Volume 4 Number 7 July 1977 Nucleic Acids Research

Molecular cloning of extensive sequences of the in vitro synthesized chicken ovalbumin structural gene

Peter Humphries*, Madeleine Cochet[1], Andrée Krust[1], Pierre Gerlinger[2], Philippe Kourilsky[3] and Pierre Chambon[1]**

[1]Lab. de Génétique Moleculaire des Cellules Eukaryotes du CNRS, U44 de L'INSERM et Inst. de Chimie Biologique, Fac. de Méd., 67085 Strasbourg Cédex, [2]Inst. d'Embryologie, Fac. de Méd., 67085 Strasbourg Cédex, [3]Dep. de Biologie Moléculaire, Inst. Pasteur, Paris, France.

Received 4 May 1977

ABSTRACT

Double-stranded DNA molecules complementary to ovalbumin chicken messenger RNA were synthesized *in vitro* and integrated into the E.coli plasmid pCR1 using an oligodG-dC tailing procedure. The resultant hybrid plasmids, amplified by transfection of E.coli, were shown by hybridization and gel electrophoresis to contain extensive DNA sequences of the ovalbumin structural gene.

INTRODUCTION

The expression of the ovalbumin gene is under hormonal control (for a review, see ref. 1) and its isolation would be very valuable for studies on the way in which steroid hormones, histones, RNA polymerases and other non-histone chromosomal proteins, interact with DNA to control higher eukaryotic transcription. Such studies would be facilitated if it were possible to obtain in large amounts, fragments of DNA containing the structural ovalbumin gene plus its flanking putative regulatory sequences. There are two interrelated methods of molecular cloning by which such fragments may at present be obtained (reviewed in ref. 2 and 3). The first, which has recently become known as "shotgun cloning" involves the introduction into a bacterium of fragments of DNA of an entire genome, followed by selection by hybridization of those bacteria harbouring the gene or genes of interest. The second approach, perhaps of lower potential biological risk, involves the preliminary purification of a gene-containing fragment prior to its amplification within the bacterium. The object of the work described here has been to amplify within E.coli, specific sequences of the ovalbumin structural gene so that these can be obtained in large amounts for studying

the ovalbumin gene organization and for use as a tool for selection of larger DNA fragments containing the ovalbumin structural gene plus its possible regulatory sequences.

MATERIALS AND METHODS

1. Enzymes.

Avian myeloblastosis virus (AMV)[a] reverse transcriptase, prepared according to the procedure of Kacian and Spiegelman (4) was kindly provided by Drs. Beard and Houts. Calf thymus terminal deoxynucleotide transferase was prepared according to the procedure of Bollum (5), with minor modifications. E.coli DNA polymerase I (fragment "A", ref. 6) was purchased from Boehringer. Restriction endonuclease EcoRI was prepared according to Green et al. (7). Restriction endonucleases, R.HpaII and R.HaeIII, were purchased from BRL (Bethesda Research Laboratories). Restriction endonuclease digestions were carried out according to the recommendations of the BRL booklet. Nuclease S1 from *Aspergillus oryzae* was purified according to the procedure of Vogt (8).

2. Synthesis of double-stranded, oligodG-elongated ovalbumin cDNA.

Ovalbumin synthesizing polysomes from laying hens were purified by an indirect immunoprecipitation technique with rabbit anti-ovalbumin and goat anti-rabbit γ-globulin (9). Immunoprecipitated polysomes [more than 98% of the mRNA present in the polysomes is ovalbumin mRNA (9)] were digested with proteinase K (Merck) and the RNA was extracted with phenol-chloroform. Polysomal RNA containing the ovalbumin messenger RNA was used directly for cDNA synthesis. cDNA was synthesized in a reaction mixture containing 50 mM Tris-HCl buffer pH 7.5, 10 mM $MgCl_2$, 2 mM dithiothreitol, 200 µg/ml actinomycin D, 36 µg/ml oligodT_{12-18} (Collaborative Research), 4.8 mg/ml immunoprecipitated polysomal RNA, 120 µM [^{32}P]dCTP (Amersham), 400 µM dGTP, dATP and dTTP and 350 units (4) per ml of reverse trascriptase. After incubation at 41°C for 1 h the reaction mixture was made 0.3 M in NaOH and incubated for a further 15 min at 68°C in order to hydrolyse the RNA. After neutralization with HCl, the reaction mixture was passed through a column of Sephadex G-50 and

the cDNA was precipitated with ethanol in the presence of 30 µg/ml of E.coli tRNA (Boehringer) carrier. The precipitate was redissolved in 200 µl of 0.2 M NaOH, 0.8 M NaCl and 5 mM EDTA and layered onto the surface of a 5-20% alkaline sucrose gradient in the same solution. After centrifugation at 39,000 rpm for 14 h in a Spinco SW60 Ti rotor, the fastest-sedimenting portion of the cDNA peak (about 30% of the total cDNA) was selected and precipitated with ethanol.

