Volume 8 Number 8 1980 Nucleic Acids Research

Isolation and characterization of genomic DNA coding for α2 Type I collagen*

G.Vogeli, E.V.Avvedimento, M.Sullivan**, J.V.Maizel,Jr.**, G.Lozano, S.L.Adams, I.Pastan and B.de Crombrugghe

Laboratory of Molecular Biology, NCI, NIH, and **Laboratory of Molecular Genetics, NICHD, NIH, Bethesda, MD 20205, USA

Received 30 January 1980

ABSTRACT

We have isolated and characterized a segment of the chick α2 collagen gene by screening a library of chick genomic fragments using as hybridization probe an α2 collagen cDNA clone. Several clones were isolated and one of them, λgCOL 204, was used for further studies. The DNA of λgCOL 204 hybridizes to a unique species of mRNA the size of α2 collagen mRNA. This mRNA can be translated into a unique polypeptide which comigrates in SDS-gel electrophoresis with pro-α2 collagen. Electron microscopic analysis by R-loop technique indicates that λgCOL 204 contains 7Kb of the α2 collagen gene. This 7 Kb piece constitutes the 3' end of the gene. The same clone also contains 9 Kb of DNA that is immediately adjacent to the 3' end of the α2 collagen gene. The cloned segment of the α2 collagen gene is interrupted by 8 intervening sequences of various lengths. The coding sequences for collagen in this clone add up to approximately 1,800 bp, which correspond to about 1/3 of α2 collagen mRNA. DNA sequence analysis of a small coding segment of λgCOL 204 reveals a characteristic collagen type sequence which encodes for an amino acid sequence identical to a sequence found in calf α2 collagen. The sequence of this region of the protein has not yet been determined for the chick α2 collagen.

INTRODUCTION

Collagens are the major structural proteins of animals (for a review, see 1). Their synthesis is carefully controlled during development. There are a variety of diseases in which collagen synthesis is deranged. In tissue culture, collagen synthesis is affected by transformation (2,3) and by treatment with various drugs (4).

To study factors that control collagen gene expression and to help understand the nature of certain diseases of collagen metabolism, it is important to isolate genes for the various collagens. We, and others, have reported the isolation of cDNA clones for α1 collagen (5,6) and α2 collagen (7,8) but cDNA clones are only the first step in analyzing collagen regulation since the regions controlling collagen expression probably lie outside the regions contained in the mature mRNAs used to make the cDNA clones and may also involve

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

the intervening sequences. Therefore, we have undertaken the isolation of the genes for collagen. Here we report the isolation and characterization of a recombinant phage λgCOL 204 which contains about one-third of the α2 collagen gene. This cloned portion of the collagen gene lies at the 3' end of the gene. In addition to α2 collagen sequences, λgCOL 204 also contains non-coding sequences adjacent to the 3' end of the α2 collagen gene.

MATERIALS AND METHODS

Biosafety precautions

The work with recombinant DNA was done in a P3 or P2 physical containment facility using the EK2 host-vector system E. coli DP50 supF/λ Charon 4A and the EK2 host-vector system E. coli C600/pBR322 in compliance with the NIH guidelines for recombinant research.

Screening of the λCharon 4A library

The library of chick genomic DNA fragments in Charon 4A (9; a gift from Drs. J. Dodgson, R. Axel, and D. Engel) was screened by the in situ hybridization method (10) as outlined by T. Maniatis et al. (11). E. coli DP50 supF (a gift from Dr. F.R. Blattner) requires 100 μg/ml diaminopimelic acid and 40 μg/ml thymine. It was grown in TBMM (10 g/l Tryptone, 5 g/l NaCl, 0.01 M $MgSO_4$, 2 g/l Maltose) and stored at 4°C in 0.01 M $MgSO_4$. Phage was stored and diluted with TMG (0.05 M Tris-HCl pH 7.5, 0.1 M NaCl, 0.01 M $MgSO_4$, 0.1 g/l Gelatine). Cells and phage were preincubated at 30°C for 10 min and plated with 0.7% Agarose (Sigma). 10^4 pfU were plated per 15 cm dish on L-Broth agar (10 g/l Tryptone 5 g/l Bacto Yeast, 5 g/l NaCl, 0.005 M $MgSO_4$, pH 7.5, 12 g/l agar) and incubated overnight at 32°C.

The plaques were lifted with two subsequent applications (the first filter for 10 min, the second filter for 20 min) of nitrocellulose filters (Schleicher and Schuell BA 85). The filters were air dried for at least 1 hr; the DNA on the filters was denatured and neutralized by letting the filters float for 10 sec each on the surface of the solutions (10), plaques facing the air, and washed by submersion. They were then dried in the vacuum oven for at least 2 hrs. at 80°C. Overnight hybridization was done at 68°C in 6 x SSC and 4 x Denhardt solution (12), with 5 μg/ml micrococcal DNA (Sigma), and 10^6 cpm ^{32}P labelled nick-translated insert DNA from pCOL 1 per filter. pCOL 1 is a recombinant plasmid which contains α2 collagen cDNA sequences (7).

