Sachs, D. H.: Isolation and antigenic characterization of the product of a third polymorphic H-2 locus, H-2L. J. Immunology 121: 1469-1472, 1978.

Hansen, T. H., Ozato, K., Melino, M. R., Coligan, J. E., Kindt, T. J., Jandinski, J. J., and Sachs, D. H.: Immunochemical evidence in two haplotypes for at least three D region-encoded molecules, D, L, and R. J. Immunology 126: 1713-1716, 1981.

Levy, R. B., and Hansen, T. H.: Functional studies of the products of the H-2L locus. Immunogenetics 10: 7-17, 1980.

Maloy, W. L., Hammerling, G., Nathenson, S. G. and Coligan, J. E.: Comparison of alloantisera and hybridoma antibody for purification of the H-2D^{b} murine histocompatibility antigen and preliminary molecular characterization of this antigen. J. Immunol. Met. 37: 287-299, 1980.

Maloy, W. L., Nathenson, S. G., and Coligan, J. E.: Primary structure of murine major histocompatibility complex alloantigens. J. Biol. Chem. 256: 2863-2872, 1981.

Maloy, W. L., and Coligan, J. E.: Primary structure of the H-2D^b alloantigen: Additional amino acid sequence information, localization of a third site of glycosylation and evidence for K and D region specific sequences. Immunogenetics, in press, 1982.

Moore, K. W., Taylor, B. T., Sun, Y. H., Eakle, K. A., and Hood, L.: DNA sequence of a gene encoding a BALB/c mouse L^d transplantation antigen. Science 215: 679-682, 1982.

Nairn, R., Nathenson, S. G., and Coligan, J. E.: Isolation, characterization and amino acid sequence studies of the cyanogen bromide fragments of the H-2D^d glycoprotein. Eur. J. Immunology 10: 495-503, 1980.

Nairn, R., Nathenson, S. G., and Coligan, J. E.: Amino acid sequence of cyanogen bromide fragment CN-C (residues 29-48) of the mouse histocompatibility antigen H-2D^d: A comparison of the amino terminal 100 residues of H-2D^d, D^b, K^d and K^b reveals discrete areas of diversity. Biochemistry 20: 4739-4745, 1981.

Sood, A. K., Pereira, D., and Weissman, S. M.: Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligonucleotide primer. Proc. Natl. Acad. Sci. USA 78: 616-620, 1981.

Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F.-W., Boyse, E. A., and Hood, L.: A pseudogene homologous to mouse transplantation antigens: Transplantation antigens are encoded by eight exons that correlate with protein domains. Cell 25: 683-692, 1981a.

Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathenson, S., and Hood, L.: Three cDNA clones encoding mouse transplantation antigens: Homology to immunoglobulin genes. Cell 24: 125-134, 1981b.

Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., and Itakura, K.: Use of synthetic oligonucleotides as hybridization probes. III. Isolation of cloned cDNA sequences for human β 2-microglobulin. Proc. Natl. Acad. Sci. USA 78: 6613-6617, 1981.

Appendix I

A Restriction Map of Rat Ribosomal DNA

Antonio A. Reyes, R. Bruce Wallace and James Bonner

Abstract

A restriction map of rat ribosomal DNA was constructed from the combined analyses of total genomic DNA and two lambda phage clones containing rat ribosomal DNA inserts. The minimum length of the repeat unit was determined as 33 kb.

Introduction

Studies on the ribosomal genes of Xenopus laevis have shown that the 18S and 28S rRNA coding regions are arranged in tandem to form a unit which occurs repeatedly in a linear array (for review, see Reference 1). The repeat unit is organized into the 18 and 28S coding regions interrupted by a large and a small spacer regions. The rRNA gene is transcribed as a 40S rRNA precursor with the 18S coding region situated near the 5'-end. A portion of the large spacer region is transcribed and forms the 5'-end, while the small spacer region is transcribed in its entirety and is found between the 18S and 28S sequences in the precursor. Most of the large spacer is untranscribed. Restriction mapping of mammalian ribosomal genes has revealed an organization similar to that of Xenopus. By the use of this technique, the length of the major repeat unit has been estimated at 38-44 kb for mouse (2,3) and 28-31 for man (2,4,5). These studies have also shown that the ribosomal repeat units are heterogeneous, as seen from the presence of several minor bands when restricted genomic DNA is hybridized to nucleic acid probes specific for the 18S or 28S sequences. The probes used in these studies were 18S or 28S rRNA, cRNA or cDNA. One limitation of this approach is that fragments which are part of the repeat unit but which consist entirely of spacer sequences cannot be easily detected. Hence, the repeat lengths predicted by this technique can only be minimum estimates.

rRNA-driven hybridization experiments have shown that the rat ribosomal genes comprise about 0.02% of the total DNA and are reiterated about 100-fold per haploid genome (6,7). A recent electron microscopic study of rat ribosomal DNA-RNA hybrids has shown an organization very similar to that in mouse and man

(6). The mean length of the repeat unit in rat was estimated to be 37.1 kb. Although no significant length heterogeneity was observed, length differences of up to 5 kb could not be ruled out as these fell within the limits of error in the measurement.

We have constructed a restriction map of the rat ribosomal genes by analyzing restriction fragments of whole genomic DNA which hybridize with 18S or 28S specific probes. The map has been confirmed by further analysis of two λCh4A clones which carry rat ribosomal DNA inserts. In addition, one of the clones has been used to probe restricted genomic DNA to identify fragments which consist entirely of spacer sequences. Two such fragments were identified, thus extending the restriction map.

We have determined the minimum length of the major repeat unit at approximately 33 kb. This is less than the 37.2 kb found from electron microscopic measurements (6). The small difference might be due to our inability to detect additional spacer sequences.

Materials and Methods

DNA, RNA and cDNA

Long rat ascites DNA (approximately 30 kb long) was obtained from J.-R. Wu. Rat 18S and 28S rRNAs were obtained from W. E. Stumph. [^{32}P]cDNA was synthesized from rRNA by the method of Berns and Jaenisch (8). Specific activities of cDNAs prepared were typically 0.5-1 x 10^8 cpm/μg. Clone λRR04 was nick-translated according to the method of Maniatis et al. (9).

Restriction Enzyme Digestions

Restriction endonucleases EcoR1, HindIII, PstI and BamHI were purchased from Boehringer-Mannheim and Bethesda Research Laboratories. Single digestions were carried out at 37°C for 1 hour in the buffer suggested by the supplier. For double digestions, the buffer composition of the initial digestion mixture was adjusted to that optimum for the second enzyme. The subsequent digestion was also at 37°C for 1 hour.

Gel Electrophoresis

Digestion mixtures were adjusted to 2% Ficoll, 0.25% SDS, 0.025% bromophenol blue before loading onto vertical gels. For horizontal gels, 10% glycerol was used instead of Ficoll. 0.5%-1% SeaKem agarose gels were run in TBE buffer (90 mM Tris, 90 mM boric acid, 0.025 mM EDTA, pH 8.3) at 2.5 V/cm for 15 hours or 10 V/cm for 4 hours. Gels were calibrated by running EcoR1 digests of phage λ DNA and HindIII digests of phage PM2 DNA. For gels of restricted genomic DNA, marker fragments were ^{32}P-labeled at the 5'-ends by polynucleotide kinase and mixed with EcoR1-digested genomic DNA. We found that marker fragments, when run by themselves, migrated slower than when mixed with genomic DNA. After electrophoresis, gel strips containing the marker-genomic DNA mixture were dried and autoradiographed.

Blots

The protocol used was essentially that described by Engel and Dodgson (10). The hybridization mixture contained 3 ng/ml of probe and was incubated at 70°C for 24 hours. Filters were exposed to pre-flashed Kodak XR-5 film with intensifying screens (Cronex Lightning Plus) for 12-24 hours at -70°C.

Screening of Rat Library

The construction of a library of rat genomic DNA fragments cloned in the EK2 vector λCh4A has been described by Sargent et al. (11). The main feature of this library is that the rat DNA inserts were generated by partial EcoR1 digestion of total genomic DNA. The library was screened for clones carrying the ribosomal genes essentially according to the method of Maniatis et al. (12). A total of 2 x 10^4 phage was screened with 18S[^{32}P]cDNA. Two positive clones, λRR03 and λRR04, were selected for further study. Phage were propagated in liquid culture as described by F. Blattner in the protocol accompanying the λCharon phage, and by Yamamoto et al. (13).