600 ng of cDNA was then incubated with 25 units of E.coli DNA polymerase I (fragment "A", ref. 6) in a reaction mixture (4.2 ml) containing 30 mM Tris-HCl buffer pH 7.5, 4 mM $MgCl_2$, 0.5 mM β-mercaptoethanol and 200 µM dATP, dCTP, dGTP and dTTP. After incubation at 30°C for 60 min, the reaction mixture was extracted twice with phenol and dialyzed extensively against H_2O, during which time the volume of the reaction mixture increased to 6.0 ml. The reaction was then made 3 mM in $ZnCl_2$, 30 mM in sodium acetate (pH 4.5) and 300 mM in NaCl, and a sufficient quantity of nuclease S1 was added to render 600 ng of single-stranded cDNA acid-soluble within 40 min. After incubation at 40°C for 80 min the reaction was extracted twice with phenol, dialyzed extensively against H_2O and lyophilized to residue. Assay of the acid precipitability of the original [^{32}P]-labelled cDNA strand before and after nuclease S1 treatment showed that approximately 50% of the cDNA had been rendered double-stranded under the above conditions. The lyophilized double-stranded cDNA (ds-cDNA) was resuspended in 250 µl of 10 mM Tris-HCl buffer pH 7.5 containing 0.8 M NaCl, 8 mM EDTA and layered onto the surface of a 5-20% neutral sucrose gradient in the same buffer. The DNA was centrifuged for 13.5 h at 30,000 rpm in a Spinco SW60 Ti rotor in order to select the fastest-sedimenting DNA molecules which were dialyzed against H_2O. 60 ng of the largest ds-cDNA in 1.6 ml H_2O was made up to 10 mM HEPES buffer pH 7.1 containing 5 mM $MgCl_2$ and 1 mM β-mercaptoethanol; 250 µM dGTP and 200 units of terminal transferase were added (1 unit of activity is defined as that amount of enzyme that will incorporate 10 pmoles of dCTP into acid-insoluble product in 5 min at 37°C in the presence of 0.4 M HEPES buffer pH 7.1, 2 mM $CoCl_2$, 2.5 mM β-mercaptoethanol, 0.13 µg/ml of oligodA_{10} and 1 mM dCTP),

bringing the volume of the final reaction mixture to 2.3 ml. Incubation for 5 min at 37°C resulted in the addition of a sufficient number of residues of dGMP to the termini of the cDNA to allow efficient molecular hybrid formation with oligodC-elongated linear pCR1 DNA as assayed by transfection of E.coli (see below). The reaction was extracted with phenol and dialyzed extensively against H_2O before use in transfection experiments.

3. Preparation of linear, oligodC-elongated pCR1 plasmid DNA.

Covalently closed circular pCR1 plasmid DNA (10) was separated from cellular DNA by equilibrium banding in caesium chloride/ethidium bromide gradients, followed by sedimentation in a neutral 5-20% sucrose gradient. 30 µg of plasmid DNA was incubated with sufficient EcoRI restriction endonuclease to convert the superhelical DNA into linear unit length molecules as assayed by electrophoresis on 1% agarose gels and by direct observation under the electron microscope. After phenol extraction and ethanol precipitation 25 µg of linear plasmid was incubated with 7 units of AMV reverse transcriptase for 70 min at 42°C in a 0.8 ml reaction mixture containing 50 mM Tris-HCl pH 9.0, 400 µM dATP, 400 µM dTTP, 10 mM $MgCl_2$ and 2 mM dithiothreitol in order to fill the cohesive termini left by the EcoRI restriction endonuclease. After phenol extraction and ethanol precipitation, 11 µg of repaired linear plasmid DNA was incubated with 18 units of calf thymus terminal transferase in 150 µl of reaction mixture containing 100 mM HEPES buffer pH 7.2, 1 mM $CoCl_2$, 1.25 mM β-mercaptoethanol and 1.26 mM dCTP. Incubation at 37°C for 20 min resulted in the addition of a sufficient number of residues of dCMP to the termini of the linear plasmid DNA to allow the formation of molecular hybrids with oligodG-elongated ds-cDNA. The elongated plasmid DNA was phenol extracted, ethanol precipitated and dialyzed against the hybridization buffer (see below).

4. Construction of molecular hybrids and transformation of E.coli

22.5 ng of oligodG-elongated ds-cDNA was mixed with 165 ng of linear oligodC-elongated pCR1 plasmid DNA in a volume of 200 µl of 10 mM Tris-HCl buffer pH 8.0, containing 0.1 M NaCl and 1 mM EDTA. The hybridization mixture was heated to 65°C for

2 h and allowed to cool slowly to room temperature. An overnight culture of E.coli C600 r_k^- m_k^- rec B^-C^- (11) was diluted into 160 ml of L-broth and incubated at 37°C until the 630 nm absorbance was 0.6. The cells were then spun down and resuspended in 80 ml of cold 50 mM $CaCl_2$ containing 50 µg/ml of thymidine. After standing for 10 min on ice the cells were again spun down and resuspended in 8 ml of cold 50 mM $CaCl_2$ containing 50 µg/ml thymidine. The cells were kept on ice. Either 20 or 40 µl of the molecular hybrid solution prepared as above was made up to 100 µl with 0.1 M Tris-HCl pH 7.0 and each sample was mixed with 0.2 ml of the bacterial suspension. The cells were allowed to stand on ice for 10 min and then maintained at 37°C for 5 min. 0.5 ml of L-broth was added and the cells were shaken at 37°C for 30-45 min. 0.2 ml aliquots of the cellular suspension was then spread over tryptone plates containing 12 µg/ml of Kanamycin.

RESULTS

1. Construction of hybrid plasmids.

Several procedures have been developed for the integration of cDNA molecules into bacterial plasmids. All of these methods require tailing the linear plasmid and cDNA with short blocks of complementary homopolymer in order to form circular hybrid molecules. The cDNA is made double-stranded either before the construction of hybrid molecules (11, 12, 13) or alternatively by *in vivo* repair following transformation of E.coli by hybrid plasmids (14). In addition cDNA-mRNA hybrid molecules can be successfully integrated into plasmid vectors (15).