After drying, the filters were autoradiographed at -70°C using preflashed film (Kodak XR5) and intensifying screens (Cronex, DuPont) (13). Only positive plaques occurring in both filters were purified by several cycles of

plating and rehybridization until more than 98% of all plaques were positive.

Isolation of Phage DNA

E. coli DP50 supF was infected with the purified phage at a m.o.i. of 10^{-3} and grown overnight in L-Broth supplemented with 0.005 M $CaCl_2$ at 32°C. After shaking the culture for an additional 15 min with chloroform, it was incubated for 1 hr at 4°C with 1 µg/ml RNase A (Worthington) and 1 µg/ml Deoxyribonuclease I (Worthington). The solution was made 1M with NaCl and the cell debris were removed by centrifugation. The phage was precipitated with 10% (w/v) polyethylene glycol at 4°C for 1 hr and collected by centrifugation. The pellet was suspended in TMG and extracted 3 times with chloroform. The phage was banded in a CsCl step gradient (1.7 g/ml, 1.5 g/ml, 1.45 g/ml in 0.02 mM Tris-HCl pH 7.5, 0.01 M $MgSO_4$) at 27,000 rpm for 2 hrs and then further purified on a continuous CsCl gradient by overnight centrifugation at 35,000 rpm. The portions containing the phage were digested with proteinase K (0.5 mg/ml) for 1 hr during dialysis against 20 mM Tris-HCl, pH 7.5, 5 mM EDTA. The DNA was extracted three times with phenol and dialyzed against 10 mM Tris-HCl pH 7.5, 1 mM EDTA. Usually a yield of 150 µg DNA per liter of culture was obtained.

Sub-cloning of λgCOL 204 into pBR322 and isolation of DNA from plasmid

Subcloning and isolation of DNA from plasmid was done as published (5). Phage DNA was digested with Hind III restriction enzyme and fragments were ligated into pBR322 and transformed into E. coli C600. One such clone, pgCOL 28-1.7, screened by colony hybridization, which contains a 1.5 Kb Hind III insert, was used for sequence analysis.

Nick translation of DNA

Nick translation (14) of the DNA was done with DNA Polymerase I (Boehringer) dCTP, and dGTP labelled with ^{32}P at the highest specific activity available (Amersham); no additional DNase was added. The incubation at 25°C was for 60 min if the DNA was used for hybridizaing to nitrocellulose or DBM filters; for 15 min if the labelled DNA had to be separated on an agarose gel for the double-transfer method.

Electron microscopy

Heteroduplexes were formed by hybridizing denatured λgCOL 204 and λCharon 4A DNA at 25°C for 12 hrs in 50% formamide, 10 mM Tris HCl, pH 7.8. RNA, enriched for chicken collagen mRNA, was hybridized to the heteroduplexes at 52°C for 3 hrs in 70% formamide, 0.5 M NaCl, 0.001 M EDTA, 0.1 M Tricine pH 8.0. The nucleic acid preparation was spread with Cytochrome c (15) and analyzed in a Phillips 300 EM. SV40 DNA was used as the internal length standard.

Analysis of fragments generated by digestion with restriction endonucleases

Digestions with the different restriction endonucleases were done as recommended by the supplier (Biolabs). The restriction fragments were analyzed on a 0.8% agarose (Sigma) gel in E-Buffer (16). The DNA fragments were transferred to either nitrocellulose filters (17) (Schleicher and Schuell) or DBM paper (18) (Schleicher and Schuell). The double-transfer method described by Pero et al. (19) was used to determine the relative location of the various restriction fragments. In such experiments 15 µg of λgCOL 204 DNA was restricted and separated on an agarose gel of 10 x 10 cm. The fragments were transferred in the usual manner to a DBM paper (18) and the paper was treated as described to saturate all further available DNA binding sites. 2 µg of λgCOL 204 DNA was restricted with a different restriction endonuclease, labelled by nick translation and the DNA fragments fractionated on a 10 x 10 cm agarose gel. This latter gel was transferred to the first blot at a 90° angle relative to the direction of migration in the first gel. The transfer was done at 68°C with 4 x SSC, 0.1% SDS. Where two restriction fragments which have sequences in common meet, the radioactive fragment hybridized to the fragment covalently bound to the DBM paper. After autoradiography the DBM paper can be reused after removing the labelled DNA by washing the paper at 70°C in 0.1 x SSC, 0.1% SDS.

Isolation of purified RNA

Total cellular RNA was isolated from calvaria and long bones and from chicken embryo fibroblasts (CEF) using 8M guanidine as described previously (20). Poly (A)-containing RNAs were selected by one or two cycles of chromatography on oligo (dT)-cellulose and then subjected to sucrose gradient centrifugation.

Isolation and translation of mRNA hybridizing to λgCOL 204 DNA

50 µg of λgCOL 204 was cleaved with the restriction endonuclease Hind III before denaturation and binding to DBM cellulose. The isolation and translation of mRNA hybridizing to λgCOL 204 was done as published previously (21).