Recombinant DNA Safety

λCh4A phage carrying rat DNA fragments were propagated under P3 + EK2

containment conditions in accordance with the NIH guidelines for recombinant DNA research.

Results

Restriction Analysis of Rat Genomic DNA

Rat genomic DNA was digested with endonucleases EcoR1, BamHI, HindIII or PstI, electrophoresed on agarose gels, transferred to nitrocellulose and hybridized with 18S or 28S cDNA probe. In each case one or a few fragments hybridized strongly to the probe, allowing identification of fragments which contain the 18S or 28S gene or both. A number of faint bands were also seen, suggesting heterogeneity of the rat ribosomal repeat unit (Fig. 1).

A partial restriction map of the rat ribosomal gene was constructed from the hybridization data (Fig. 2). (The sites b1 and h1 were determined by other experiments; see below). The positions of the restriction sites along the map were further checked by double digestion experiments (Fig. 3).

Restriction Analysis of Cloned Ribosomal DNA Sequences

Two λCh4A clones containing rat ribosomal DNA inserts were isolated by screening a rat genomic library (11) with 18S cDNA probe. The rat DNA inserts in the phage library were generated by partial EcoR1 digestion of total DNA. Clones λRR03 and λRR04 were digested with EcoR1 and hybridized with 18S or 28S cDNA probe. λRR03 contains fragments r1-r2 (10 kb) and r2-r3 (6 kb), while λRR04 contains r1-r2 only (Fig. 4). Restriction analysis of the two clones using HindIII, BamHI and PstI confirmed the map obtained from genomic DNA.

Search for Spacer Fragments

λRR04 was labeled by nick-translation and hybridized to restricted genomic DNA to determine the location of restriction sites to the left of r1. It was hoped that the detection of h1-h2 and b1-b2 would yield the length of the major repeat unit (i.e., h1 = h4 and b1 = b6). However, h1-h2 and b1-b2 were shorter than expected (Fig. 5).

Fragments predicted from the restriction map to hybridize with r1-r2 were identified in the blot, specifically b2-b3, b3-b4, h2-h3 and r1-r2 itself. The other major bands in the BamH1 and HindIII lanes were assigned to fragments containing only spacer DNA, namely b1-b2 and h1-h2. In the case of EcoR1-restricted DNA, no other major fragment was expected in addition to r1-r2. However, two other bands with intensities only slightly less than that of r1-r2 were seen. These bands are 5.8 kb and 1.7 kb in length. Several faint bands were also seen in all three lanes.

Discussion

We have combined the analyses of total genomic DNA and λCh4A clones carrying rat ribosomal DNA inserts to construct a restriction map of rat ribosomal DNA. We have also used one of the clones to probe for restriction fragments which are part of the repeat unit but do not contain 18S or 28S coding sequences. The results of these experiments suggest that the minimum length of the repeat unit is about 33 kb, corresponding to p1-r4. This is smaller than the 37.2 kb measured from electron microscopic studies (6).

It is interesting to note that sites corresponding to b3, p2, r2 and h3 in the 18S gene, and b4, b5 and r3 in the 28S gene are also present in the same positions relative to each other in mouse and man (2-5). In particular, Cory and Adams (3) have estimated that r2-h2 is 0.1 kb in mouse. In rat, the same fragment is 0.1-0.2 kb long. These findings suggest that the ribosomal coding sequences have been strictly conserved during evolution.

Hybridization of genomic DNA to 18S cDNA, 28S cDNA or fragment r1-r2 reveals the presence of several minor bands. This suggests the existence of other families of repeat units which are structurally different from the major ribosomal repeat family. It has been hypothesized that length heterogeneity of the ribosomal repeat unit is caused by variations in the length of the non-transcribed spacer, specifically

by the variable number of repetitive simple sequences (15). The presence of simple DNA repeat sequences in the non-transcribed spacer of Xenopus (16,17) and mouse (18,19) has been reported. Whether an analogous situation exists in the rat remains to be seen. It is interesting to note that the results shown in Fig. 5 suggest that r1-r2, which contains a portion of the non-transcribed spacer, is homologous with other sequences found elsewhere in the genome, though not necessarily in the rRNA repeat unit.

Another possible explanation for heterogeneity is the occurrence of base changes along the gene which results in the presence of specific restriction sequences in some but not in other repeat families. One could speculate that such mutations are more likely to be found in the non-transcribed spacer region of the repeat unit, where there is less pressure to maintain specific DNA sequences.

References

1. Federoff, N. V.: Cell 16: 697-710, 1979.
2. Arnheim, N., and Southern, E. M.: Cell 11: 363-370, 1977.
3. Cory, S., and Adams, J. M.: Cell 11: 795-805, 1977.
4. Wilson, G. N., Hollar, B. A., Waterson, J. R., and Schimkel, R. D.: Proc. Natl. Acad. Sci. USA 75: 3567-3571, 1978.
5. Krystal, M., and Arnheim, N.: J. Mol. Biol. 126: 91-104, 1978.
6. Stumph, W. E., Wu, J.-R., and Bonner, J.: Biochemistry 18: 2864-2871, 1979.
7. Holmes, D. S., and Bonner, J.: Biochemistry 13: 841-848, 1974.
8. Berns, A., and Jaenisch, R.: Proc. Natl. Acad. Sci. USA 73: 2448-2452, 1976.
9. Maniatis, T., Jeffrey, A., and Kleid, D. G.: Proc. Natl. Acad. Sci. USA 72: 1184-1188, 1975.
10. Engel, J. D., and Dodgson, J. B.: J. Biol. Chem. 253: 8239-8246, 1978.
11. Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., and Bonner, J.: Proc. Natl. Acad. Sci. USA 76: 3256-3260, 1979.
12. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A.: Cell 15: 687-701, 1978.
13. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., and Treiber, C.: Virology 40: 734-744, 1970.
14. Southern, E. M.: J. Mol. Biol. 98: 503-517, 1975.
15. Wellauer, P. K., Dawid, J. B., Brown, D. D., and Reeder, R. H.: J. Mol. Biol. 105: 461-486, 1976.
16. Boseley, P., Moss, T., Machler, M., Portmann, R., and Birnstiel, M. L.: Cell 17: 19-31, 1979.
17. Botchan, P., Reeder, R. H., and Dawid, J. B.: Cell 11: 599-607, 1977.
18. Grummt, I., and Gross, H. J.: Molec. gen. Genet. 177: 223-229, 1980.
19. Arnheim, N., and Kuhn, M.: J. Mol. Biol. 134: 743-765, 1979.

Fig. 1. Restriction pattern of rat ribosomal genes. Restricted genomic DNA (10 μg) was fractionated on a 0.5% agarose gel. After transfer to nitrocellulose, fragments were hybridized to 18S or 28S[^{32}P]cDNA. In this and following figures, R indicates EcoR1; H, HindIII; B, BamHI and P, PstI. The positions of some molecular weight markers are shown.

Fig. 2. Restriction map of rat ribosomal DNA. The boxes represent the 18S and 28S genes. The rat DNA inserts in λRR03 and λRR04 are also shown.

Fig. 3. Double digestion patterns of rat ribosomal genes. Restricted genomic DNA (10 μg) was fractionated on 1% agarose gels, transferred to nitrocellulose and hybridized to 18S or 28S[^{32}P]cDNA probes.

Fig. 4. Restriction pattern of λRR03 and λRR04. Phage DNA (200 ng) was digested with restriction enzymes and fractionated on 1% agarose gels. Probes were [^{32}P] 18S and 28S cDNA.

Fig. 5. Hybridization of fragment r1-r2 to restricted genomic DNA. Restricted genomic DNA (10 μg) was fractionated on a 0.5% agarose gel. Fragments were hybridized to [^{32}P]λRR04.