The method employed for the integration of ovalbumin cDNA into the E.coli plasmid pCR1 was essentially the same as that used by Rougeon et al. (11) for the molecular cloning of rabbit β-globin cDNA. As described in Materials and Methods and outlined in Fig. 1, single-stranded cDNA was synthesized from purified ovalbumin mRNA using reverse transcriptase. The longest cDNA molecules with size approaching that of ovalbumin mRNA were selected by sedimentation in alkaline 5-20% sucrose gradients. This so-called "full length" cDNA has a length of approximately 1930 nucleotides as determined by electrophoresis on 3% polyacrylamide gels containing 99% formamide (see Fig. 6). "Full length" cDNA was made

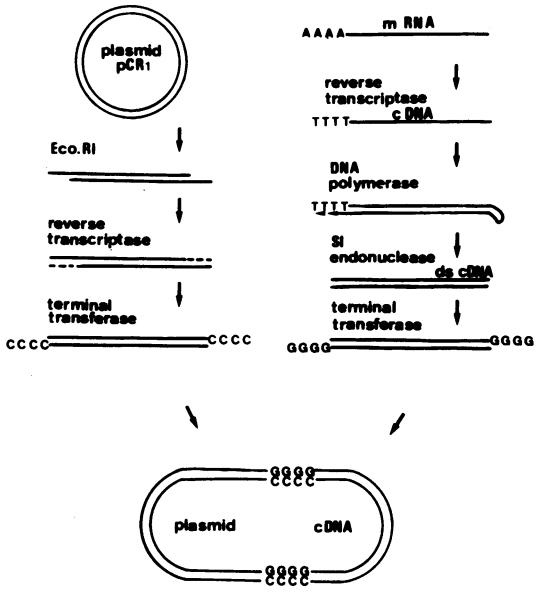

FIGURE 1 Schematic diagram illustrating the synthesis of ds-cDNA from purified chicken ovalbumin mRNA and the construction of ovalbumin plasmid DNA hybrids

double-stranded by incubation with fragment A of E.coli DNA polymerase I (6). As observed by others (12, 16), the cDNA was able to serve as a template for DNA polymerase without the addition of a primer, although only approximately 50% of the cDNA could be made double-stranded as assayed by resistance of the parental strand to nuclease S1. Non-primer dependent synthesis of ds-cDNA is presumably due to the ability of the 3'-OH terminus of the cDNA to fold back upon itself to form a small hairpin structure (16, 17). This terminal hairpin, together with any other single-stranded regions were then removed by incubation with nuclease S1. The ds-cDNA was found to sediment through a 5-20% neutral sucrose gradient as a wide peak with an average length of about 1000 nucleotide pairs (not shown). In order to maximize the probability of obtaining clones containing "full-length" ds-cDNA, only the fastest sedimenting portion of the ds-cDNA peak was selected. To each 3' OH-terminus of this material oligodG was added using calf thymus terminal transferase (18, 19).

In order to serve as a molecular cloning vehicle, the E.coli plasmid pCR1 (11)was cleaved at its single EcoRI restriction target, and the short cohesive termini of the linear molecu-

le were reparied using reverse transcriptase (Fig. 1 and Materials and Methods). The object of this step was to provide more accessible 3'-OH termini for the subsequent action of terminal transferase, and also to attempt to reconstitute the EcoRI restriction sites on either side of the integrated ds-cDNA (11). The linear plasmid was then elongated with oligodC at each 3'-OH terminus using calf thymus terminal transferase. OligodC-elongated linear pCR1 DNA and ovalbumin oligodG-elongated ds-cDNA were hybridized together and used to transform E.coli.

Only circular plasmid molecules are able to infect E.coli and in principle the only means by which an oligodC-elongated plasmid DNA molecule can circularize, is by combining with oligodG-elongated ovalbumin ds-cDNA. Moreover, since the plasmid carries the resistance determinant of Kanamycin, bacteria harbouring recombinant plasmids can readily be selected by plating transformant cultures on agar containing Kanamycin. Complete details of the method used for transformation of E.coli with hybrid plasmids are given in Materials and Methods. Briefly, oligodG-elongated ds-cDNA was hybridized with approximately a sevenfold weight excess of oligodC-extended plasmid DNA. In a typical experiment, hybrids formed between 22.5 ng of ds-cDNA and 165 ng of linear plasmid gave rise to 284 transformants. Experiments employing higher plasmid to ds-cDNA ratios did not produce any additional enhancement of the transformation efficiency. In control experiments, 165 ng of oligodC-extended pCR1 DNA alone gave rise to 16 transformants, while 1.5 ng of superhelical plasmid DNA alone gave rise to 280 transformants. Hence hybrid DNA molecules were able to transform E.coli at about 1% of the efficiency of wild-type plasmid DNA molecules.

2. Analysis of recombinant clones.

A preliminary analysis of the recombinant clones was carried out as described by Rougeon et al. (11). DNA was extracted from colonies inoculated in L-broth and subjected to an overnight incubation in chloramphenicol. The DNA was immobilized on nitrocellulose filters and hybridized with [^{3}H] ovalbumin cDNA. This experiment showed that about 75% of the colonies contained an ovalbumin sequence (not shown).

Since, however, this sort of analysis provides no information as to the length of the integrated sequence, the DNA of some of the colonies was examined with respect to its capacity to protect "full length" ovalbumin cDNA molecules from nuclease S1 digestion. As shown in Table 1, most of the transformants selected showed clearcut evidence of integrated ovalbumin DNA, with levels of protection ranging from more than 5% to 85% of the input cDNA.