Characterization of λgCOL 204 DNA by hybridization to RNA

Total RNA was separated on methyl mercuric hydroxide agarose gels, transferred to DBM paper, and hybridized with nick-translated λgCOL 204 as published previously (18,21).

DNA sequencing

The 1.5 Kb Hind III fragment which is located close to the 5' side of the genomic insert (Fig. 3) was cut with the restriction endonuclease Ava II. The fragments were end-labeled with polynucleotide kinase and [^{32}P] α-ATP and then

separated on a 10% polyacrylamide gel (22). A 220 bp fragment was eluted, the strands separated and both strands were sequenced following the method of Maxam and Gilbert (27). The entire sequence of this fragment will be published elsewhere.

RESULTS

We used a 200 bp cloned cDNA containing sequences corresponding to the 3' end of α2 collagen mRNA as hybridization probe to screen a chick genomic library. This library, which was constructed by Dodgson, Axel, and Engel according to the method of Maniatis et al. (11), using λCharon 4A as vector, contains chicken DNA fragments of approximately 15 to 22 Kb length. The fragments were generated by partial digestion of chick reticulocyte DNA: one-half of the DNA was digested with Alu I endonuclease and the other half with Hae III endonuclease. Five independent clones were isolated after screening about 10^6 plaques. After preliminary R-loop analysis, one clone was chosen for further study because it contained the largest segment hybridizing with collagen RNA. This clone was designated λgCOL 204 and was characterized by electron microscopy and mapped by restriction enzyme analysis.

Electron microscope analysis

Analysis of R-loops by electron microscopy gives information on the length and the distribution of coding and intervening sequences in genomic DNA. To favor the hybridization of collagen RNA to single-stranded DNA and, hence, visualize more clearly the coding and intervening sequences, we first formed a heteroduplex between λgCOL 204 DNA and the DNA of the parent λCharon 4A. The single-stranded regions so created were then hybridized with RNA enriched for Type I Collagen mRNA under R-loop conditions. A typical picture is shown in Fig. 1.

There are eight coding sequences separated by seven intervening sequences. Measurements of 20 recombinant molecules are summarized in Table 1. Fig. 3A presents a diagrammatic summary of these measurements. The eight coding sequences are clustered toward the long arm of the λCharon 4A vector leaving around 9 Kb toward the smaller arm of λCharon 4A which do not appear to contain coding sequences. λgCOL 204 has around 1800 bp of coding sequences and around 5000 bp of intervening sequences. λgCOL 204 contains about one-third of coding sequences of the α2 collagen gene because the total size of α2 collagen mRNA and, hence, of the coding sequence in the α2 collagen gene is about 5000 nucleotides (21).

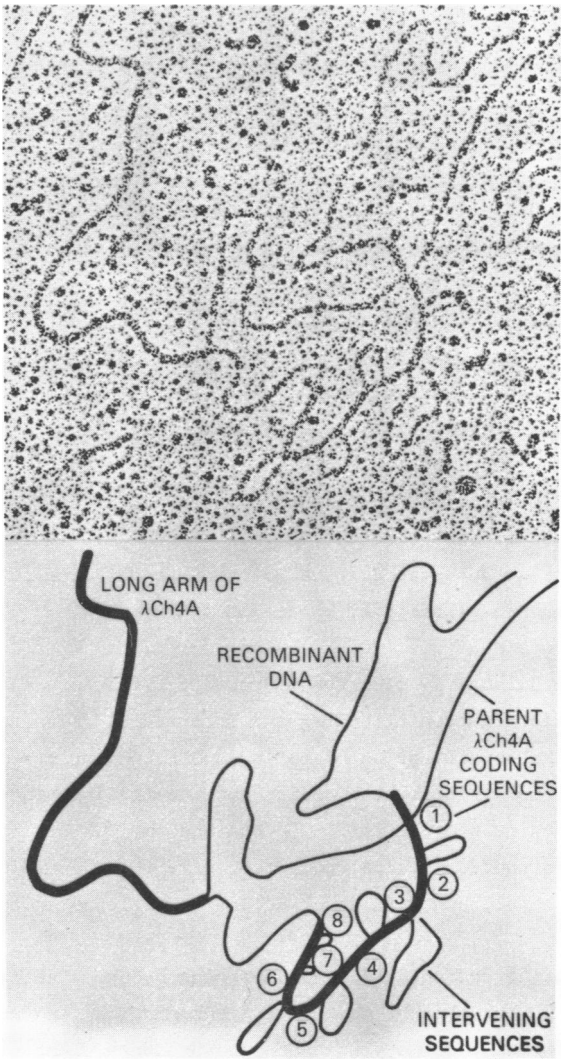

Fig. 1. Electron microscopy of λgCOL 204 λgCOL 204 DNA was hybridized with λCharon 4A DNA. In this heteroduplex the common Charon 4A sequences (the long arm and the short arm) hybridize, whereas the recombinant DNA of λgCOL 204 is rendered single stranded. To this heteroduplex, RNA enriched for Type I collagen was hybridized. The eight coding sequences are numbered.