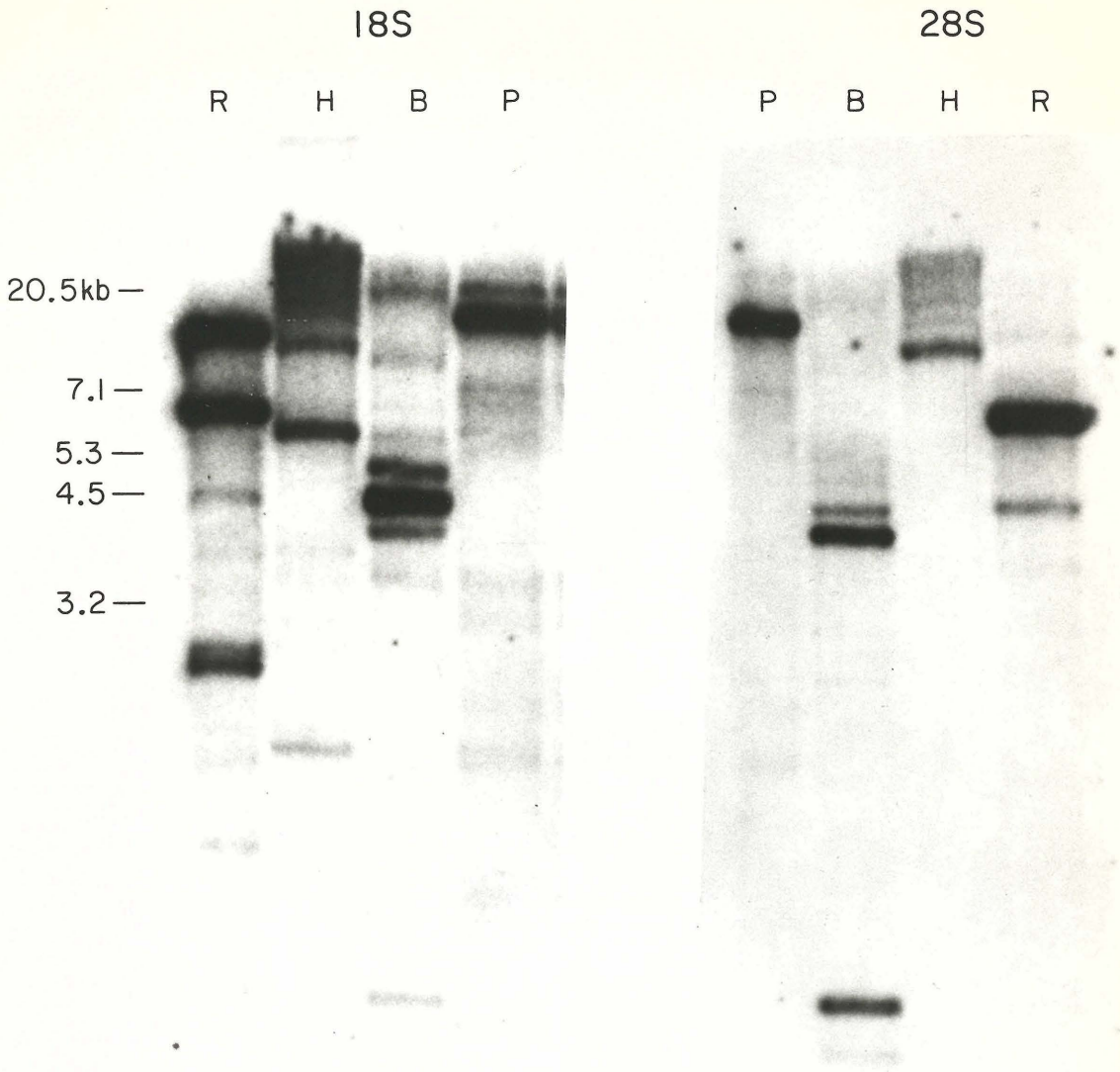

Fig.1

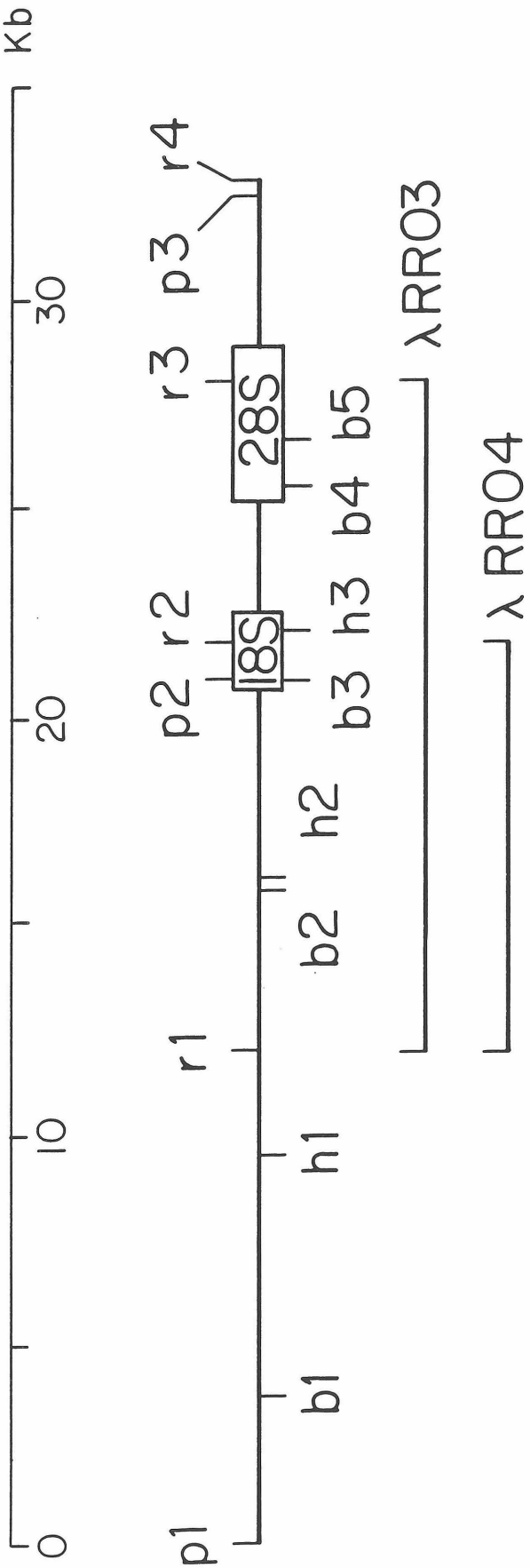
Kb
0
10
20
30
p1
r1
p2
r2
18S
r3
28S
p3
r4
b1
h1
b2
h2
b3
h3
b4
b5
λRRO3
λ RRO4