Ovalbumin-specific DNA sequences should be integrated into the plasmid at the position of its EcoRI restriction site, and

TABLE 1

Plasmid designation	% protection of "full length" cDNA from nuclease S1 digestion	Plasmid designation	% protection of "full length" cDNA from nuclease S1 digestion
pCR1ov2.1	85	pCR1ov3.20	50
2.5	65	3.21	10
2.6	60	3.22	≤5
2.10	65	3.23	25
2.16	20	3.25	10
2.17	25	3.26	≤5
2.23	35	3.27	50
3.1	≤5	3.28	20
3.2	45	3.29	55
3.3	≤5	3.32	10
3.4	10	3.33	≤5
3.5	15	3.34	15
3.6	30	3.35	50
3.7	70	3.36	≤5
3.8	≤5	3.37	35
3.9	≤5	3.38	10
3.10	≤5	3.39	≤5
3.11	≤5	3.40	32
3.14	60	3.41	≤5
3.15	15	3.42	40
3.16	15	3.43	≤5
3.17	≤5	3.44	45
3.19	≤5		

Total DNA was extracted from 10 ml cultures of 45 hybrid clones. 10 µl of DNA (between 1 and 2 µg) was mixed with 90 µl of 0.3 M NaOH and heated to 100°C for 15 min. The solutions were then neutralized with HCl and made 0.3 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% SDS. Approximately 33 pg of [^{3}H]cDNA (specific activity 15,000 cpm per ng) was added and the solutions were heated at 100°C for 2 min and then at 68°C for 24 h. The mixtures were then incubated with nuclease S1, precipitated with TCA and counted.

two EcoRI sites should have been reconstituted on either side of the integrated ds-cDNA. The DNA of wild-type plasmid pCR1, which is about 13,000 base pairs in length as observed under the electron microscope (unpublished observations) is cleaved by restriction enzyme R.HpaII into a series of fragments which can be resolved into more than 10 bands on 3% polyacrylamide-0.5% agarose gels (Fig. 2). The slowest moving band (see right arrow, Fig. 2, slot c) contains two fragments (not resolved on this print), R.HpaII A and B, which are each approximately 700 base pairs in length when compared with standard HaeIII fragments of SV40 DNA (20) used as markers (slot a). The slower moving fragment, arbitrarily designated fragment A, is cleaved by EcoRI into a fragment (A1) approximately 600 base pairs in length (see left arrow, Fig. 2, slot b) and an additional fragment (A2) which is too small to be detected on this gel. It can be concluded that R.HpaII fragment A of pCR1 DNA contains an EcoRI cleavage site approximately 100 base pairs in from one end, and that this fragment should carry the integrated ovalbumin sequences in hybrid plasmids, and hence should increase in size (see also Fig. 6).

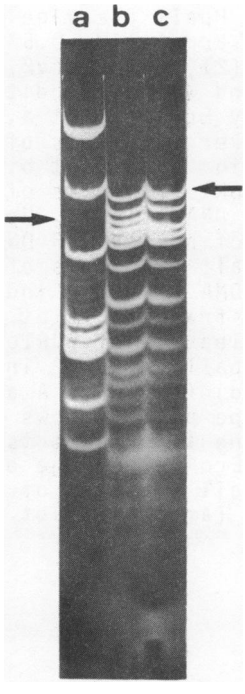

FIGURE 2 Electrophoresis on 3% acrylamide-0.5% agarose gels (21) of endonuclease R.HpaII fragments of wild-type pCR1 DNA with and without further cleavage by EcoRI. Slot a, HaeIII digestion fragments of SV40 DNA used as size markers; slot b, HpaII + EcoRI digestion fragments of pCR1 DNA; slot c, HpaII digestion fragments of pCR1 DNA. The arrows indicate those fragments of DNA which are affected by EcoRI cleavage (see text).

Fig. 3 shows the R. HpaII cleavage pattern of the hybrid clone pCR1ov2.16 on a 3% acrylamide-0.5% agarose gel. This hybrid protects about 20% of the "full length" cDNA from nuclease S1 digestion (Table 1). It will be observed that the plasmid R.HpaII fragment A (slot d, arrow 2) disappears to be replaced in the hybrid DNA molecule by a slower migrating band which is 2000 base pairs in length (Fig. 3, slot c, arrow 1). However, fragment B (see above) also disappears (slot d, arrow 2), presumably because fragments A and B are linked in the hybrid molecule, the HpaII cleavage site connecting the two fragments being either deleted, or rendered non-functional during the integration of ds-cDNA. As shown in Fig. 3, slot b, the new pCR1ov2.16 HpaII fragment is cleaved by EcoRI to produce two smaller fragments of DNA, 1400 (arrow 3) and 600 (arrow 4) base pairs long, respectively. The smaller fragment migrates at the same position as the larger EcoRI cleavage product of R.HpaII wild-type fragment A (Fig. 2, slot b, left arrow). It is concluded that pCR1ov2.16 carries only one reconstituted EcoRI cleavage site. Generally, the strategy of regenerating the EcoRI cleavage sites so that the integrated ds-