Restriction map

λgCOL 204 DNA was digested with several restriction endonucleases. Fig. 2A gives the pattern of the fragments on an agarose gel after digestion with

Table I: Coding and Intervening Sequences in λgCOL 204 (α2 collagen)

Coding Sequence No	Mean Length bp	Standard Deviation bp	Standard Error bp
1	409	133	31
-	614	103	23
2	260	47	10
-	1262	126	26
3	202	44	9
-	452	57	13
4	254	61	13
-	456	72	15
5	183	41	9
-	673	59	12
6	135	36	8
-	312	76	18
7	136	108	27
-	203	58	14
8	214	46	11
-	1106	144	43

20 molecules have been analyzed. The coding sequences are numbered as in Fig. 1. The non-transcribed region at the 3' side of the gene (beyond coding sequence 1) extends 9000 bp to the small arm of λCharon 4A.

Hind III (lane 2) and Eco R1 (lane 4). The DNA on the gel was transferred to nitrocellulose filters and then hybridized with the nick-translated cloned α2 collagen cDNA probe used in screening the library. Lane 3 is an autoradiogram of the Hind III digest of lane 2, and lane 5 is an autoradiogram of the Eco R1 digest of lane 4. The cloned cDNA probe hybridizes to two Hind III fragments of 3.8 Kb and 6.7 Kb. Since this cloned cDNA does not contain a Hind III site, its sequence must be interrupted in the genome by an intervening sequence. Fig. 2A, lane 5 indicates that the cloned cDNA segment hybridizes only to one Eco RI fragment of 3.3 Kb.

The order of the restriction fragments was established by the double-transfer method (see Methods). Non-radioactive λgCOL 204 DNA was digested by Hind III and transferred to DBM paper. ^{32}[P] labeled λgCOL 204 DNA was digested by Eco RI, fractionated, and transferred at a right angle to the DBM sheet containing the unlabeled DNA under appropriate hybridization conditions. The radioactive Eco RI fragments hybridized to the non-radioactive Hind III

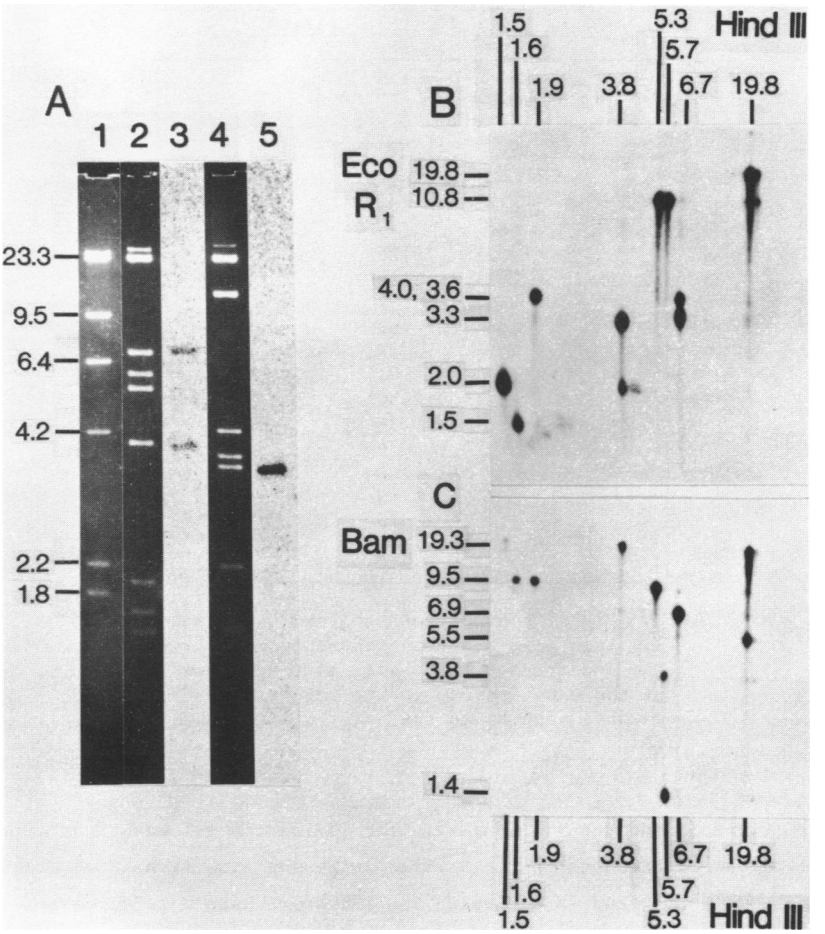

Figure 2. Restriction analysis of λgCOL 204 DNA Fig. 2A: Lane 1: Size markers, λDNA digested with Hind III restriction endonuclease. Lane 2: λgCOL 204 DNA digested with Hind III. Lane 2: Autoradiogram from Lane 2, hybridized with the nick-translated insert from pCOL 1. Lane 4: λgCOL 204 DNA digested with Eco R1. Lane 5: Autoradiogram from Lane 4, hybridized with the nick-translated insert from pCOL 1.

