Construction and Characterization of cDNA Encoding the α2 Chain of Chicken Type IX Collagen†

Yoshifumi Ninomiya,‡ Michel van der Rest,§ Richard Mayne,∥ Guillermina Lozano,‡ and Bjorn Reino Olsen*,‡

Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Rutgers Medical School, Piscataway, New Jersey 08854, Genetics Unit, Shriners Hospital, Montreal, Canada H3G 1A6, and Departments of Anatomy and Biochemistry, University of Alabama, Birmingham, Alabama 35294

Received December 20, 1984

ABSTRACT: We have isolated and characterized a cDNA encoding the carboxy-terminal half of one of the polypeptide subunits of a novel disulfide-bonded collagen found in hyaline cartilage. This collagen has been given the type assignment type IX, and it has several unusual characteristics. First, the polypeptide subunits are shorter than α-chains of the fibrillar collagens types I, II, and III. Second, type IX molecules are heterotrimers of three genetically distinct polypeptide subunits. Third, type IX molecules contain three triple-helical collagenous domains interspersed with noncollagenous domains. When chicken cartilage collagens are extracted with pepsin, type IX collagen is cleaved and gives rise to the triple-helical fragments HMW and LMW. The identification of the cDNA reported here is based on a comparison of the amino acid composition of tryptic peptides derived from LMW with the composition of tryptic peptides predicted from the nucleotide sequence of the cDNA. We also show that the amino-terminal sequence of one of the subunits of LMW is identical with the sequence predicted from the nucleotide sequence of the cDNA. Finally, we demonstrate that the amino-terminal amino acid sequence of a tryptic peptide isolated from one of the subunits of HMW is identical with a sequence predicted from the cDNA. We have given the polypeptide chain encoded by the cDNA reported here the name α2(IX), and we show that it is homologous to the α1(IX) chain previously characterized by us.

Cartilaginous tissues contain a unique set of collagenous proteins (Mayne & von der Mark, 1983). Of these collagens, type II collagen with molecules consisting of three identical α1(II) chains (Miller, 1976) is the most abundant, constituting about 85% of the total tissue collagen. In addition, hyaline cartilage from both mammals and birds contains several minor collagenous components. Among these is a disulfide-bonded collagenous protein that we have recently characterized as type IX collagen (Ninomiya et al., 1984; van der Rest et al., 1985). Amino acid analysis of tryptic peptides and limited amino acid sequence analyses indicate that type IX collagen contains three genetically distinct chains (Mayne et al., 1984). Analysis of the nucleotide sequence of a clone isolated from a cDNA[1] library made with chick embryo sternal cartilage mRNA has allowed the deduction of the amino acid sequence of one of the three type IX collagen chains (Ninomiya & Olsen, 1984). This chain, α1(IX), contains three collagenous regions separated by noncollagenous sequences. This suggests that type IX collagen molecules form three triple-helical domains separated by non-triple-helical domains (van der Rest et al., 1985). In this model, the molecules would be composed of three homologous but nonidentical polypeptide chains. To test this model and examine the degree of homology between these polypeptide subunits, we have screened a chick cartilage cDNA library for additional type IX collagen cDNAs. In this report we describe the isolation and characterization of one such clone. On the basis of nucleotide sequence analysis of the cDNA we show that it codes for a polypeptide that indeed is homologous to the α1(IX) chain. A comparison of the amino acid composition of tryptic peptides predicted from the nucleotide sequence of the cDNA and the composition of tryptic peptides derived from the triple-helical fragment LMW isolated by pepsin extraction of cartilage strongly suggests that the cDNA contains the coding information for a subunit of LMW. Definitive proof that this is indeed correct is provided by the identity between the partial amino acid sequence of one of the subunits of LMW and the cDNA-derived sequence. The polypeptide chain that corresponds to the new cDNA is distinct from the previously identified α1(IX) chain. We have, therefore, given the new chain the designation α2(IX).

Materials and Methods

Preparation of RNA, Synthesis of cDNA, and Screening of Recombinant Clones. RNA was extracted from 17-day-old chick embryo sternal cartilage with guanidine hydrochloride (Adams et al., 1977). Poly(A)+ RNA was isolated by oligo(dT)–cellulose chromatography and fractionated by sucrose gradient centrifugation (Ninomiya et al., 1984). The synthesis and molecular cloning of cDNA were according to the method of Okayama & Berg (1982) as described previously (Ninomiya et al., 1984). To screen for cDNAs specific for minor chondrocyte collagens, we selected transformants containing plasmids whose inserts had a large number of sites for the restriction endonuclease *Sau*96I and whose inserts did not show any hybridization to type II or α1(IX) collagen cDNA probes. Transformants containing such plasmids were then characterized further by using their plasmid inserts as probes for

† This study was supported by Research Grants AM 21471, AM 34059, and AM 30481 from the National Institutes of Health, Medical Research Council of Canada Grant MA 7796, and the Shriners of North America.

‡ Rutgers Medical School.

§ Shriners Hospital.

∥ University of Alabama.

[1] Abbreviations: bp, base pair(s); cDNA, complementary DNA; HPLC, high-performance liquid chromatography; mRNA, messenger RNA; oligo(dT), oligo(thymidylic acid); poly(A), poly(adenylic acid); 1 × SSC, 0.15 M NaCl in 0.015 M sodium citrate buffer, pH 7.0.

0006-2960/85/0424-4223$01.50/0

RNA transfer blot analyses. The use of poly(A)+ RNA from sternal cartilage and calvaria for the analyses allowed identification of cDNAs specific for cartilage mRNAs.