Fig. 2

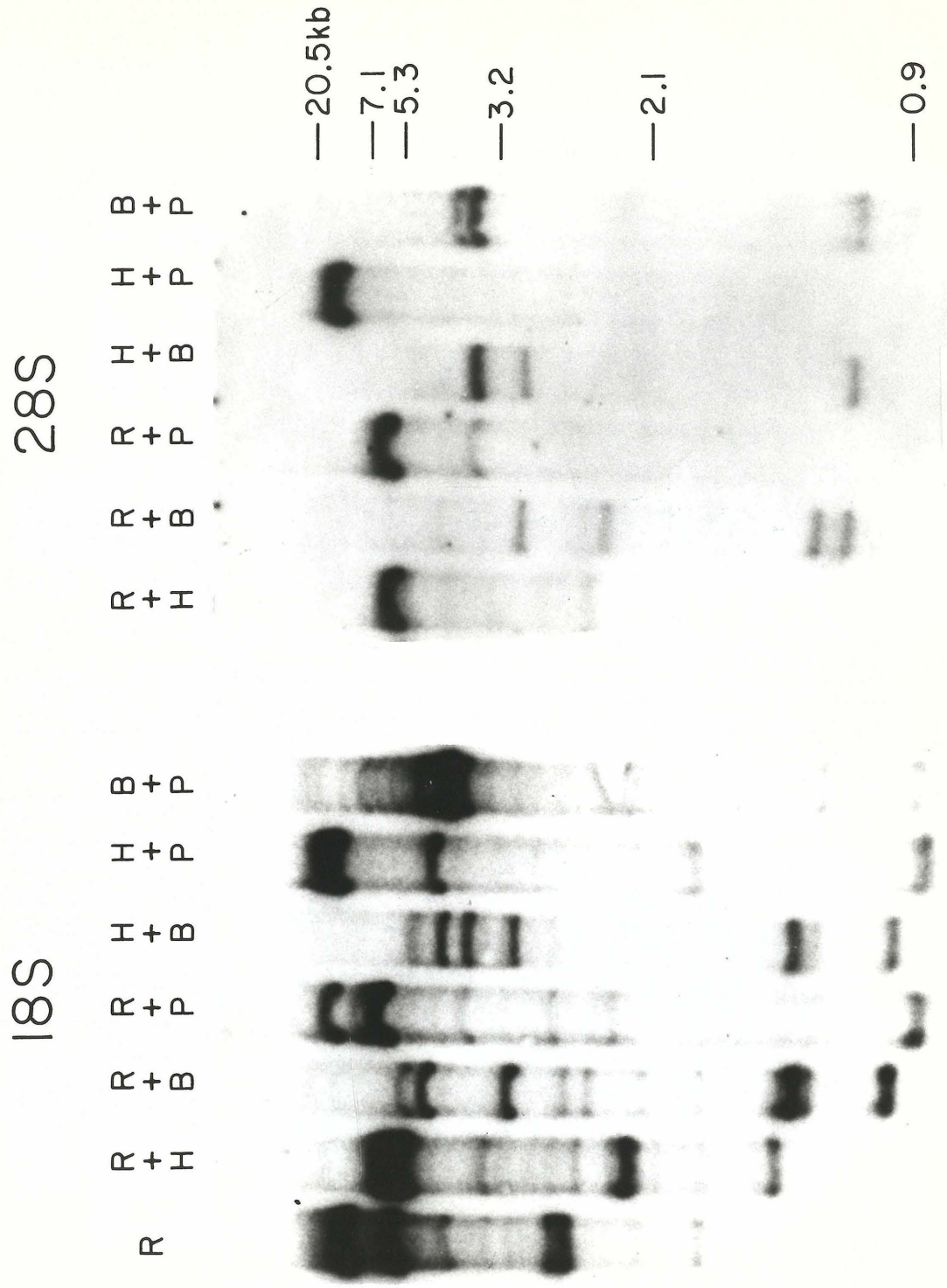

Fig. 3

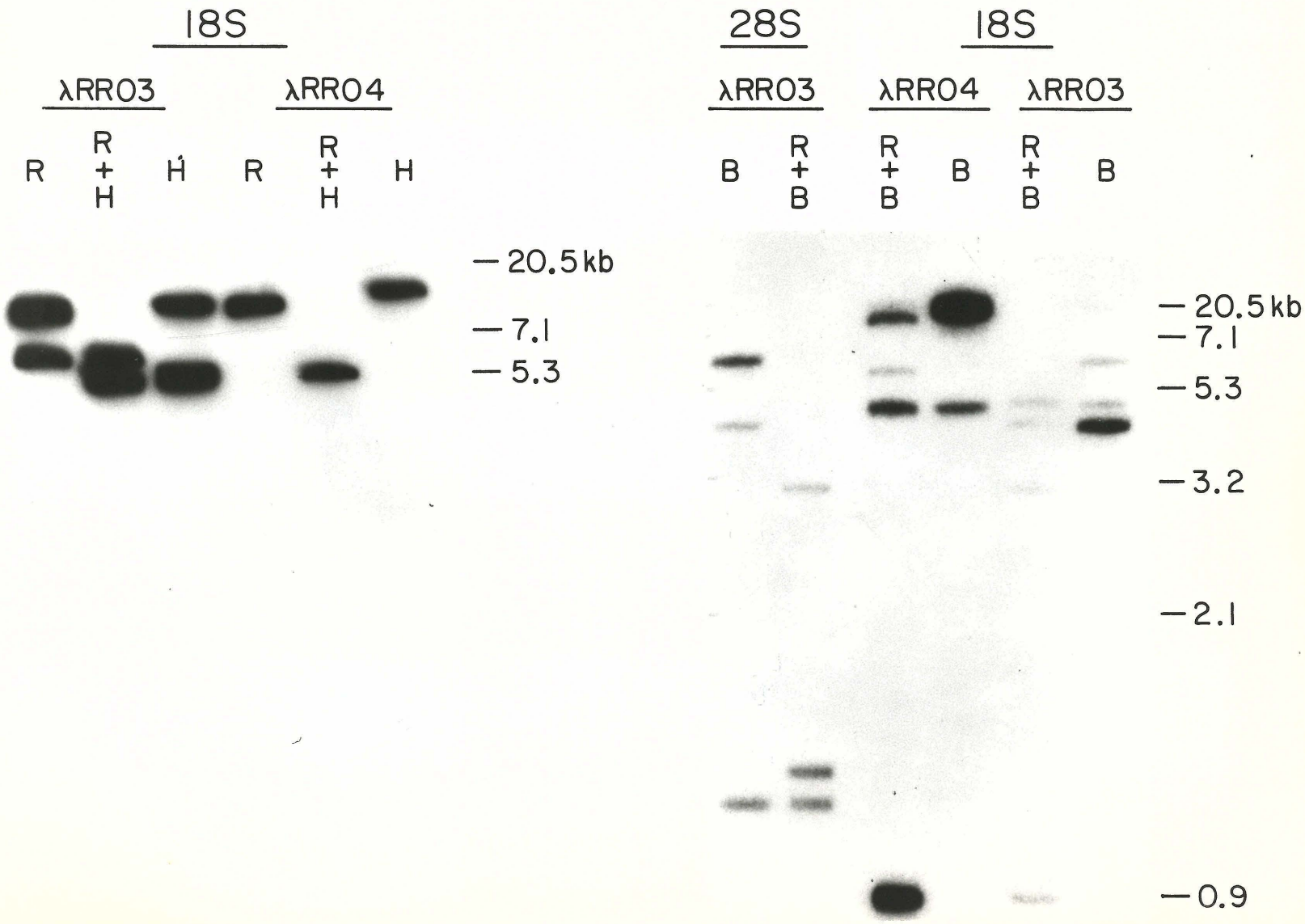
18S
λRR03
λRR04
R
R + H
H
R
R + H
H
— 20.5 kb
— 7.1
— 5.3
28S
18S
λRR03
λRR04
λRR03
B
R + B
R + B
B
R + B
B
— 20.5 kb
— 7.1
— 5.3
— 3.2
— 2.1
— 0.9