Fig. 2B: Autoradiogram of a double transfer of Hind III restriction fragments against nick-translated Eco R1 fragments.

Fig. 2C: Autoradiogram of a double transfer of Hind III restriction fragments against nick-translated Bam H1 fragments of λgCOL 204.

The numbers indicate length of restriction fragments in Kb.

fragments where any two restriction fragments meet that have sequences in common. For instance, the 3.3 Kb Eco RI fragment has sequences in common with the 6.7Kb and the 3.8Kb Hind III fragments; hence, the 6.7 and 3.8 Kb Hind III fragments are contiguous. The data presented in Fig. 2B and 2C, together with the known location of Eco Rl, Bam HI, and Hind III sites in the λ portion of λgCOL 204, enabled us to deduce the order of restriction fragments presented in Figure 3B. Examination of both the R loop measurements and the restriction map indicates that the coding sequences must, therefore, lie with-

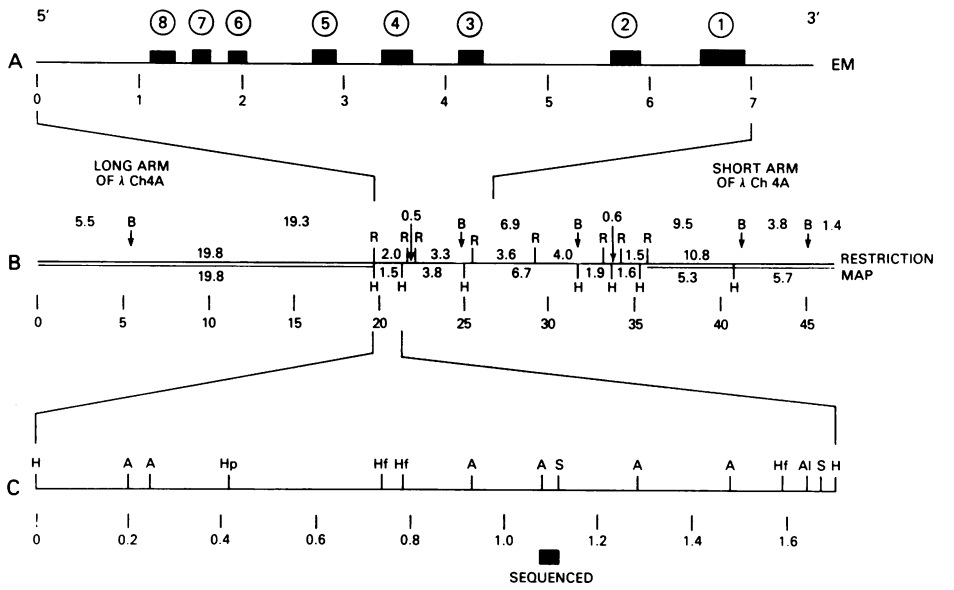

Figure 3. Restriction map of λgCOL 204 Fig. 3A: Diagram of the data from Table 1; the black boxes indicate coding sequences. The numbering of the coding sequences was begun at the 3' end of the gene, since that was the order in which they were identified.

Fig. 3B. Map of the restriction fragments of λgCOL 204. The double line indicates the sequence of the parent phage λgCharon 4A. The exact site of the various restriction fragments is given in Fig. 2.

Fig. 3C: Map of the restriction fragments of the 1.5 Kb Hind III fragment. The sequence between the Ava II and the Sau 3A cleavage site is reported in Fig. 5. A detailed analysis by polyacrylamide gel electrophoresis indicated that the 1.5 Kb Hind III fragment contained approximately 1700 bp. This discrepency could be due to different mobilities in agarose compared to polyacrylamid gels.

Abbreviations: Numbers in circle: coding sequence corresponding to numbers in Fig. 1 and Table I; all other numbers indicate sizes in Kb. The following restriction endonucleases have been used: B: Bam Hl, R: Eco Rl, H: Hind III, A: Ava II, Hp: Hap II, Hf: Hin fII, S: Sau 3A, Al: Alu I.

in the Hind III restriction fragments of 1.5 Kb, 3.8 Kb, and 6.7 Kb.

Hybridization of λgCOL 204 DNA to collagen RNA

Our previous results have indicated that the cloned cDNA fragment of pCOL 1, specifically hybridizes to α2 collagen mRNA (7). This 5000-nucleotides-long mRNA species is found in chick embryo fibroblasts (CEF) but its levels are greatly reduced in CEF transformed by Rous Sarcoma Virus. Fig. 4B shows that λgCOL 204 DNA also hybridizes to a RNA that is about 5000 bases long (lane 2) and is thus the size of α2 collagen mRNA. Like pCOL 1 DNA, λgCOL 204 DNA also hybridizes to a RNA which is larger. It has been argued previously that this might be a precursor for α2 collagen mRNA (21). As with pCOL 3, λgCOL 204 DNA does not hybridize to RNA from Rous Sarcoma Virus infected chick embryo fibroblasts (lane 3).