RNA Transfer Blot and Nucleotide Sequence Analyses. For most RNA blot analyses, poly(A)+ RNAs were electrophoresed on 0.8% agarose gels in the presence of formamide/formaldehyde, transferred onto nitrocellulose filters, and hybridized to nick-translated DNA as described (Thomas, 1980). In some experiments RNA was electrophoresed on 1% agarose gels containing methylmercury hydroxide (Alwine et al., 1977). Methylmercury hydroxide was removed, and the gels were prepared for transfer as described by Alwine et al. (1977), except that the RNA was transferred to New England Nuclear gene screen filters in 25 mM potassium phosphate, pH 7.0. After hybridization in 3 × SSC and 1% sarkosyl at 65 °C, the filters were washed twice in 3 × SSC and 0.5% sarkosyl and twice in 3 × SSC at 65 °C (Overbeek et al., 1981).

Nucleotide sequence analysis of recombinant plasmids was performed by using the chemical cleavage method as well as the dideoxy chain termination technique. For Maxam and Gilbert sequencing, restriction fragments were labeled at their 5′ ends with calf intestinal alkaline phosphatase and T4 polynucleotide kinase (Maxam & Gilbert, 1980). For dideoxy sequencing (Sanger et al., 1977) restriction fragments were cloned in two orientations into M13mp18 and M13mp19 by standard techniques (Norrander et al., 1983) and then sequenced by using a 17-nucleotide primer complementary to M13 sequences.

Isolation and Fractionation of HMW and LMW. Collagens were solubilized from chick sterna by limited pepsin digestion and the fragments HMW and LMW prepared by fractional salt precipitation as described previously (Reese & Mayne, 1981). Isolation of the C3 component of HMW was accomplished by molecular sieve chromatography and high-performance liquid chromatography (HPLC) after reduction and alkylation of LMW (van der Rest et al., 1985). Reduction and alkylation of LMW and separation of the components LMW1, LMW2A, LMW2B, and LMW3 were as described (Mayne et al., 1984).

Cleavage of Peptides with Trypsin. The peptides were dissolved in 0.2 M ammonium bicarbonate at a concentration of 1 mg/mL. Trypsin (Worthington) dissolved in water (1 mg/mL) was added to give an enzyme:substrate ratio of 1:30. After digestion at 37 °C for 4 h the reaction was terminated by addition of a drop of glacial acetic acid. Samples were then dried in a Speedvac concentrator.

High-Performance Liquid Chromatography (HPLC). The tryptic peptides were fractionated as described previously (van der Rest & Fietzek, 1982). The instrument used was from Beckman and consisted of a Model 334 chromatograph, a Model 160 UV monitor, equipped with a zinc lamp, and a CR1B data system. The column was a C_{18} Vydan TP201 (4.6 × 250 mm) from the Separation Group, and it was protected with a guard column filled with pellicular C_{18} resin (Waters Associates). Samples were dissolved in 9 mM trifluoroacetic acid before being injected into the column. Elution was achieved with an aqueous linear gradient of acetonitrile (0–32%) over 90 min at 1 mL/min.

Amino Acid Analysis. The samples were hydrolyzed in 6 N HCl containing 0.05% β-mercaptoethanol for 20 h under nitrogen. The analyses were performed on a Dionex D-500 amino acid analyzer as described previously (Ninomiya et al., 1984). *S*-(Carboxymethyl)cysteine was estimated by using the color factor measured for L-cystine. Tryptophan was not quantitated.

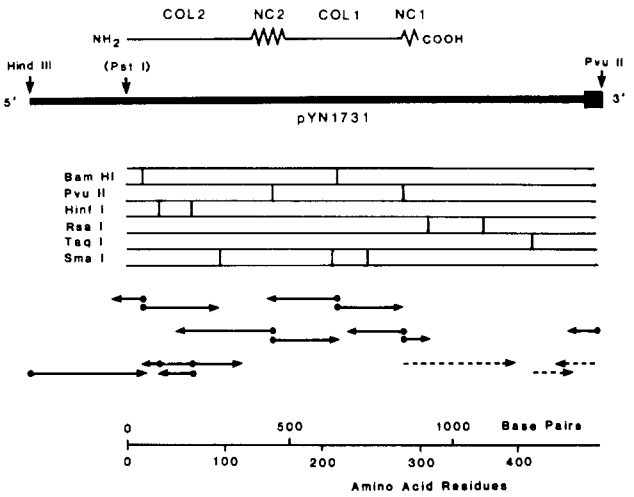

FIGURE 1: Restriction endonuclease map of the α2(IX) cDNA insert. The direction of transcription is from left to right, as indicated by the 5′ and 3′ notations. The strategy of nucleotide sequencing of the pYN1731 insert is indicated with dots representing the position of 5′-end labeling and arrows showing the direction and length of the sequence analysis. The regions sequenced by the dideoxy technique are indicated by dotted lines. The cDNA insert produced by the method of Okayama & Berg (1982) was cut from the vector by digesting the recombinant plasmid pYN1731 with *Hin*dIII and *Pvu*II. During the cloning procedure the *Pst*I site between the linker fragment and the cDNA insert was altered. Therefore, the insert could not be cut off with *Pst*I and *Pvu*II, and the fragment released with *Hin*dIII and *Pvu*II contains the linker fragment and the cDNA insert. In the figure, only the insert map is shown. At the top of the figure, the positions of the different domains within the translation product of α2(IX) mRNA are indicated.