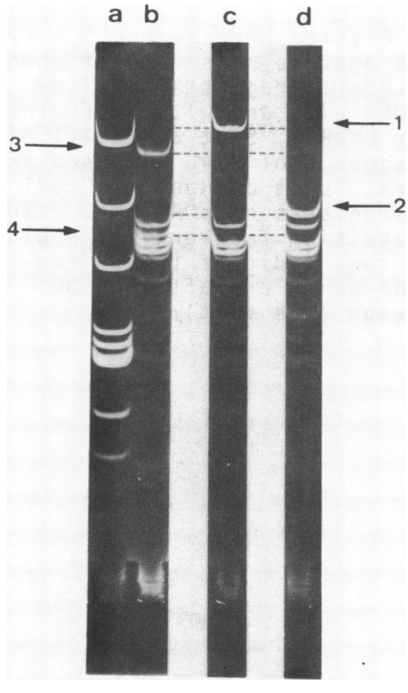

FIGURE 3 HpaII cleavage pattern on 3% acrylamide-0.5% agarose gels (21) of pCR1ov2.16 DNA with and without additional cleavage by EcoRI. Slot a, HaeIII marker fragments of SV40 DNA (see Fig. 2); slot b, HpaII + EcoRI fragments of pCR1ov2.16 DNA; slot c, HpaII fragments of pCR1ov2.16 DNA; slot d, HpaII fragments of pCR1 wild-type DNA. Arrow 1 indicates the larger fragment (slot c) produced by cleavage of pCR1ov2.16 DNA with HpaII. Arrow 2 indicates (slot d) fragments A and B of wild-type pCR1. Arrows 3 and 4 indicate the two fragments obtained by EcoRI cleavage of the largest HpaII fragment of pCR1ov2.16 (arrow 1, slot c).

cDNA could be accurately re-excised, was not sucessful. Out of 10 hybrid plasmids analysed in detail, only pCR1ov2.16 carries one EcoRI cleavage site and none had two EcoRI cleavage sites.

In Fig. 4, the R.HpaII fragment patterns of pCR1ov2.16 and pCR1ov2.1 are compared on 2% agarose gels. In the case of plasmid pCR1ov2.1, the HpaII fragments A and B (slot d, arrow) of wild-type pCR1 are also missing and a much longer new fragment of about 3400 base pairs is obtained (Fig. 4, slot c). Since the original R.HpaII fragments A and B are each approximately 700 base pairs long, it can be concluded that in pCR1ov2.1 the length of the integrated sequence is in the region of 2000 base pairs.

The most likely interpretation of the HpaII mapping data is summarized in Fig. 5. Ovalbumin-specific DNA sequences are integrated in the longest HpaII fragment A. During the transfection the repaired EcoRI sites are generally not conserved and in addition the HpaII site nearest to the site of ds-cDNA integration becomes non-functional. Additional support of those conclusions is given by the observation that the longer HpaII fragments of pCR1ov2.1 is as efficient as total pCR1ov2.1 DNA in pro-

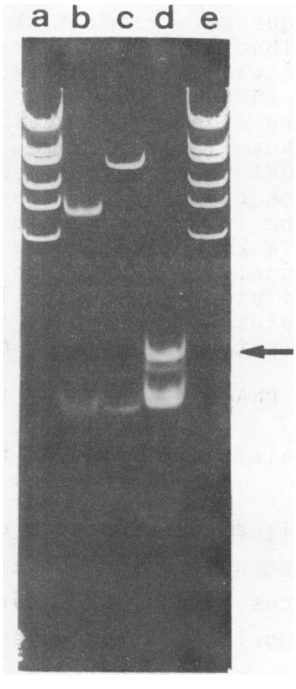

FIGURE 4 Analysis of 2% agarose gels of HpaII digests of pCR1, pCR1ov2.16 and pCR1ov2.1 DNA. Slots a and e, EcoRI fragment markers of adenovirus type 2 DNA (22), slots b, c and d, HpaII fragments of pCR1ov2.16, pCR1ov-2.1 and pCR1 wild type DNA, respectively. The arrow points to HpaII A and B fragments of pCR1 wild-type DNA.