Isolation and translation of α2 collagen mRNA by hybridization to λgCOL 204 DNA

λgCOL 204 DNA was restricted with Hind III restriction endonuclease, denatured and covalently bound to DBM cellulose. The derivatized cellulose powder was hybridized to total poly(A)-containing RNA from calvaria and long bones. After washing off the unhybridized RNA, the bound mRNA was eluted and translated in an in vitro reticulocyte system. This RNA directs the synthesis of one polypeptide (Fig. 4A, lane 3) which comigrates on SDS acrylamide gels with pro-α2 collagen. The only other bands which can be detected are also synthesized by the reticulocyte system in the absence of added RNA (Fig. 4A, lane 1).

Sequence analysis

To prove that the isolated λgCOL 204 contains α2 sequences, we sequenced part of the DNA and compared this sequence with the known amino acid sequence for pro-α2 collagen. We chose to sequence a segment of the 1.5 Kb Hind III fragment next to the long arm of λCharon 4A (Fig. 3B).

Comparison of the R-loop data and the results of the restriction enzyme analysis suggests that the 1.5 Kb Hind III fragment contains two coding sequences. The 1.5 Kb Hind fragment was subcloned in pBR322 and designated pgCOL 28-1.7. A restriction map of the fragment was determined and is presented in Fig. 3C. A detailed analysis by polyacrylamide gel electrophoresis showed that the 1.5 Kb Hind III fragment contains approximately 1700 bp. We have not corrected the data in Fig. 2 and Fig. 3B for this discrepancy which might be due to different mobilities of restriction fragments in agarose compared to acrylamide gel dependent on the sequence. Partial cleavage with Eco RII endonuclease indicated several neighboring Eco RII sites at two loca-

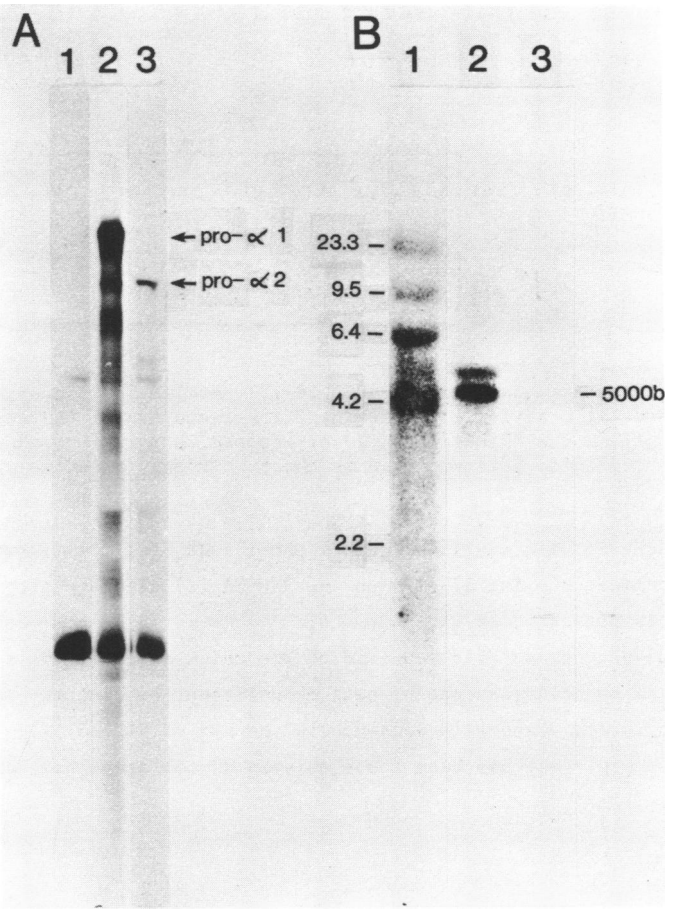

Figure 4A: Translation of mRNA hybridizing to λgCOL 204. Autoradiogram of proteins synthesized in a reticulocyte system with ^{35}S methionine and separated by SDS acrylamide gel electrophoresis. Lane 1: No RNA added. Lane 2: Protein synthesis with total RNA added. Lane 3: Protein synthesis with RNA hybridized to λgCOL 204.

Figure 4B: Hybridization of nick-translated λgCOL 204 to mRNA from chicken embryo fibroblasts separated on methyl mercuric hydroxide agarose gels. Lane 1: λwt digested with Hind III used as size marker, numbers indicate size in Kb. Lane 2: RNA from chicken embryo fibroblasts. Lane 3: RNA from chicken embryo fibroblasts infected with Rous sarcoma virus.

tions in the 1.5 Kb fragment (results not shown). Eco RII sites might be expected in regions where a proline codon (CCX) would be adjacent to a glycine codon (GGX) since the Eco RII endonuclease recognition site is 5'CCTGG 3'.