Amino Acid Sequence Analysis. Amino acid sequences of LMW2A and LMW2B were determined by automated Edman degradation in the Beckman 890C protein/peptide sequencer as previously described (Seidah et al., 1981). The analyses were generously carried out by Nabil G. Seidah and Michel Chretien. The amino acid sequence of a tryptic peptide from the C3 component of HMW was determined in an Applied Biosystems gas-phase sequencer. The phenylthiohydantoin derivatives of the amino acids were identified by HPLC as described by Lazure et al. (1983).

RESULTS

Isolation and Characterization of the cDNA pYN1731. One of the cDNA clones, pYN1731, showed the presence of several sites for the restriction endonuclease *Sau*96I, and it hybridized to cartilage-specific RNA. Since many of the fragments generated by *Sau*96I cleavage of pYN1731 were multiples of 9 base pairs in length and this feature is characteristic for DNA sequences that code for collagenous peptides, pYN1731 was characterized in more detail. Figure 1 shows a restriction endonuclease cleavage map of the insert of the cDNA. The insert is about 1450 bp long. With the nick-translated insert of pYN1731 as probe, RNA transfer blot analysis of sternal cartilage poly(A)+ RNA showed that the cDNA hybridized to two mRNAs migrating close together that were between 2500 and 3000 bases in length (Figure 2). Thus pYN1731 is distinctly different from cDNAs that encode pro α1(II) or α1(IX) collagen chains (Figure 2). Nucleotide sequencing of the insert of pYN1731 showed that the coding strand of the cDNA contained about 130 adenylic acid residues at the 3′ end preceded by about 500 nucleotides in the 3′ nontranslated sequence between the first in-phase translational stop codon and the poly(A)+ tail. As seen in Figures 1 and 3, the amino acid sequence deduced from the nucleotide se-

Table I: Amino Acid Composition of Tryptic Peptides Isolated from LMW2B[a]

peptide fractions	7, 8	55	70	14	64	73[d]
S-CM-Cys	*b*					2.0 (2)
Hyp		3.4[c]	1.3		5.0	5.3
Asp + Asn	2.2 (2)		1.1 (1)		1.1	1.6 (1)
Thr	1.2 (1)					
Ser	0.2		0.2		0.3	2.6 (1)
Glu + Gln	1.4 (1)	1.8 (2)	1.1 (1)	1.9 (2)	0.3	1.8 (1)
Pro		4.2 (10)	1.1 (2)		1.1 (7)	3.3 (8)
Gly	3.8 (3)	10.1 (11)	7.6 (7)	2.0 (2)	8.0 (8)	11.0 (8)
Ala		2.3 (3)	2.3 (2)		1.1 (2)	6.6 (4)
Val		0.7 (1)	1.0 (1)			1.0 (1)
Met		0.3 (2)			0.2 (1)	
Ile		0.2	1.9 (2)		1.9 (2)	0.4
Leu		1.7 (2)	2.1 (2)		2.0 (2)	2.5 (2)
Tyr						
Phe						1.0 (1)
His		0.8 (1)			2.0 (2)	
Hyl			1.0[c]	0.9		
Lys	0.2		1.0 (2)	(1)	1.0 (1)	
Arg	2.0 (2)	1.0 (1)		2.0 (2)		1.0 (1)

[a] Values in parentheses are calculated from the sequence predicted from pYN1731. [b] Amounts less than 0.1 are not given. [c] Note that hydroxyprolyl and hydroxylysyl residues are found as prolyl and lysyl residues in the predicted sequence. [d] Peptide 73 was not pure since the composition of this fraction showed variation among different HPLC runs. This probably explains the discrepancies between the composition shown in the table and the predicted values. The predicted amounts for each residue in peptide 73 are calculated on the basis of the assumption that pepsin has cleaved the peptide bond between a leucyl and a glycyl residue within the NC1 domain of α2(IX).

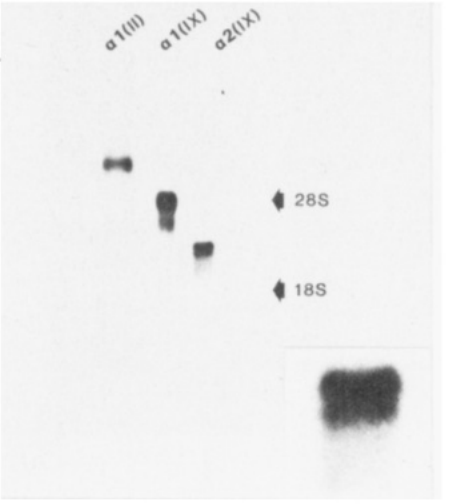

FIGURE 2: RNA transfer blot analysis of cloned cDNAs. Chicken sternal cartilage poly(A)+ RNAs were electrophoresed either in a 0.8% agarose gel [in the case of probes for α1(II) and α1(IX) collagens] or on a 1% methylmercury–agarose gel [in the case of the α2(IX) collagen probe] as described under Materials and Methods, blotted onto nitrocellulose, and hybridized to nick-translated inserts from different recombinant plasmids. As indicated in the text, mRNAs that hybridized to the α2(IX) collagen probe migrated as two bands close together (see higher magnification in inset, lower right-hand corner). Chicken sternal cartilage 18S and 28S RNAs were used as RNA size markers as indicated. The numbers of bases assigned to the size markers were 1900 and 4100.

quence could be divided into two collagenous domains with a glycyl residue in every third position and two noncollagenous domains. When compared to the amino acid sequence of α1(IX) chains (Figure 3), the sequence derived from pYN1731 is seen to be organized into similar domains as the α1(IX) sequence. Of interest is that both sequences contain two imperfections in the triplet structure within their COL1 domains (Figure 3). One of the two imperfections consists of the dipeptide sequence Gly-Arg inserted between Gly-X-Y triplets. The second imperfection consists of a deletion of a glycyl residue within the triplet structure of COL1 (Figure 3).