Fig. 4

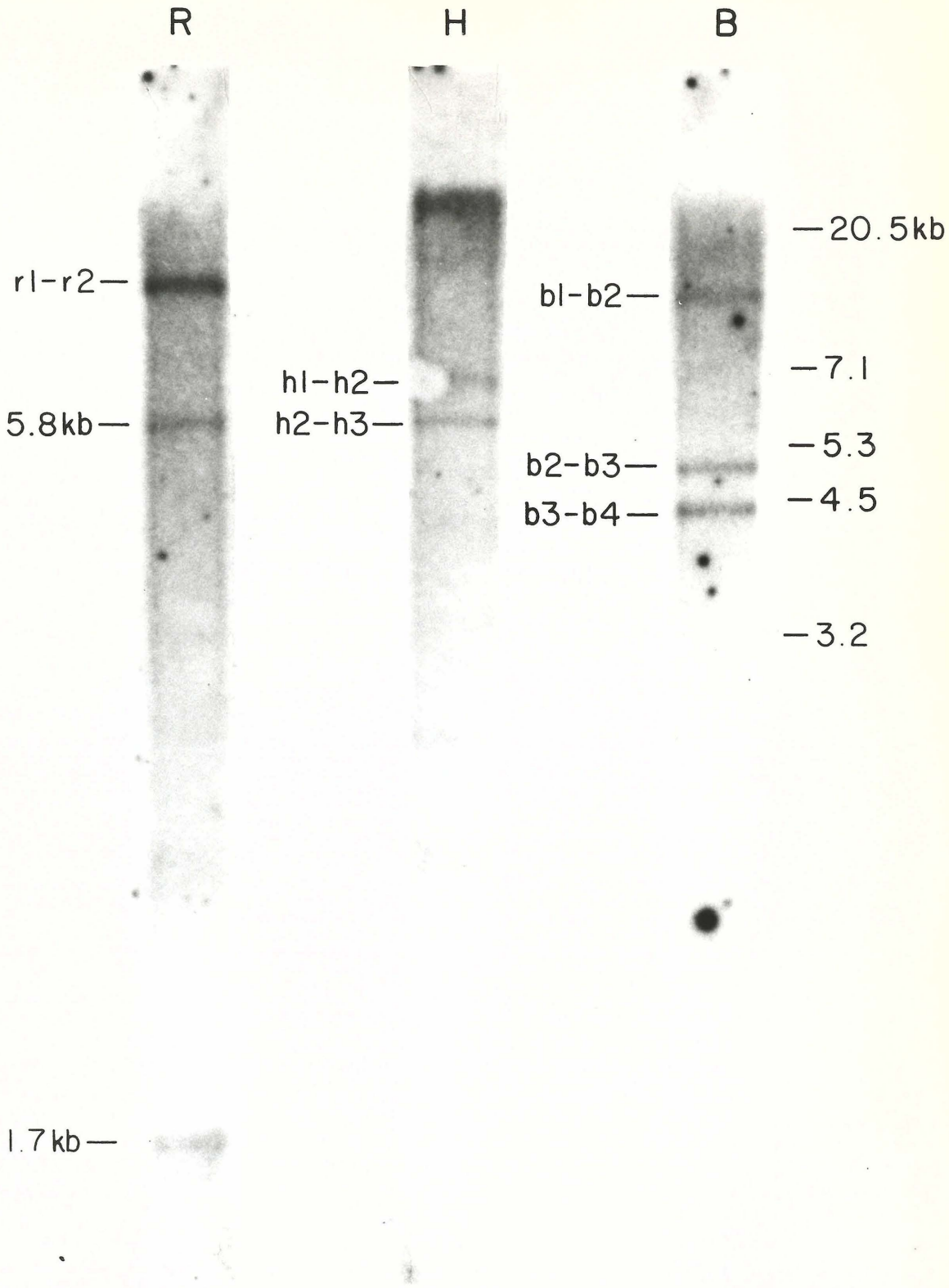

Fig. 5

Appendix II

Proc. Natl. Acad. Sci. USA
Vol. 76, No. 7, pp. 3256–3260, July 1979
Biochemistry

The rat serum albumin gene: Analysis of cloned sequences

(genome libraries/cDNA clones/R-loops/Southern blots)

THOMAS D. SARGENT, JUNG-RUNG WU, JOSE M. SALA-TREPAT*, R. BRUCE WALLACE†, ANTONIO A. REYES, AND JAMES BONNER

Division of Biology, California Institute of Technology, Pasadena, California 91125

Contributed by James F. Bonner, April 27, 1979

ABSTRACT **The rat serum albumin gene has been isolated from a recombinant library containing the entire rat genome cloned in the λ phage Charon 4A. Preliminary R-loop and restriction analysis has revealed that this gene is split into at least 14 fragments (exons) by 13 intervening sequences (introns), and that it occupies a minimum of 14.5 kilobases of genomic DNA.**

Recent advances in recombinant DNA technology have made it possible to obtain virtually any desired single-copy genomic sequence in cloned form, provided an appropriate probe is available. We have used these techniques to isolate the rat serum albumin gene. Serum albumin synthesis is one of the major characteristics of vertebrate liver. Observation of the activity and state of this gene during development and in adult tissues should be informative as to the process of terminal differentiation. Albumin synthesis is essentially constitutive, but does respond significantly to a variety of stimuli (1). It is also expressed to variable extents in different hepatoma cell lines (2). The availability of cloned albumin genomic DNA will greatly facilitate the study of this variable expression, particularly at the level of transcript processing.

Determination of the sequence organization of the albumin gene is also of interest, especially with regard to the disposition of repetitive elements and intervening sequences. Although regulatory (3) and evolutionary (4) significance has been postulated, the functional role, if any, of these striking features of eukaryotic genomes remains unknown. The comparative studies that will be possible as other genes are extracted from the rat and related species can be expected to provide considerable insight into this fascinating problem.

MATERIALS AND METHODS

Rat Genome Library. High molecular weight liver DNA was extracted from an adult male Sprague–Dawley rat (Simonsen Labs, Gilroy, CA) by the method of Blin and Stafford (5) and aliquots were digested with *Eco*RI (Boehringer Mannheim) under conditions adjusted to cleave either one-third or one-fifth of the *Eco*RI sites in an equivalent amount of bacteriophage λ DNA. The fragments resulting from this partial digestion were sedimented through a 10–30% sucrose gradient; the material between 10 and 20 kilobases (kb) was recovered by ethanol precipitation. A sample of this rat DNA (2.5 μg) was ligated with 8.5 μg of a preparation of Charon 4A "cloning fragments" (6, 7). This recombinant DNA was packaged *in vitro* by using extracts from defective λ lysogens provided by N. Sternberg (6). The method used was that of Hohn and Murray (8). Approximately 2,000,000 independent clones were obtained. The library was amplified 100,000-fold by subconfluent plating on *Escherichia coli* strain DP50SupF (9).

cDNA Clones. cDNA was synthesized from purified albumin mRNA as described (10). This cDNA contained a small amount of full-length material and had a number average size of approximately 1000 nucleotides. It was rendered double-stranded by sequential treatment with *E. coli* DNA polymerase I and S1 nuclease (11). The resulting DNA had a number average size of 600 nucleotide pairs. An average of 10 dCMP residues were polymerized per 3′ end by terminal transferase (ref. 12; W. Rowekamp and R. Firtel, personal communication). Forty nanograms of the tailed albumin cDNA was mixed with 200 ng of pBR322 DNA that had been cleaved with *Pst* I and similarly polymerized with dGMP residues at the 3′ ends (a gift of W. Rowekamp) and induced to cocircularize by incubation at 42°C for 4 hr followed by 16 hr of slow cooling to 4°C. *E. coli* strain χ1776 was transformed with this mixture (13, 14). Several hundred colonies were obtained on appropriate selective media. These were screened by the filter colony hybridization method of Grunstein and Hogness (15), with ^{32}P-labeled albumin cDNA as a probe. The most intensely reacting clones were selected and plasmid DNA was prepared. Proof of their identity was obtained by comparison of partial nucleotide sequences to the existing amino acid sequence of rat serum albumin (ref. 16; T. D. Sargent and J. Posakony, unpublished data).

Screening. Approximately 1,000,000 plaques from the rat library were screened by a modification of the method of Benton and Davis (17), using as a probe nick-translated albumin cDNA clones. Nine different genome clones were obtained, three of which, λRSA14, λRSA30, and λRSA40, are the subject of the present report.

R-Loops. Fifty nanograms of recombinant phage DNA that had been digested with either *Eco*RI or *Hin*dIII was mixed with 40 ng of purified albumin mRNA in a total volume of 20 μl of 70% recrystallized formamide/0.4 M NaCl/5 mM EDTA/80 mM Pipes (1,4-piperazinediethanesulfonic acid), pH 7.4 (18) and incubated at 50–53°C for 24 hr. The hybrids were spread for electron microscopy by the modified Kleinschmidt method (19). The grids were rotary shadowed with platinum and palladium (80/20) and viewed in a Phillips 300 electron microscope. DNA contour lengths were measured with a Hewlett–Packard digitizer.

Blots. Rat liver DNA from the same preparation used to make the genomic library was digested three times with a 6-fold excess of either *Eco*RI or *Hin*dIII, extracted with phenol, and precipitated with ethanol. Ten micrograms of the digested DNA was fractionated on an 8-mm-thick 0.8% agarose slab gel buffered with 50 mM Tris/18 mM NaCl/2 mM EDTA/20 mM sodium acetate, pH 7.4. Electrophoresis was at 2.5 V/cm for 16 hr. Clone DNA was digested once with a 10-fold excess of restriction endonuclease and 0.2 μg was electrophoresed on a 2-mm-thick gel. Transfer to nitrocellulose (Millipore) was es-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase.