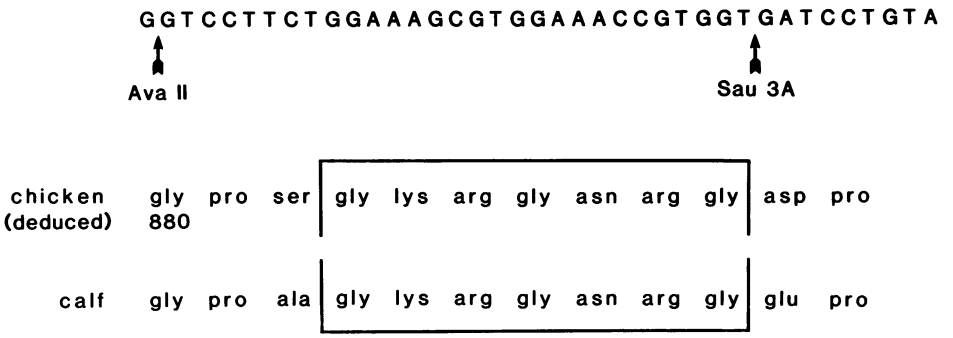

Figure 5. Sequence analysis of pgCOL 28-1.7 Nucleotide sequence of a part of the insert at position 1.1 Kb (Fig. 3C) between Ava II (A) and Sau 3A (S). The first gly in the chicken pro-α2 corresponds to amino acid 880 of the mature collagen. The amino acid sequence in the box is only found in α2 collagen.

The DNA sequence of a small segment of the 1.5 Kb Hind III fragment at position 1.1 between the Ava II (A) and the Sau 3A (S) cleavage site was determined and is presented in Fig. 5.

The deduced amino acid sequence of the chick α2 collagen is compared with the known amino acid sequence of calf α2 collagen (P. Fietzek, personal communication). The amino acid sequence in the box of Fig. 5: gly - lys - arg - gly - asn - arg - gly has been found only in α2 collagens but not in other collagens.

DISCUSSION

We have isolated and characterized the recombinant Charon 4A phage λgCOL 204 which contains DNA sequences for the chick α2 collagen. Evidence that this DNA contains part of the α2 collagen gene is as follows: 1) λgCOL 204 hybridizes to an RNA species the size of α2 collagen messenger RNA. This RNA is present in normal chick embryo fibroblasts (CEF) but not in CEF transformed by Rous Sarcoma Virus. We have previously shown that the levels of α2 collagen mRNA are reduced in RSV transformed CEF (21). 2) λgCOL 204 hybridizes to a unique mRNA which directs the synthesis of a polypeptide the size of pro-α2 collagen. This polypeptide has been shown to be pro-α2 collagen (7). 3) A partial DNA sequence of a Hind III 1.5 Kb fragment gives a deduced amino acid sequence which corresponds to a sequence that has only been found in α2 collagen.

The 6000 bp of the α2 collagen gene which are present in λgCOL 204 must

be located at the 3' end of the α2 collagen gene, because the cloned cDNA probe (pCOL 1, 7) used to detect the genomic clone is derived from the 3' end of the α2 collagen mRNA. DNA sequence analysis of pCOL 1 (Yamamoto, T., personal communication) indicates that it does not contain sequences corresponding to the central helical region of α2 collagen. The location of pCOL 1 at the 3' end is in agreement with the data of Lehrach et al., who found a similar DNA fragment within a larger cDNA clone (8). Hence, the 3.8 Kb and the 6.7 Kb Hind III fragments to which the cloned pCOL 1 cDNA hybridizes contain coding sequences corresponding to the C-terminal portion of collagen. Since the 1.5 Kb Hind III fragment contains coding sequences corresponding to the helical part of the α2 collagen, the orientation of the α2 collagen gene within λgCOL 204 is as given in Fig. 3A. The 9kb DNA segment between the last 3' proximal coding sequence and the short arm of λ is, thus, a sequence which flanks the α2 collagen gene at its 3' end. We do not know what gene lies at the 3' side of the α2 collagen gene or the size of the intergenic distance. Preliminary experiments (F. Eden, personal communication) indicate that the DNA at the 3' side of the chick α2 gene contains a short repetitive squence of low-repetitive frequency.

The recombinant phage λgCOL 204 contains, within its 8 coding sequences, a total of 1800 bp coding for α2 collagen. The intervening sequences make up at least 5000 bp. The coding regions towards the 3' end are larger than the coding regions in the helical portion of the protein. How many base pairs from the 3' end code for amino acids and how many are from the non-translated 3' region of the mRNA is not yet known. Sequence analysis and comparison with the protein structure will establish which part of the mRNA is not translated into protein.

Since the mRNA for α2 collagen has about 5000 bases, we assume that the whole gene is around 25-35 kb long and contains about 50 intervening sequences. We have recently isolated several other recombinant clones which cover most of the remaining DNA of the α2 collagen gene. The analysis of these clones confirms the existence of numerous intervening sequences within the α2 collagen gene. The continuity of the coding sequences of the α2 collagen gene is interrupted by intervening sequences, as are other eukaryotic genes which have been examined (15, 23-27). It is interesting to note that the fibroin gene of Bombyx mori, which also contains repeating glycine codons and is 16 kb long, has only one intervening sequence of 1.1 kb near the 5' end of the gene (24).