The cDNA pYN1731 Codes for One of the Polypeptide Subunits of Type IX Collagen. We have previously shown that type IX collagen gives rise to the HMW and LMW collagenous fragments during pepsin extraction of chicken hyaline cartilage by demonstrating that the cDNA pYN1738 contains the coding information for one of the subunits of HMW (C2) and one of the subunits of LMW (LMW3) (van der Rest et al., 1985). In this study, we compared the amino acid compositions of tryptic peptides derived from LMW2 with amino acid compositions predicted for tryptic fragments of the polypeptide chain encoded by pYN1731. Such a comparison demonstrated that many of the tryptic peptides recovered from LMW2 could be accounted for within the pYN1731-derived sequence. However, the comparison was complicated by the fact that the LMW2 subunit of LMW gave rise to two peaks, designated 2A and 2B, when reduced and alkylated LMW was subjected to CM-cellulose chromatography (Mayne et al., 1984). When tryptic peptides derived from each of the two peaks were fractionated by reverse-phase HPLC, fractions 2A and 2B gave very similar peptide patterns at both the start and the end of the gradient, but in the center of the gradient a family of five peptides was observed, which for peak 2B eluted slightly later in the gradient (Mayne et al., 1984). Table I shows the amino acid compositions of the tryptic peptides from LMW2B and the compositions predicted from the DNA sequence of pYN1731. The peptides are labeled according to their elution time (in minutes) from the HPLC column. Except for the flow-through fractions 7 and 8, the order in which the peptide compositions in the table are arranged follows the sequence in which the peptides occur along the predicted sequence (Figure 3).

Amino acid sequence analysis of the LMW2A and LMW2B components by automated Edman degradation provided definitive proof that LMW2 is encoded by pYN1731. The sequences obtained from LMW2A and LMW2B also provided an explanation for the difference between the peptide material of peaks 2A and 2B. As shown in Figure 4, the amino-terminal amino acid sequences of 2A and 2B are derived from the same common polypeptide. However, due to pepsin cleavage at two different sites within the polypeptide, two fragments having a different amino-terminal end are generated. The sequences obtained from LMW2A and LMW2B are both found within the sequence derived from pYN1731 (residues 598–618, Figure 3).