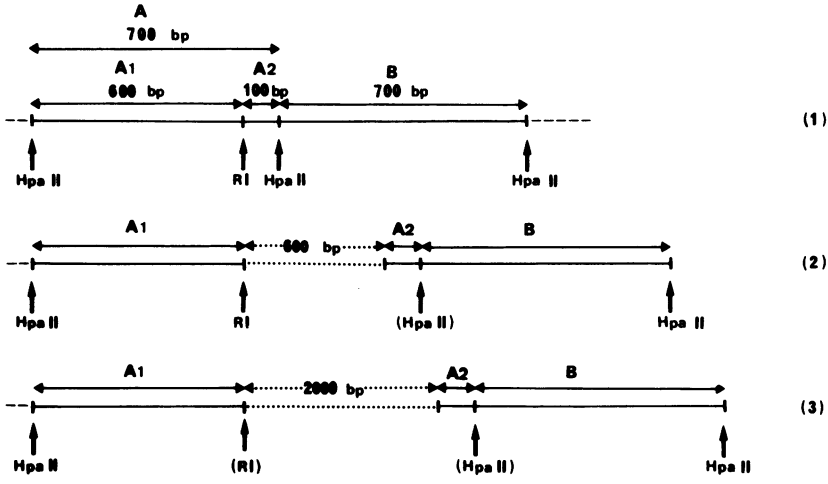

FIGURE 5 Possible topography of the integrated ovalbumin ds-cDNA. 1) The topography of two R.HpaII largest fragments of pCR1 wild-type DNA. The single EcoRI cleavage site of pCR1 DNA is located approximately 100 base pairs from the right hand cleavage site of R.HpaII fragment A. Cleavage of R.HpaII fragment A results in the production of two fragments, A1 and A2, which are approximately 600 and 100 base pairs (Fig. 2). 2) The topography of the R.HpaII fragments of pCR1ov2.16 DNA containing the integrated ovalbumin DNA sequence. Fragment A1 (600 base pairs) lies on the left of the ovalbumin DNA while fragment A2 lies to the right. One EcoRI cleavage site, located to the left of the integrated ovalbumin ds-cDNA has been reconstituted. The R.HpaII site to the right hand side of the ovalbumin DNA has been rendered non-functional, thus linking fragments A2 and B and the integrated ovalbumin ds-cDNA to produce a new fragment of 1400 base pairs. 3) The topography of the R.HpaII fragment of pCR1ov2.1 DNA containing the integrated ovalbumin DNA sequence. Fragments A1 and A2 + B flank the integrated sequence, as in pCR1ov2.16. Neither of the EcoRI sites on either side of the integrated DNA have been reconstituted. The dotted lines correspond to ovalbumin gene sequences plus oligodG-dC flanking tails. Non functional R.HpaII and EcoRI sites are between brackets.
(1) wild-type pCR1 DNA, (2) pCR1ov2.16 DNA, (3) pCR1ov2.1 DNA.

tecting "full length" ovalbumin cDNA against digestion by nuclease S1 (unpublished).

Since the exact length of the oligodG-dC sequences on either side of the integrated ds-cDNA are not known, the exact length of the ovalbumin-specific sequences in pCR1ov2.1 can not be determined simply from the above gel electrophoreses (see below).

3. Fidelity and size of the inserted ovalbumin gene fragment.

High fidelity of the procedures used for synthesizing and amplifying the ovalbumin gene is important for many possible applications, i.e.: construction of a physical map of the ovalbumin structural gene and its flanking sequences in cellular DNA using restriction endonucleases, use of the amplified ovalbumin gene sequences as specific hybridization probes, production of ovalbumin protein in bacteria. Since neither the ovalbumin protein nor the ovalbumin mRNA or cDNA have been as yet sequenced, we have tested the fidelity of the procedures used in the synthesis and cloning of the ovalbumin gene by using nuclease S1 which under appropriate conditions cleaves DNA duplex at mismatches which might possibly involve a single base-pair (23).

As described in legend to Fig. 6, [^{3}H]-labelled "full length" ovalbumin cDNA was hybridized with ovalbumin mRNA or with the HhaI pCR1ov2.1 fragment A which contains the integrated ovalbumin sequence (unpublished result). After digestion with nuclease S1, 90% of the cDNA hybridized to ovalbumin mRNA and 75% of the cDNA hybridized to HhaI fragment A was acid-insoluble. The size of the nuclease S1 resistant cDNA was analyzed by polyacrylamide gel electrophoresis under denaturating conditions. The efficiency of the nuclease S1 treatment was demonstrated by the complete transformation, with the same amount of nuclease S1, of 2 µg of superhelical SV40 DNA into open circular and linear forms (24, 25). Fig. 6 A and B indicate that there was no detectable variation of the length of the cDNA when the cDNA-ovalbumin mRNA hybrid was treated with nuclease S1. This result suggests that the ovalbumin cDNA is a rather faithful transcript of the ovalbumin RNA, although it should be pointed out that it is presently unknown whether nuclease S1 cleaves as efficiently a mismatched DNA:RNA hybrid as a mismatched DNA duplex. In any case, our result is in good agreement with previous evidences that cDNAs are accurate transcripts of mRNAs (26, 27, 28). From the data shown in Fig. 6 A and B a length of about 1930 nucleotides was determined for "full length" ovalbumin cDNA. Examination of Fig. 6 C indicates a small but definite length reduction of the cDNA after its hybridization with the HhaI pCR1ov2.1 fragment A and nuclease S1 digestion. A length of about 1730 was calculated