* Present address: Laboratoire d'Enzymologie, Centre National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France.
† Present address: Department of Biology, City of Hope National Medical Center, Duarte, CA 91010.

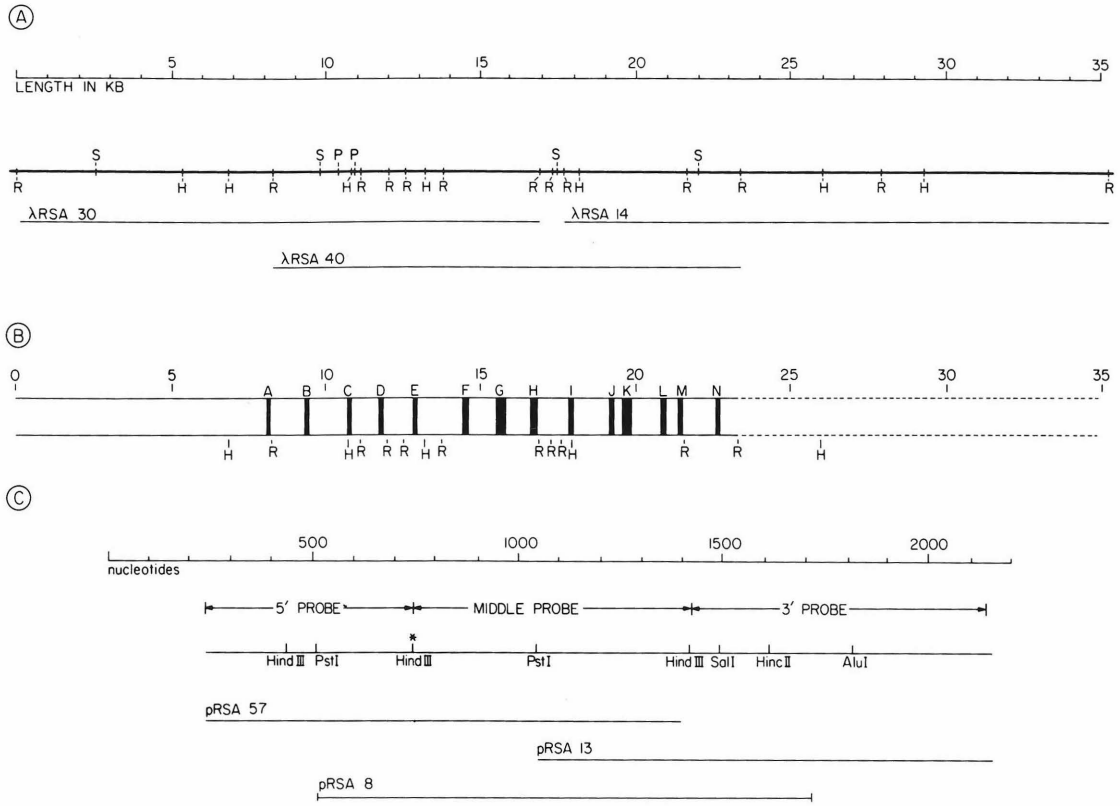

FIG. 1. (*A*) Restriction site map for three albumin genome clones, λRSA30, λRSA40, and λRSA14. R, *Eco*RI; H, *Hin*dIII, S, *Sac* I; P, *Pst* I. Molecular weights were estimated by comparison with phage PM2 DNA digested with *Hin*dIII. (*B*) Intron/exon map of the albumin gene deduced from R-loops. Black bars are exons; white bars are introns. Dashed line indicates region that does not react with albumin cDNA and was not tested for R-loop formation. Scale same as in *A*. (*C*) Restriction site map for three albumin cDNA clones, pRSA57, pRSA8, and pRSA13. Asterisk indicates *Hin*dIII site not found in genomic DNA (see text).

sentially as described by Southern (20). Washing after hybridization was done at 62–63°C with a descending series of salt concentrations from 1.0 to 0.1 M NaCl in Denhardt's solution (21).

RESULTS

The recombinant DNA methodology that we have used involves considerable manipulation of DNA, including ligation of a mixture of restriction fragments and several rounds of replication in the bacterial host. This creates the potential for two particularly serious artifactual modifications of genomic sequences: ligation of noncontiguous restriction fragments and genetic rearrangement during propagation. The partial restriction library approach used in this study provides an effective mechanism for detecting the former artifact—independently generated clones confirm the legitimacy of the restriction map of shared DNA since the probability of any given spurious ligation event occurring more than once in the production of a recombinant genome library is negligible. Fig. 1*A* shows the map of restriction sites for *Eco*RI, *Hin*dIII, and *Sac* I. The region included in clone λRSA40 is confirmed by the maps of λRSA14 and λRSA30.

Verification of nonoverlapping regions and exclusion of gross genetic artifacts such as deletions or rearrangements depend upon comparisons made between the cloned sequences and the rat genome, which are accomplished by use of "Southern blots" of genomic and cloned rat DNA. Fig. 2 shows the pattern obtained when rat DNA is digested with *Eco*RI (lane A) or *Hin*dIII (lane B), fractionated by electrophoresis, transferred to nitrocellulose, and driven by cloned albumin cDNA labeled with ^{32}P by nick translation (22). The *Eco*RI digestion results in seven bands complementary to the albumin probe. Band d is quite faint, but is clearly visible in the original and in blots probed with albumin cDNA (unpublished data). This is presumably due to the absence of some 3′-terminal mRNA sequence from the cDNA clones. A mixture of two different plasmid clones, pRSA13 and pRSA57, which includes approximately 85% of the albumin mRNA sequence complexity, is used as a probe in these experiments (unpublished data). The restriction site map for these cDNA clones is shown in Fig. 1*C*. Since there are no *Eco*RI sites present in the probe sequence, the genome blot suggests that either the gene exists in multiple divergent copies or is interrupted by sequences not present in the mRNA or conceivably both. Evidence will be presented that shows that the albumin gene is in fact interrupted. Fig. 2, lane B, shows the result of a similar experiment with rat DNA digested with *Hin*dIII. A total of seven bands can be visualized. Four of these, labeled a, b, c, and d, are consistent with the restriction map obtained from the genome clones, as are all but the largest band in Fig. 2, lane A. The remaining bands, indicated by asterisks, are unexpected. These anomalous bands appear with variable intensity in different experiments and are

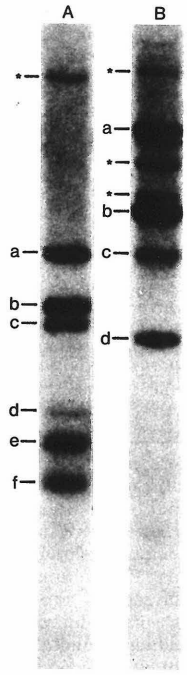

FIG. 2. Rat genome blots. *Eco*RI-digested (lane A) and *Hin*dIII-digested (lane B) rat liver DNA. Electrophoresis, filter transfer, and hybridization conditions are given in *Materials and Methods*. The sizes of the material in the lettered bands (in kb) are as follows. Lane A: a = 3.9, b = 3.1, c = 2.7, d = 1.6, e = 1.3, and f = 1.0 Lane B: a = 7.9, b = 4.9, c = 3.9, and d =2.5. Bands marked by asterisks are not present in the genome clones.

probably due to partial digestion. However, their appearance also is consistent with the hypothesis that there are multiple, slightly divergent albumin genes in the rat.