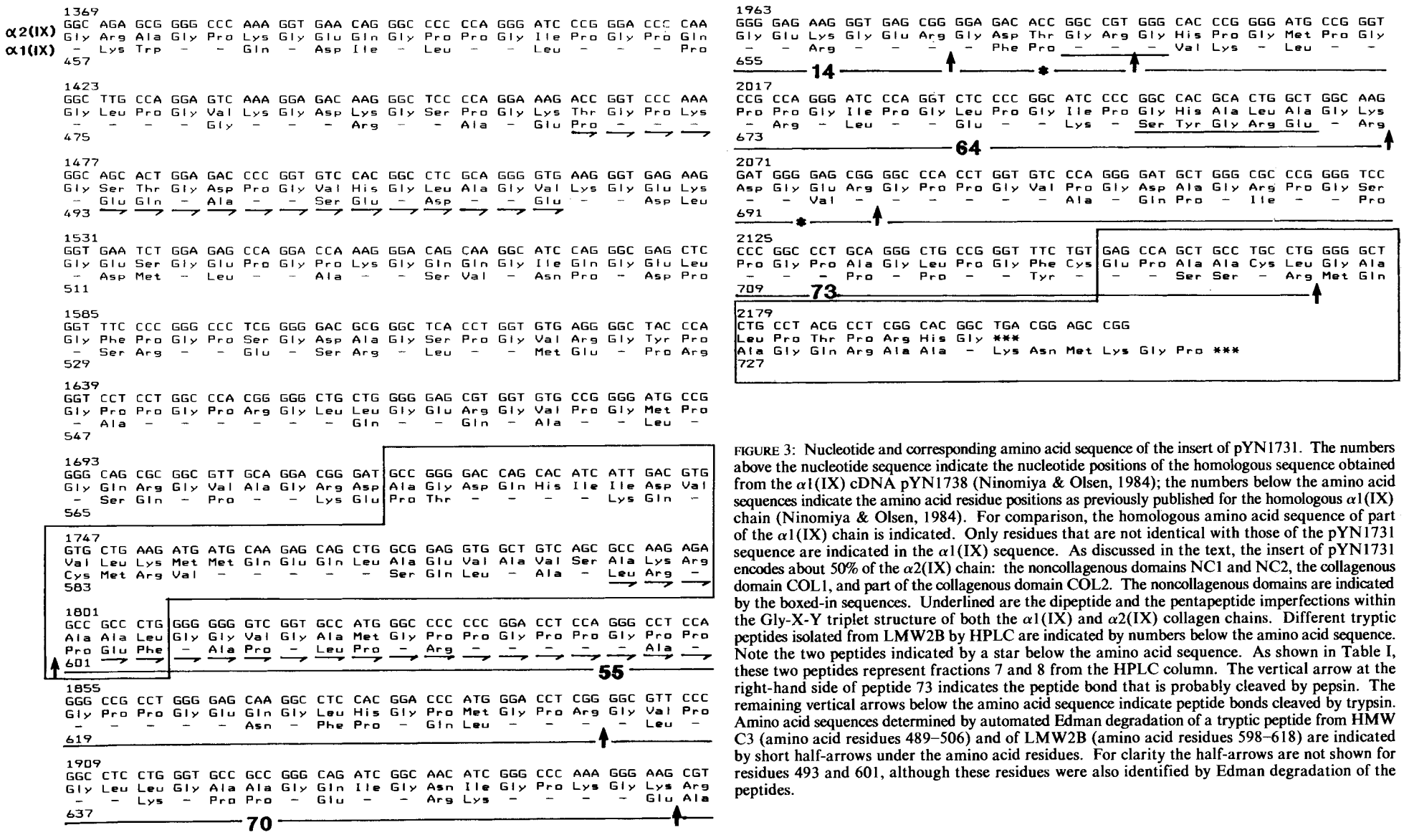

FIGURE 3: Nucleotide and corresponding amino acid sequence of the insert of pYN1731. The numbers above the nucleotide sequence indicate the nucleotide positions of the homologous sequence obtained from the α1(IX) cDNA pYN1738 (Ninomiya & Olsen, 1984); the numbers below the amino acid sequences indicate the amino acid residue positions as previously published for the homologous α1(IX) chain (Ninomiya & Olsen, 1984). For comparison, the homologous amino acid sequence of part of the α1(IX) chain is indicated. Only residues that are not identical with those of the pYN1731 sequence are indicated in the α1(IX) sequence. As discussed in the text, the insert of pYN1731 encodes about 50% of the α2(IX) chain: the noncollagenous domains NC1 and NC2, the collagenous domain COL1, and part of the collagenous domain COL2. The noncollagenous domains are indicated by the boxed-in sequences. Underlined are the dipeptide and the pentapeptide imperfections within the Gly-X-Y triplet structure of both the α1(IX) and α2(IX) collagen chains. Different tryptic peptides isolated from LMW2B by HPLC are indicated by numbers below the amino acid sequence. Note the two peptides indicated by a star below the amino acid sequence. As shown in Table I, these two peptides represent fractions 7 and 8 from the HPLC column. The vertical arrow at the right-hand side of peptide 73 indicates the peptide bond that is probably cleaved by pepsin. The remaining vertical arrows below the amino acid sequence indicate peptide bonds cleaved by trypsin. Amino acid sequences determined by automated Edman degradation of a tryptic peptide from HMW C3 (amino acid residues 489–506) and of LMW2B (amino acid residues 598–618) are indicated by short half-arrows under the amino acid residues. For clarity the half-arrows are not shown for residues 493 and 601, although these residues were also identified by Edman degradation of the peptides.

```
LMW 1:          Leu-Arg-Lys-Pro-Leu-Ser-Pro-Gly-Met-Thr-Gly-Arg-Hyp-Gly-Pro-Ala-Gly-Pro-Hyp-

LMW 2A:                                  Gly-Gly-Val-Gly-Ala-Met-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-

LMW 2B:            Ala-Lys-Arg-Ala-Ala-Leu-Gly-Gly-Val-Gly-Ala-Met-Gly-Pro-Hyp-Gly-Pro-Hyp-Gly-Pro-Hyp-

pYN1731:-Ala-Val-Ser-Ala-Lys-Arg-Ala-Ala-Leu-Gly-Gly-Val-Gly-Ala-Met-Gly-Pro-Pro-Gly-Pro-Pro-Gly-Pro-Pro-
                        600                                                 610
```

FIGURE 4: Amino acid sequence analysis of LMW1, LMW2A, and LMW2B by automated Edman degradation. The corresponding derived sequence for the peptide encoded by pYN1731 is given below the three sequences. The numbers below the derived sequence indicate the amino acid residue positions as shown in Figure 3. As discussed in the text, the sequences from LMW2A and LMW2B are derived from the same polypeptide chain. However, due to two different pepsin cleavage sites within the $\alpha 2$(IX) chains, two LMW2 components (2A and 2B) are generated during pepsin extraction of LMW from cartilage. Original data for the sequence analysis by Edman degradation are available upon request.