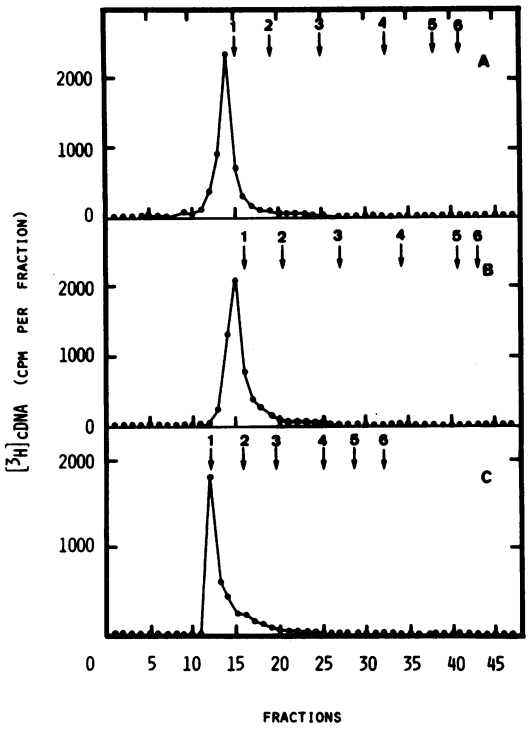

FIGURE 6 Fidelity and size of ovalbumin DNA sequences inserted in pCR1ov2.1 clone. The HhaI restriction enzyme DNA fragment A (about 2400 base pairs) which contains the ovalbumin gene sequences inserted in pCR1ov2.1 was obtained as described elsewhere (manuscript in preparation) and "full length" single-stranded HhaI fragment A molecules were further purified by sedimentation through a 5-20% sucrose alkaline gradient containing 0.2 M NaOH, 0.9 M NaCl and 25 mM EDTA (SW60 Ti Spinco rotor, 34,000 rpm, 20°C, 14 h). 4 ng of "full length" ovalbumin [^{3}H] cDNA (40 ng/ml, 15,000 cpm/ng) (Materials and Methods) was hybridized (24 h at 37°C under paraffin oil, in siliconized tubes) with denatured (20 min at room temperature in 0.1 M NaOH) HhaI DNA fragment A (3 µg/ml) (panel A and C) or with immunoprecipitated ovalbumin mRNA (200 µg/ml) (panel B) in a hybridization mixture containing 32% deionized formamide, 40 mM NaCl, 2 mM EDTA and 50 mM PIPES buffer pH 7.6 in a final volume of 100 µl. At the end of the hybridization period the samples were diluted 10 times with a buffer containing 30 mM sodium acetate pH 4.9, 3 mM $ZnSO_4$, 100 mM NaCl, 1 µg/ml denatured sonicated calf thymus DNA, and incubated for 1 h at 37°C in the absence (panel A) or in the presence (panels B and C) of 32 units of purified nuclease S1 (Materials and Methods). After ethanol precipitation and centrifugation the pellets were dissolved in 40 µl of 10 mM Tris-HCl pH 7.5, 1 mM EDTA. Aliquots of 10 µl were brought to 70% deionized formamide and 7 mM sodium phosphate pH 6. After addition of 0.2 µg of a Hind II + III partial digest of [^{14}C]SV40 DNA (27,000 cpm/ µg) the mixture was heated at 70°C for 10 min, and electrophoresed in 3% polyacrylamide gels containing 99% formamide (30) for 14 h at 40 V (panel C) or 17 h at 60 V (panels A and B). The gels were then cut into 2 mm slices, dissolved in 0.75 ml H_2O_2 and counted for [^{3}H] and [^{14}C] as described in ref. 30. The arrows indicate the position of some of the SV40 DNA fragments whose lengths have been determined by comparing their migration with that of mouse satellite EcoRII fragments (21); 1, 2, 3, 4, 5 and 6 correspond to the position of fragments of 1770, 1225, 830, 510, 360 and 260 nucleotides, respectively.

from the relative migration of the SV40 restriction enzyme markers. The slight shoulder on the right side of the cDNA peak is most likely due to the presence of some nicked molecules in the HhaI restriction DNA fragment, since it was much more pronounced when the HhaI fragment was not purified on an alkaline sucrose gradient before hybridization with cDNA (see legend to Fig. 6). Since we did not find any discrete smaller fragment of cDNA on polyacrylamide gels of higher percentage (not shown), we conclude that pCR1ov2.1 contains a sequence of about 1730 nucleotides which shows no gross differences with the corresponding sequence of the natural ovalbumin gene. It should indeed be pointed out that single base deletions similar to that recently observed by Browne et al. (31) might not have been cleaved during nuclease S1 treatment, since it has not been demonstrated that all single base pair mismatches are actually cleaved by nuclease S1 treatment (23).

DISCUSSION

In the work described in this paper, ovalbumin messenger RNA has been converted into a double-stranded DNA copy which was then integrated into a bacterial plasmid by an oligodG-dC tailing method. In order to accurately re-excise the ds-cDNA, an attempt was made to reconstitute the EcoRI restriction targets on either side of the integrated cDNA sequence (16). This strategy however met with little success, which could be due either to the close proximity to the EcoRI restriction sites of stable blocks of oligodG-dC thus rendereing them non-functional, or to the fact that sequence re-arrangements take place in or around the reconstituted EcoRI targets during *in vivo* repair. Maniatis et al. (12) have observed cross-overs between tracts of oligodA and oligodT during the analysis of hybrid plasmids, which renders the second of the above possibilities more probable. However, a limited number of hybrid plasmids do retain EcoRI cleavage sites, and it is possible that the efficiency of cleavage site reconstitution would be increased if an *in vitro* repair step, using a combination of reverse transcriptase and DNA ligase, was included in the experimental design.

In spite of the fact that two sucrose gradient size-selec-

tion steps were included in the preparation of ds-cDNA, it is interesting to note that some of the hybrid plasmids, for example pCR1ov 3.4, 3.21 and 3.25 have incorporated only short lengths of ovalbumin specific DNA. The most reasonable explanation for this observation is the presence of endonucleolytic activity in terminal transferase preparations. Under the conditions used for homopolymer extension of ds-cDNA and linear plasmid, all of our enzyme preparations, purified to near homogeneity according to the procedure of Bollum (5), contain low levels of endonucleolytic activity. Sucrose gradient sedimentation and DNA-cellulose chromatography failed to completely remove this contaminating activity. Internal nicking could result in internal homopolymer addition to the ds-cDNA, and it is possible that some recombinant clones were formed by non-terminal hybridization events, the excess ds-cDNA being subsequently deleted during transfection of E.coli.

Many of the hybrid clones however (i.e. pCR1ov2.1, 2.5, 2.10 and 3.7) contain extensive sequences of ovalbumin-specific DNA. Since the ovalbumin protein requires only about 1100 coding nucleotides, it is very likely that clone pCR1ov2.1, in which about 1730 base pairs of ovalbumin DNA have been integrated, contains in fact the entire coding sequence of the ovalbumin structural gene (for references, see 29). This clone should be very useful to elucidate the ovalbumin gene organization and to study its expression *in vivo* and *in vitro*.

The present cloning experiments were carried out in a negatively pressurized laboratory especially equipped for this type of work and conforming to the NIH P3 containment level. Biohazards associated with the experiments described in this publiction have been examined previously by the French National Control Committee.