Fig. 3 illustrates analogous experiments performed with DNA from the various genomic clones. In these sets of blots the albumin cDNA probe has been cleaved with *Hin*dIII and electrophoretically fractionated into a 5′, "middle," and 3′ probe (Fig. 1C). The patterns generated are entirely consistent with that seen with whole rat liver DNA. The sizes are given in the legend to Fig. 2. This suggests that there has been little or no disruption of individual restriction fragments during the cloning procedure although the resolution limit of the genome blots is probably about 100 nucleotides and small modifications or point mutations would most likely be undetected. A curious aspect of these results is revealed in Fig. 3, gels F and G. Both the 5′ and middle albumin mRNA probes react with the same *Hin*dIII fragment (band d), which implies that the *Hin*dIII site separating these two probes (indicated by an asterisk in Fig. 1C) does not exist in the genome. The trivial explanation of probe cross contamination is unlikely because there is no analogous reaction with *Eco*RI-digested λRSA40 (Fig. 3, gels B–D). Sequence rearrangement is not responsible because the phenomenon is observed with two different genome clones, λRSA30 and λRSA40, and two different cDNA clones, pRSA13 and pRSA8 (data not shown). Furthermore, this unexpected pattern is seen when 5′ and middle albumin probes are used to drive *Hin*dIII genome blots (unpublished data). Polymorphism cannot be excluded, but seems unlikely because Sprague–Dawley rats are highly inbred. It is conceivable that this restriction site is created in the mRNA by the fusion of two adjacent exons. Final resolution of this question must await determination of the nucleotide sequence of the relevant genomic DNA.

The pattern of hybridizations with the various probes shown in Fig. 3 also establishes colinearity of the genomic and mRNA sequences. The polarity of the cDNA clones was established by

Table 1. Intron and exon lengths from R-loops formed with albumin mRNA and albumin genome clone restriction fragments

Exon	Length (mean ± SD)	*n*	Intron	Length (mean ± SD)	*n*
A	102 ± 29	12	AB	927 ± 64	6
B	111 ± 31	11	BC	1370 ± 37	21
C	95 ± 35	10	CD	946 ± 145	10
D	148 ± 34	10	DE	978 ± 42	10
E	108 ± 32	12	EF	1458 ± 88	19
F	163 ± 34	13	FG	920 ± 65	13
G	245 ± 39	15	GH	778 ± 66	14
H	189 ± 53	13	HI	908 ± 62	10
I	133 ± 26	10	IJ	1161 ± 90	11
J	131 ± 27	16	JK	297 ± 88	11
K	239 ± 31	14	KL	1011 ± 74	16
L	146 ± 28	13	LM	434 ± 94	13
M	108 ± 29	14	MN	1054 ± 163	13
N	125 ± 58	10			

Sizes in nucleotide pairs of exons and introns in the albumin genomic clones. Introns are designated by two letters, indicating adjacent exons. Lengths were determined by comparison with simian virus 40 DNA contour lengths, as measured from electron micrographs. These data were used to construct the schematic diagram shown in Fig. 1*B*. *n*, Number of samples measured for each intron or exon.

nucleotide sequence determination (T. Sargent and J. Posakony, unpublished data) and, thus, the direction of transcription, from left to right in Fig. 1, can be inferred. These blot hybridization data, in conjunction with the restriction maps, suggest that within the cloned locus there exists only one albumin mRNA sequence which is interrupted at least five times and is dispersed over a minimum of 14 kb.

To verify these observations and improve the resolution beyond the limits of a cursory restriction analysis, we digested DNA from the genome clones with either *Eco*RI or *Hin*dIII, mixed it with purified albumin mRNA, and incubated the mixture under conditions suitable for "R-loop" formation (18). The hybrid structures that formed were then visualized by electron microscopy. The result of these experiments was the identification of a total of 13 interruptions in the albumin mRNA sequence. Selected electron micrographs along with interpretive drawings of the R-loops are shown in Fig. 4. The various hybrids that formed could be identified by their contour lengths, determined by comparison with simian virus 40 DNA spread on the same grids. This permits ordering of R-loop structures by consulting the genome clone restriction maps, but does not specify the correct orientation of the fragments. In most cases the polarity can be deduced from the *Hin*dIII fragment R-loops by identification of the molecules containing the left or partial right arms of the Charon 4A vector. In one case, shown in Fig. 4, second from left, it was necessary to assume that the end lacking a visible DNA "branch" corresponds to the *Hin*dIII site located in the mRNA sequence, that is, within exon C. If this assumption is incorrect, then the albumin gene is subdivided into 15 rather than 14 exons, and the order of exons C, D, and E is reversed. The results obtained from *Eco*RI-digested clones are more difficult to interpret due to the larger number of fragments and and their similar sizes. However, the structures visualized are consistent with the *Hin*dIII-digested R-loops. Certain hybrids do not form efficiently, apparently due to close proximity to a restriction cleavage, and these tend to appear with one but not the other digestion. Also, the temperature optima for hybrid formation with the various fragments differed over a few degrees and had to be adjusted accordingly. The data from which the exon/intron map was deduced are summarized in Table 1. These values are in some

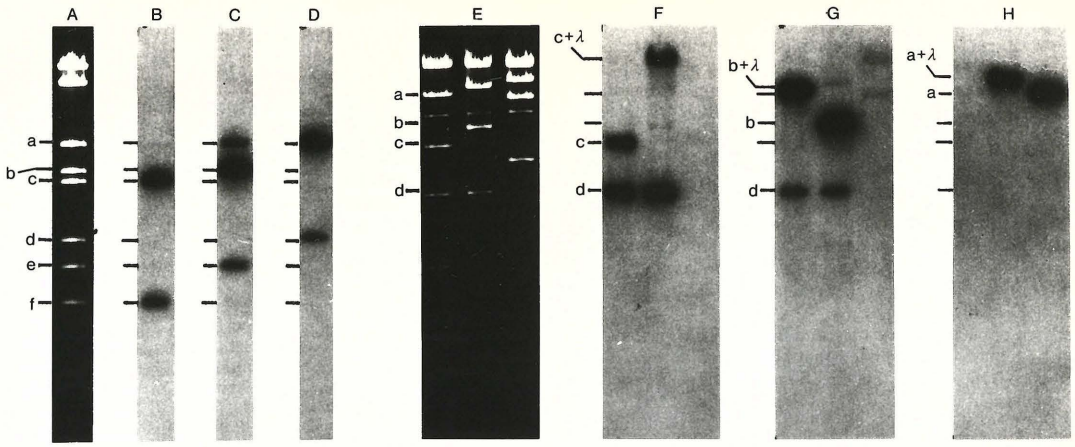

FIG. 3. Albumin genome clone blots. Gel A, photograph of gel of *Eco*RI-digested λRSA40 DNA. Gels B–D, triplicate blots of DNA shown in gel A probed with 5′, middle, and 3′ fragments of albumin cDNA clones, respectively (see Fig. 1*C*). Gel E, photograph of gel of *Hin*dIII-digested λRSA30 (left lane), λRSA40 (middle lane), and λRSA14 (right lane). Gels F–H, triplicate blots of DNA shown in gel E hybridized with 5′ (gel F), middle (gel G), and 3′ (gel H) probes as above. The sizes of the material in the lettered bands are the same as that given for the corresponding bands in Fig. 2. The order of fragments in the map (Fig. 1*A*) is as follows, from left to right: *Eco*RI, c, f, e, b, a, d; *Hin*dIII, c, d, b, a.

cases the sum of two measurements which terminate on opposite sides of the same restriction site. All of the structures represented by the diagram in Fig. 1*B* are supported by a minimum of 10 measurements made from different unambiguous examples. However, it is important to bear in mind the uncertainty inherent in data of this type. Aside from possible interpretive errors, small exons and introns might not be detectable by electron microscopy. Furthermore, very small hybrids might not be stable under the conditions used for hybridization or spreading of the DNA. DNA displacement loops alongside exons were not visible in most cases. Presumably this is due to the collapse of single-stranded DNA, which may be a property of smaller R-loops prepared by the methods we have used (23).

Close examination of the RNA involved in the R-loop structures reveals a continuous translocation of duplex regions from the 5′ end to the 3′ end of the mRNA, supporting the conclusion drawn from the blot experiments that there is only one albumin gene in the cloned complex and that it is colinear with the mRNA sequence, although interrupted. Another interesting observation is that exons A and N seem very near the termini of the mRNA. Very little, if any, RNA extends beyond the 5′ side of exon A. This implies that unless a tiny "leader" exon is separated from the rest of the gene by a huge intron or there is more than 8 kb of RNA cleaved from the 5′ end of a primary transcript, the initiation of transcription must be located somewhere on the 5′ end of clone λRSA30. Similarly, the small "whisker" of RNA visible on the 3′ side of exon N is probably mostly poly(A) (10), suggesting that the albumin mRNA coding sequence terminates at this point in the genome. Neither of the *Eco*RI fragments located 3′ to this terminal exon reacts with albumin cDNA (unpublished data). Since these fragments represent over 10 kb of genomic DNA, it is unlikely that there are any albumin exons not contained on the genomic clones.