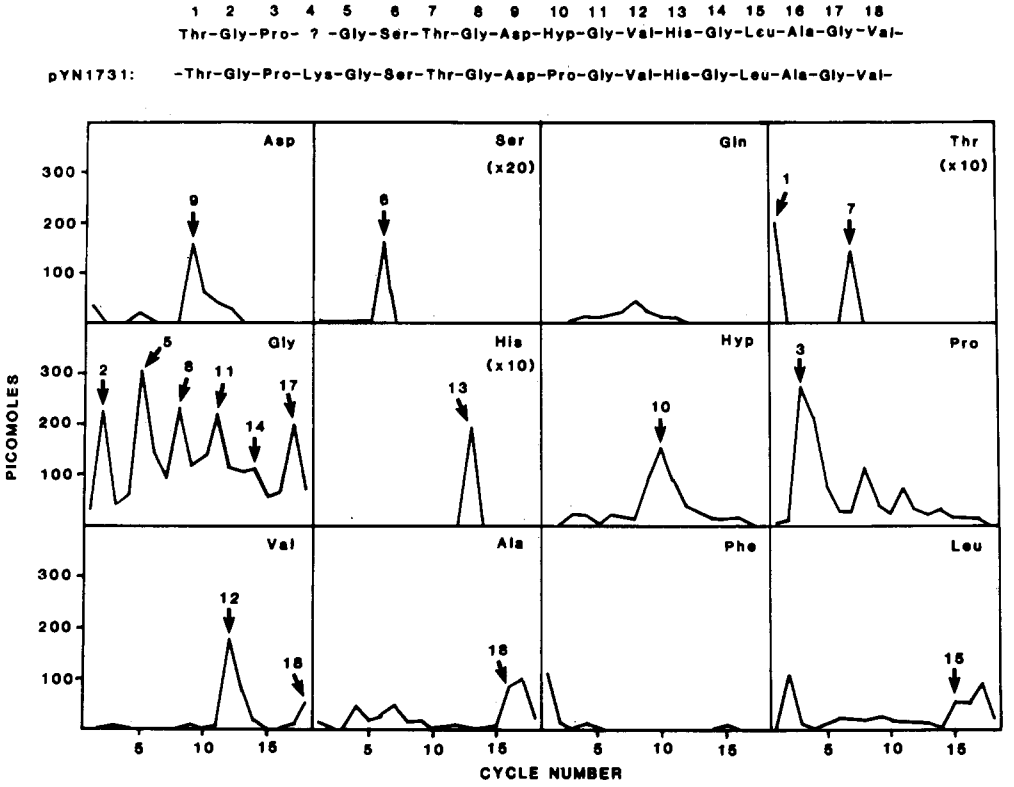

FIGURE 5: Amino acid sequence analysis of a tryptic peptide isolated from the C3 component of HMW (van der Rest et al., 1985). As described in the text, the peptide elutes at 60 min when a tryptic digest of C3 is fractionated by HPLC as described (van der Rest et al., 1985). For the sequence analysis about 500 pmol was subjected to 18 degradation cycles in the gas-phase sequencer. The sequence determined from the run is indicated at the top of the figure, together with the cycle numbers. Also indicated is a sequence derived from the nucleotide sequence of the cDNA pYN1731. Because only nonaqueous phases of phenylthiohydantoin (PTH) derivatives were examined, the amino acid residue in cycle 4 could not be determined. The recovery (in picomoles) of PTH derivatives of each of 12 amino acid residues is given for each cycle along the left-hand side of the diagrams for each amino acid. For Thr and His residues a scaling factor of 10 was used in drawing the diagrams; for Ser residues a scaling factor of 20 was used. The cycle numbers are indicated along the abscissas of the diagrams for each kind of amino acid residue. The assignment of different residues to each of the 18 cycles (except cycle 4) is indicated by the numbers placed above the peaks in the diagrams for each of the 12 different amino acids. As indicated in the text, the peptide was not entirely pure, but the results allowed identification of a major and a minor sequence. Only the assignment of residues to the major sequence is indicated. Thus, although phenylalanine, leucine, alanine, and proline were found in significant amounts in cycles 1, 2, 8, 11, 16, and 17, they were not assigned to the major sequence since other residues were found in larger quantities in the same cycles.

As shown in Figure 4, the amino acid sequence analysis of the components LMW1, LMW2A, and LMW2B indicated that the subunits of LMW are derived from three polypeptide chains of different primary structures. Previously, we have shown (van der Rest et al., 1985) that the LMW3 component is part of the $\alpha 1$(IX) collagen chain and that it is connected to the C2 component of HMW. To identify the component of HMW that is connected to the LMW2 component within intact type IX collagen, we compared amino acid compositions of tryptic peptides isolated from the C3, C4, and C5 components of HMW (van der Rest et al., 1985) with compositions predicted from the nucleotide sequence of pYN1731. This comparison (data not presented) suggested that one of the tryptic peptides from HMW C3 was encoded by pYN1731. This tryptic peptide [eluting at the 60-min position from the HPLC column; see van der Rest et al. (1985)] was, therefore, subjected to amino acid sequence analysis. The results of this analysis, shown in Figure 5, demonstrated that although the peptide was not entirely pure, a major amino acid sequence could be identified. With the exception of cycle 4 (see Figure 5 legend), the analysis allowed identification of the first 18 amino acid residues of the major sequence. This sequence is identical with a sequence (residues 489–506, see Figure 3) predicted from the nucleotide sequence of pYN1731 (Figure 5). We conclude, therefore, that the C3 component of HMW and the LMW2 component are contained with the same polypeptide chain of type IX collagen.

Discussion

The results presented here demonstrate conclusively that

the collagenous fragment LMW isolated from hyaline cartilage by pepsin extraction (Reese & Mayne, 1981) contains three genetically distinct polypeptide subunits. Previously, we have reported (van der Rest et al., 1985) that the LMW3 subunit represents part of the α1 chain of type IX collagen. Here, we show that the LMW2A and LMW2B components are pepsin-resistant products of a second chain of type IX collagen and that the cDNA pYN1731 contains the coding information for about 50% of that chain. We propose to give the chain encoded by pYN1731 the designation α2(IX). The primary structure of the LMW1 component is clearly different from those of LMW3, LMW2A, and LMW2B. LMW1, therefore, is a fragment of the third chain of type IX collagen, and we will refer to this as α3(IX).