ACKNOWLEDGEMENTS

We thank the Viral Cancer Program, National Cancer Institute, for gifts of AMV DNA polymerase, Dr. Jolles (Rhone-Poulenc, France) for a gift of actinomycin D., Dr. M. Bellard for a gift of purified nuclease S1. We also thank Dr. F. Rougeon for many helpful discussions and suggestions, Dr. P. Oudet for his help with the electron microscopy and Miss C. Lambs for excellent technical assistance. This work was supported by grants from the INSERM (CRT N° 76.5.468), the CNRS (ATP N° 2117) and the Fonda-

tion pour la Recherche Médicale Française.

*Peter Humphries was supported by a long-term EMBO fellowship.

Present address : The Beatson Institute for Cancer Research
Laboratory of Molecular Pathology
Glasgow, Scotland, G.B.

**To whom reprint requests should be sent

[a]ABBREVIATIONS

AMV, avian myeloblastosis virus; cDNA, single-stranded DNA complementary to ovalbumin mRNA; ds-cDNA, double-stranded cDNA; SV40, Simian Virus 40.

REFERENCES

1. Palmiter, R.D. (1975) Cell 4, 189-197.
2. Murray, K., Murray, N.E. and Brammar, W.J. (1975) FEBS Proceedings 10th Paris Meeting 38, 193-207.
3. Cohen, S.N. (1975) Scientific American 233 (July) 25-33.
4. Kacian, D.L. and Spiegelman, S. (1973) Methods Enzymol. 29E, 150-173.
5. Bollum, F.J., Chang, L.M.S., Tsiapalis, C.M. and Dorson, J.W. (1974) Methods Enzymol. 29E, 70-78.
6. Klenow, H., Overgaard-Hansen, K. and Patkar, S.A. (1971) Eur. J. Biochem. 22, 371-381.
7. Green, P.J., Bettack, M.C., Goodman, H.M. and Berger, H.W. (1974) Methods Mol. Biol. 7, 87-105.
8. Vogt, V.M. (1973) Eur. J. Biochem. 33, 192-200.
9. Shapiro, D.J., Taylor, J.N., McKnight, S.G., Palacios, R., Gonzalez, C., Kiely, M.L. and Schimke, R.T. (1974) J. Biol. Chem. 249, 3665-3671.
10. Covey, C., Richardson, D. and Carbon, J. (1976) Mol. Gen. Genet. 145, 155-158.
11. Rougeon, F., Kourilsky, P. and Mach, B. (1975) Nucl. Acids Res. 2, 2365-2378.
12. Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.G. (1976) Cell 8, 163-182.
13. Higuchi, R., Paddock, G.V., Wall, R. and Salser, W. (1976) Proc. Natl. Acad. Sci. USA 73, 3146-3150.
14. Rabbits, T.H. (1976) Nature 260, 221-225.
15. Wood, K.O. and Lee, J. (1976) Nucl. Acids Res. 3, 1961-1971.
16. Rougeon, F. and Mach, B. (1976) Proc. Natl. Acad. Sci. USA 73, 3418-3422.
17. Efstratiadis, A., Kafatos, F.C., Monam, A.M. and Maniatis, T. (1976) Cell 7, 279-288.
18. Lobban, P.E. and Kaiser, A.D. (1973) J. Mol. Biol. 78, 453-471.
19. Jackson, D.A., Symons, R.H. and Berg, P. (1972) Proc. Natl. Acad. Sci. USA 69, 2904-2909.
20. Yang, R.C.A., Van der Voorde, A. and Fiers, W. (1976) Eur. J. Biochem. 61, 101-117.

21. Bellard, M., Oudet, P., Germond, J.E. and Chambon, P. (1976) Eur. J. Biochem. 70, 543-553.
22. Petterson, U., Mulder, C., Delius, H. and Sharp, P. (1973) Proc. Natl. Acad. Sci. USA 70, 200-204.
23. Schenk, T.E., Rhodes, C., Rigby, P.W.J. and Berg, P. (1975) Proc. Natl. Acad. Sci. USA 72, 989-993.
24. Hossenlopp, P., Oudet, P. and Chambon, P. (1974) Eur. J. Biochem. 41, 397-411.
25. Beard, P., Morrow, J.F. and Berg, P. (1973) J. Virol. 12, 1303-1313.
26. Poon, R., Paddock, G.V., Heindell, H., Whitcome, P., Salser, W., Kacian, D., Bank, A., Gambino, R. and Ramirez, F. (1974) Proc. Natl. Acad. Sci. USA 71, 3502-3506.
27. Marotta, C.A., Forget, B.G., Weissman, S.M., Verma, I.M., McCaffrey, R.P. and Baltimore, D. (1974) Proc. Natl. Acad. Sci. USA 71, 2300-2304.
28. Salser, W., Bower, C., Browne, D., Adli, F.E., Fedoroff, N., Frey, K., Heindell, H., Paddock, G., Poor, R., Wallace, B. and Whitcome, P. (1976) Fed. Proc. 35, 23-35.
29. Cheng, C.C., Brownlee, G.G., Carey, N.H., Doel, M.T., Gillam, S. and Smith, M. (1976) J. Mol. Biol. 107, 527-547.
30. Brison, O. and Chambon, P. (1976) Anal. Biochem. 75, 402-409.
31. Browne, J.K., Paddock, G.V., Liu, A., Clarke, P., Heindell, H.C. and Salser, W. (1977) Science 195, 389-391.