It has been proposed (29) that the different coding sequences of a gene

correspond to functional domains of the polypeptide. According to this model, a gene or a unit of transcription could be assembled during evolution from different genetic segments. Such an assemblage would make genes evolve faster if coding sequences were embedded in much larger non-coding (intervening) sequences, provided mechanisms exist to splice the mRNAs of different coding sequences together. Once a gene has attained an evolutionary favorable dimension, the intervening sequences could decrease the probability of illegitimate recombination between genes with similar coding sequences but with different types of regulation.

Acknowledgement

We would like to thank Drs. T. Maniatis, F. Blattner, and L. Enquist for gifts of materials; Drs. J. Seidman and L. Enquist for advice; and Mr. R. Steinberg for help with illustrations.

*Abbreviations

α2 collagen, α2 type I collagen; SSC, 0.15M NaCl, 0.015 M Na citrate; DBM, diazobenzyloxymethyl; bp, base pairs; SDS, sodium dodecyl sulfate; Kb, kilo base pairs.

REFERENCES

1. Ramachandran, G.N., and Reddi, A.H., eds., (1976) Biochemistry of Collagen, Plenum Press, New York
2. Levinson, W., Bhatnagar, R.S., and Liu, T.-Z. (1975) J. Natl. Cancer Inst. 55, 807-810
3. Howard, B.H., Adams, S.L., Sobel, M.E., Pastan, I., and de Crombrugghe, B. (1978) J. Biol. Chem. 253, 5869-5874
4. Mayne, R., Vail, M.S., and Miller, E.J. (1975) Proc. Natl. Acad. Sci. USA 72, 4511-4515
5. Yamamoto, T., Sobel, M.E., Adams, S.L., Avvedimento, E.V., DiLauro, R., Pastan, I., de Crombrugghe, B., Showalter, A., Pesciotta, D., Fietzek, P., Olsen, B. (1980) J. Biol. Chem. (in press)
6. Lehrach, H., Frischauf, A.M., Hanahan, D., Wozney, J., Fuller, F., Boedtker, H. (1979) Biochemistry 18, 3146-3152
7. Sobel, M.E., Yamamoto, T., Adams, S.L., DiLauro, R., Avvedimento, E.V., de Crombrugghe, B., and Pastan, I. (1978) Proc. Natl. Acad. Sci. USA 75, 5846-5850
8. Lehrach, H., Frischauf, A.M., Hanahan, D., Wozney, J., Fuller, F., Crkvenjakov, R., Boedtker, H., and Doty, P. (1978) Proc. Natl. Acad. Sci. USA 75, 5417-5421
9. Dodgson, J.B., Strommer, J., and Engel, J.D. (1979) Cell 17, 879-887
10. Benton, W.D. and Davies, R.W. (1977) Science 196, 180-182
11. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, E., (1978) Cell 15, 687-701
12. Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun. 23, 641-646
13. Laskey, R.A. and Mills, A.D. (1975) Eur. J. Biochem. 56, 335-341

14. Maniatis, T., Jeffrey, A., and Kleid, D.G., Proc. Nat. Acad. Sci. USA (1975) 72, 1184-1188
15. Tilgham, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P., and Weissman, C. (1978) Proc. Natl. Acad. Sci. USA 75, 1309-1313
16. Loening, U.E., Biochem, J. (1969), 113, 131-138
17. Southern, E.M., J. Mol. Biol. (1975) 98, 503-517
18. Alwine, J.C., Kemp, D.J., and Stark, G.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5350-5354
19. Pero, J., Hannett, N.M., Talkington, C. (1979) J. Virology 31, 156-171
20. Adams, S.L., Sobel, M.E., Howard, B.H., Olden, K., Yamada, K.M., de Crombrugghe, B., and Pastan, I. (1977) Proc. Natl. Acad. Sci. USA 74, 3399-3403
21. Adams, S.L., Alwine, J.C., de Crombrugghe, B., and Pastan, I. (1979) J. Biol. Chem. 254, 4935-4938
22. Maxam, A.M., Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
23. Royal, A., Garapin, A., Cami, B., Perrin, F., Mandell, J.L., LeMeur, M., Bregegegre, F., Gannon, F., LePennec, J.P., Chambon, P., Kourilsky, P., (1979) Nature 279, 125-132
24. Tsujimoto, Y., Suzuki, Y. (1979) Cell 16, 425-436
25. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., and Chambon, P. (1979) Nature 282, 567-574
26. Tucker, P.W., Marcu, K.B., Blattner, F.R. (1979) Science 206, 1303-1306
27. Max, E.E., Seidman, J.G., and Leder, P. (1979) Proc. Natl. Acad. Sci. USA 76, 3450-3454
28. Gilbert, W. (1979) Nature, 271, 501