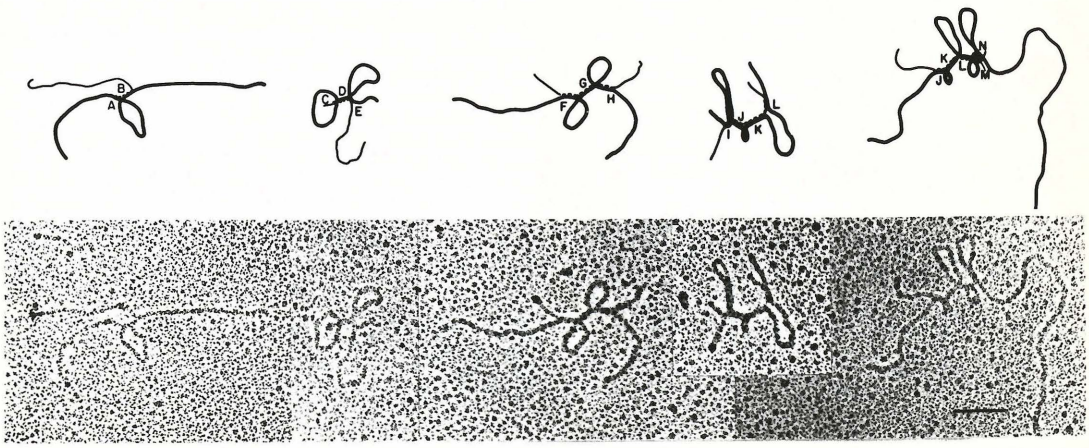

FIG. 4. Selected electron micrographs along with interpretive drawings of R-loops. The restriction fragments involved in the R-loops are, from left to right, *Hin*dIII fragments c, d, and b, *Eco*RI fragment a, and *Hin*dIII fragment a + λ (from λRSA40). In the tracings at the top of the figure, heavy solid lines indicate DNA and thin lines represent albumin mRNA. Bar at bottom right represents 500 nucleotide pairs.

DISCUSSION

The isolation of the albumin genomic sequences in cloned form makes it possible to determine the structure of the albumin gene—or, more properly, its DNA component—to the nucleotide level of resolution if desired. However, it is important to recall that the cloned gene is no longer functional rat DNA but nonfunctional bacteriophage DNA. The possibility that rearrangements might occur is not trivial and could lead to major misinterpretation. We have argued that the results shown are not artifactual because they were derived from independently generated clones and because the patterns generated by two different restriction enzymes are very similar when either clone or genomic DNA is digested and analyzed by the Southern blot technique.

Based upon the R-loop map and the various blots shown, it can be tentatively concluded that all of the albumin coding sequence is included within the 35 kb of contiguous cloned rat DNA and that it is divided into at least 14 segments or "exons" by 13 intervening sequences or "introns." The possibility remains that additional exons or introns exist that are too small to be easily detectable by electron microscopy. The possibility of multiple variants of the albumin gene within the individual rat used to generate the library has not been excluded. However, we have isolated a total of nine albumin clones from the library and, with the exception of one obvious ligation artifact, all contain linear permutations of the same rat DNA restriction fragments. It seems unlikely that other, different albumin genes would have escaped inclusion in the rat library or detection by plaque hybridization.

There seems to be no simple pattern to the size or position of the exons in this gene. The interruptions occur along the entire length of the mRNA sequence. The middle *Hin*dIII site in the mRNA has been aligned with amino acid number 181 by nucleotide sequence determination (T. Sargent and J. Posakony, unpublished data). This would mean that approximately 200–300 nucleotides of untranslated RNA must reside at the 3′ end of the albumin mRNA (Fig. 1*C*). If exon 14 does in fact contain the 3′ end of the mRNA, then one and possibly two introns may be located within the untranslated sequence (Table 1).

Assuming that we have not inadvertently cloned an inactive variant of the albumin gene, the arrangement of introns and exons has obvious ramifications regarding the processing of the albumin mRNA precursors. First, the transcription unit is at least 14.5 kb in length and, by analogy to other systems (24, 25), the primary transcript can be expected to be of the same size. This is 3 times larger than the value reported by Shafritz and coworkers (26), suggesting that the 26S species they detected represents accumulated processing intermediates. Second, there should be a minimum of 13 different nuclear intermediates that have been processed to some extent. If multiple pathways exist, then there could be many more species.

We thank Merrie Jo Johnson, Marie Carter, and Leila Gonzalez for excellent technical assistance, and Mark Davis, Doug Engel, Tom Maniatis, and James Posakony for their advice and comments. This work was supported in part by U.S. Public Health Service Grants GM 13262 and GM 20927. T.D.S. was supported by a predoctoral fellowship from the National Science Foundation and R.B.W. by a fellowship of the Canadian Medical Research Council. J.M.S.-T. was on leave from Laboratoire d'Enzymologie, Centre Nationale de la Recherche Scientifique, Gif-sur-Yvette, France.

1. Peters, T., Jr. (1975) in *The Plasma Proteins*, ed. Putnam, P. W. (Academic, New York), Vol. 1, pp. 133–181.
2. Tse, T. P. H., Morris, H. P. & Taylor, J. M. (1978) *Biochemistry* **17,** 3121–3128.
3. Britten, R. J. & Davidson, E. H. (1969) *Science* **165,** 349–357.
4. Gilbert, W. (1978) *Nature (London)* **271,** 501.
5. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303–2308.
6. Sternberg, N., Tiemeier, D. & Enquist, L. (1977) *Gene* **1,** 255–280.
7. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. & Smithies, O. (1977) *Science* **196,** 161–169.
8. Hohn, B. & Murray, K. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3259–3263.
9. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
10. Sala-Trepat, J. M., Dever, J., Sargent, T. D., Thomas, K., Sell, S. & Bonner, J. (1979) *Biochemistry*, in press.
11. Higuchi, R., Paddock, G. V., Wall, R. & Salser, W. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3146–3150.
12. Roychoudhury, R., Jay, E. & Wu, R. (1976) *Nucleic Acids Res.* **3,** 101–116.
13. Curtiss, R., III, Pereira, D. A., Hsu, J. C., Hull, S. C., Clarke, J. E., Maturin, L. J., Sr., Goldsmith, R., Moody, R., Inoue, M. & Alexander, L. (1977) in *Proceedings of the 10th Miles International Symposium*, eds. Beers, R. F., Jr. & Bassett, E. G. (Raven, New York), pp. 45–56.
14. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3727–3731.
15. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3961–3965.
16. Isemura, S. & Ikenaka, T. (1978) *J. Biochem.* **83,** 35–48.
17. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
18. White, R. L. & Hogness, D. S. (1977) *Cell* **10,** 177–192.
19. Davis, R. W., Simon, M. & Davidson, N. (1971) *Methods Enzymol.* **21D,** 413–428.
20. Southern, E. (1975) *J. Mol. Biol.* **98,** 503–517.
21. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23,** 641–646.
22. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1184–1188.
23. Dugaiczyk, A., Woo, S. L. C., Lai, E. C., Mace, M. L., Jr., McReynolds, L. & O'Malley, B. W. (1978) *Nature (London)* **274,** 328–333.
24. Tilghman, S. M., Curtiss, P. J., Tiemeier, D. C., Leder, P. & Weissman, C. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1309–1313.
25. Catterall, J. F., Stein, J. P., Lai, E. C., Woo, S. L. C., Dugaiczyk, A., Mace, M. L., Means, A. R. & O'Malley, B. W. (1979) *Nature (London)* **278,** 323–327.
26. Strair, R. K., Yap, S. H., Nadal-Ginard, B. & Shafritz, D. A. (1977) *J. Biol. Chem.* **253,** 1328–1331.