In a previous study (van der Rest et al., 1985), we showed that the cDNA pYN1738 (Ninomiya & Olsen, 1984) encoded both the C2 component of HMW (Reese & Mayne, 1981) and the LMW3 component of LMW. This provided definitive evidence that the HMW and LMW collagenous fragments are derived from the same parent molecule. Here, we show that the LMW2 component is covalently attached to the C3 component of HMW within the intact type IX collagen molecule. Therefore, the α2(IX) chain must contain the sequences of the C3 component of HMW and the LMW2 components. On the basis of results obtained in a previous study of the subunit structure of HMW (Reese et al., 1982) it was concluded that the C4 component is located at the amino terminus of HMW. Therefore, it is likely that C4 is attached to the C3 component within the α2(IX) chain. However, we cannot rule out the possibility that the nick between C4 and C3 in HMW is not due to pepsin but is the result of some processing step during biosynthesis of type IX collagen.

In early sequence studies of type I collagen it was noted that several amino acids are nonrandomly distributed between X and Y positions of the collagenous triplet structure $(\text{Gly-X-Y})_n$ [for references, see Bornstein & Traub (1979)]. The basis for this selectivity is probably steric hindrance and intramolecular side chain to side chain and side chain to backbone interactions that stabilize the triple-helical structure [see Bornstein & Traub (1979)]. It is, therefore, not surprising to find the same nonrandomness within the collagenous domains of α1(IX) and α2(IX) collagen chains. For example, glutamic acid, leucine, and phenylalanine are preferentially found in the X position in the collagenous domains of α1(IX) and α2(IX), as they are in α1(I) and α2(I) chains, whereas arginine is preferentially found in the Y position in all four chains. Certain amino acid residues show interesting differences in their distributions: In α1(I) and α2(I) chains aspartic acid and serine are almost evenly distributed between the X and Y positions (Bornstein & Traub, 1979); in type IX collagen chains these residues are preferentially found in the X position.

What is the degree of homology between the α1(IX) and α2(IX) collagen chains? Comparison of the amino acid sequences, including the invariant glycyl residues within the collagenous domains, reveals an overall homology of 61% for COL1 and 56% for the 117 residues of COL2 reported here (Figure 3). This is clearly lower than the 66% homology between the triplet sequences of α1(I) and α2(I) collagen chains but not very different from the homology between type I and type II collagen chains. Of course, type IX collagen chains are distinctly different from those of fibrillar collagens types I, II, and III in that they contain several collagenous sequence domains interspersed with noncollagenous domains. This difference between type II collagen and the fibrillar collagens extends to the exon/intron structure of the genes, and we have recently shown (Lozano et al., 1985) that type IX collagen genes belong to a novel class of vertebrate collagen genes distinctly different from that of fibrillar collagens.

It has recently been reported (Bruckner et al., 1985) that type IX collagen contains covalently bound glycosaminoglycans and that (Vaughan et al., 1985) in fact type IX collagen may be identical with the proteoglycan Pg-Lt (Noro et al., 1983). It has also been shown (Vaughan et al., 1985) that all three of the polypeptide chains of type IX collagen contain mannose-rich carbohydrate side chains linked to asparaginyl residues. We have searched the sequences reported here (Figure 3) for potential carbohydrate attachment sites of the structure Asn-X-Thr/Ser and find that no such sites are present in the carboxy-terminal half of α1(IX) and α2(IX) collagen chains. In addition, no such sites are present in the more extended sequence previously published for the α1(IX) chain (Ninomiya & Olsen, 1984). We conclude, therefore, that if such oligosaccharide attachment sites are indeed present in type IX collagen, they must be localized [at least for the α1(IX) and α2(IX) chains] in the amino-terminal region for which cDNA-derived amino acid sequences are not yet available.

REFERENCES

Adams, S. L., Sobel, M. E., Howard, F. H., Olden, K., Yamada, K. M., de Crombrugghe, B., & Pastan, I. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3399–3403.

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5350–5354.

Bornstein, P., & Traub, W. (1979) *Proteins (3rd Ed.) 4*, 411–632.

Bruckner, P., Vaughan, L., & Winterhalter, K. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2608–2612.

Lazure, C., Seidah, N. G., Chretien, M., Lallier, R., & St. Pierre, S. (1983) *Can. J. Biochem. 61*, 287–292.

Lozano, G., Ninomiya, Y., Thompson, H., & Olsen, B. R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Mayne, R., & von der Mark, K. (1983) in *Cartilage* (Hall, B. K., Ed.) Vol. 1, pp 181–214, Academic Press, New York.

Mayne, R., van der Rest, M., Weaver, D. C., & Butler, W. T. (1984) *J. Cell. Biochem.* (in press).

Miller, E. J. (1976) *Mol. Cell. Biochem. 13*, 165–192.

Ninomiya, Y., & Olsen, B. R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3014–3018.

Ninomiya, Y., Showalter, A. M., & Olsen, B. R. (1983) in *Limb Development and Regeneration* (Kelley, R. O., Goetinck, P. F., & MacCabe, J. A., Eds.) Part B, pp 183–192, Alan R. Liss, New York.

Ninomiya, Y., Showalter, A. M., van der Rest, M., Seidah, N. G., Chretien, M., & Olsen, B. R. (1984) *Biochemistry 23*, 617–624.

Noro, A., Kimata, K., Oike, Y., Shinomura, T., Maeda, N., Yano, S., Takahashi, N., & Suzuki, S. (1983) *J. Biol. Chem. 258*, 9323–9331.

Overbeek, P. A., Merlino, G. T., Peters, N. K., Cohn, V. H., Moore, G. P., & Kleinsmith, L. J. (1981) *Biochim. Biophys. Acta 656*, 195–205.

Reese, C. A., & Mayne, R. (1981) *Biochemistry 20*, 5443–5448.

Reese, C. A., Weidemann, H., Kuhn, K., & Mayne, R. (1982) *Biochemistry 21*, 826–829.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Seidah, N. G., Rochemont, J., Hamelin, J., Lis, M., &

Chretien, M. (1981) *J. Biol. Chem. 256*, 7977–7984.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
van der Rest, M., & Fietzek, P. (1982) *Eur. J. Biochem. 124*, 491–496.
van der Rest, M., Mayne, R., Ninomiya, Y., Seidah, N. G., Chretien, M., & Olsen, B. R. (1985) *J. Biol. Chem. 260*, 220–225.
Vaughan, L., Winterhalter, K. H., & Bruckner, P. (1985) *J. Biol. Chem. 260*, 4758–4763.

Amplified DNA of the Novikoff Hepatoma Nucleolus Is Arranged in a 7.3-Kilobase Tandem Repeat†

David L. Parker, Deborah L. Mroczka, and Lawrence I. Rothblum*
Department of Pharmacology, Baylor College of Medicine, Houston, Texas 77030
Received December 26, 1984

ABSTRACT: We have reported previously the cloning and characterization of a nucleolar-localized 5.8-kilobase (kb) *Eco*RI fragment that is approximately 50-fold more highly reiterated in Novikoff hepatoma tumor cells than in normal rat liver [Parker, D. L., Busch, H., & Rothblum, L. I. (1981) *Biochemistry 20*, 762]. In the present study, the arrangement of these 5.8-kb *Eco*RI segments within the Novikoff hepatoma genome was investigated. Through the use of "indirect" restriction site mapping, partial restriction enzyme digestions, and molecular cloning, we have determined that the 5.8-kb *Eco*RI fragment and a 1.5-kb *Eco*RI fragment together constitute a 7.3-kb unit. The 7.3-kb unit is present in the hepatoma genome as a tandem repeat and constitutes the unit of the DNA that has been amplified. Studies on the arrangement of homologous sequences in the normal rat genome indicate that the amplified DNA may have been derived by a rearrangement and amplification of the nontranscribed spacer of the ribosomal DNA (rDNA) repeat.

Novikoff hepatoma cells have been shown to contain a moderately repetitive (4000–10 000 copies per haploid genome), 5.8-kilobase (kb) *Eco*RI fragment that is nucleolar localized yet does not hybridize to ribosomal RNA (rRNA) (Parker et al., 1981). Reassociation studies, using a cloned version of the 5.8-kb fragment as probe, demonstrated that these sequences were approximately 50-fold more abundant in Novikoff hepatoma cells than in normal rat liver DNA. Comparative Southern blots of *Eco*RI-digested normal rat liver and Novikoff hepatoma DNA failed to reveal the presence of a homologous 5.8-kb fragment in the liver. Two less intense hybrids of lower molecular weight (4.5 and 4.1 kb) were detected in roughly equal intensities in both the tumor and nontumor DNAs (Rothblum et al., 1982). The relative abundance of the 5.8-kb fragment, and its specific subnuclear localization (nucleolus), suggested that it may have evolved through the rearrangement and amplification of a piece of nucleolar DNA.

Amplified DNA segments have generally been found to exist in clusters [Schmike et al., 1978; Miller et al., 1979; reviewed by Schimke (1984) and by Stark & Wahl (1984)], and within a given cell population there exist subpopulations with varying degrees of multiplicity of specific genes (Wahl et al., 1979b; Alt et al., 1976). In many of the examined cases, with the exception of the amplified ribosomal DNA (rDNA) genes described by Miller et al. (1979) and the chorion protein genes (Spradling & Mahowald, 1980), the cells or cell lines that contain the amplified genes have been selected for by multiple rounds of exposure to a specific metabolic inhibitor. As a consequence of these manipulations, the amplified genes contain several different "novel joints", which may be the products of several rounds of amplification and recombination (Caizzi & Bostock, 1982; Schimke, 1984).

The examination of the genomic arrangement of the 5.8-kb fragment within the hepatoma genome was undertaken to determine the structure of the amplified DNA. We have also attempted to isolate the homologue of the 5.8-kb fragment in the normal rat genome. Several clones from the normal rat genome were isolated that very closely matched the restriction enzyme digestion pattern of the amplified fragment. Some of these clones represent the ribosomal nontranscribed spacer.

Materials and Methods

DNA Isolation and Restriction Analysis. All DNA isolations were performed as described previously (Parker et al., 1979). Restriction enzyme digestions were performed as recommended by the suppliers.

Partial restriction enzyme digestions were performed by the addition of 1 unit of enzyme/5 μg of DNA and aliquots removed at 3-min intervals. Southern transfers and hybridizations were performed as described by Wahl et al. (1979a).

^{32}P Labeling and Cloning of DNA Fragments. All radiolabeled DNA fragments were prepared by "nick translation" (Mackey et al., 1977). Cloning was performed by using the gt-WES system provided to us by Dr. P. Leder (Tiemeier et al., 1979) or by using a partial *Eco*RI rat genome library cloned in Charon 4A (generously provided by Drs. Tom Sargent and J. Bonner and their colleagues). Colony screening was performed essentially as described by Benton & Davis (1977). In some screenings, the stringency was increased by carrying out the final washes at 65 °C and in 0.1 × standard

† This work was supported by Cancer Research Program Grant CA-10893, P4, awarded by the National Institutes of Health, DHEW, and NIGMS Grant BRSG-79-P8. D.L.P. and D.L.M. were predoctoral fellows of the Houston Pharmacological Center (GMO-7405-0).

0006-2960/85/0424-4229$01.50/0