Cloning of soybean leghemoglobin structural gene sequences synthesized in vitro

Ebbe Truelsen, Kirsten Gausing, Bjarne Jochimsen, Poul Jørgensen and Kjeld A. Marcker

Department of Molecular Biology and Plant Physiology, University of Aarhus, DK-8000 Aarhus C, Denmark

Received 20 April 1979

ABSTRACT

Double-stranded soybean leghemoglobin DNA was synthesized from leghemoglobin mRNA isolated from soybean nodules. The dsDNA was inserted into the Bam HI site of plasmid pBR322 using the poly-dAT-joiner method. A cloned DNA fragment of one recombinant plasmid was isolated and characterized by restriction endonuclease digestion. The restriction cleavage map and the DNA sequence of a selected part of the inserted DNA are in complete accordance with the amino-acid sequence of soybean leghemoglobin.

INTRODUCTION

Leghemoglobin is synthesized exclusively in the nitrogen fixing root nodules which develop due to the symbiotic association of a bacterium (*Rhizobium*) with legumes. In soybean nodules leghemoglobin may account for about 20% of the total protein synthesis¹. Recent evidence strongly suggest that leghemoglobin is a plant-gene product²⁻³. Since the leghemoglobin genes are only activated in response to a bacterial function it follows that in this particular case an eukaryotic gene somehow must be under the control of a prokaryotic function. This raises the possibility that genetic analysis of a prokaryotic organism might be applied in a study of the activation of a gene from a higher eukaryote. Leghemoglobin is encoded in a polyA(+) 9S mRNA species which is translated on cytoplasmic ribosomes³. In soybean nodules there are three major forms of leghemoglobin called leghemoglobin a, c₁, and c₂, respectively. The difference between leghemoglobin a and c is six amino acids while leghemoglobin c₁ and c₂ differ in one amino acid only⁴. In order to characterize the organisation of the leghemoglobin genes in the soybean genome we describe here the cloning, identification,

characterization, and partial sequencing of a synthetic DNA copy of soybean leghemoglobin mRNA

MATERIALS AND METHODS

Enzymes. Avian myoblastosis virus reverse transcriptase was provided by Dr. J.W.Beard through the Offices of Programme, Resources and Logistics, National Cancer Institute of the U.S.A. Endonuclease S1 from Aspergillus cryzae was from Sigma. DNA polymerase was from Boehringer. Terminal transferase was from Bethesda Research Laboratories, Rockville. Restriction endonuclease PstI was a gift from Dr. H.Mayer, Braunschweig. All other restriction endonucleases were from New England Biolabs. T4 polynucleotide kinase was from New England Biolabs. Proteinase K was from Merck. Oligo -dT-cellulose Type T2 was from Collaborative Research Inc. ^{125}I -NaI, ^3H -dTTP (49 Ci/mmol) and α - ^{32}P -dATP (350 Ci/mmol) and γ - ^{32}P -ATP (3000 Ci/mmol) were from The Radiochemical Centre, Amersham. ^3H -polyU (35 mCi/mmol) was from Miles Research Laboratories.

Isolation of mRNA. Soybean seeds (var. Brávalla) were germinated in moist vermiculite. 4 days old seedlings were infected with Rhizobium japonicum (strain 61A69). Plants were grown for approximately four weeks in a green house after which the nodules were collected and immediately immersed in liquid nitrogen. The nodules were ground to a fine powder under liquid nitrogen and then extracted for 10 min with 4 volumes of buffer containing 150 mM Tris/acetate pH 8.5, 50 mM KCl, 20 mM Mg acetate, 200 mM sucrose, 1 mM aurine tricarboxylic acid⁵, 5 mM β -mercaptoethanol, 100 $\mu\text{g}/\text{ml}$ cycloheximide, and 2 $\mu\text{g}/\text{ml}$ of polyvinyl sulphate. After centrifugation for 10 min in a Sorvall centrifuge (15,000xg) the supernatant was collected and layered on top of a 1.5 M sucrose solution containing 10 mM Mg acetate, 40 mM KCl, 50 mM Tris/acetate pH 8.5, 100 $\mu\text{g}/\text{ml}$ cycloheximide, and 2 $\mu\text{g}/\text{ml}$ polyvinyl sulphate. The polysomes were collected by centrifugation for 75 min at 50,000 rpm in a Beckman Ti60 rotor. Polysomes were digested with proteinase K (500 $\mu\text{g}/\text{ml}$) for 1 h at 30°C in a buffer containing 10 mM Tris/HCl pH 7.5, 100 mM NaCl, 5 mM EDTA and 1% SDS. PolyA(+) mRNA was then isolated by chromatography at 30°C on oligo-dT-cellulose. Finally total polyA(+) mRNA was frac-

tionated by centrifugation through a 5-20% sucrose gradient containing 10 mM Tris pH 7.5, 5 mM EDTA pH 7.5, and 0.5% SDS. Centrifugation was at 30,500 rpm (SW41 Beckman rotor) for 17 h at 23°C. After centrifugation the sucrose gradient was monitored by measuring absorption at 260 nm and by hybridization with ³H-labelled polyU. A distinct peak of UV absorbing material which coincided with a peak hybridizing with ³H-polyU was seen in the 9-10S area of gradient. As already shown by others this mRNA corresponds to leghemoglobin mRNA¹. Peak fractions from this area were pooled and precipitated with alcohol. At this stage leghemoglobin mRNA is estimated to be 30-40% pure¹.

Protein synthesis in vitro. mRNA activity was assayed in the wheat-germ system according to Roberts and Paterson⁶ except that reaction mixtures in addition contained 50 μM spermine and 250 μM spermidine.

Synthesis of cDNA. The cDNA for 9S mRNA was prepared using avian myoblastosis virus reverse transcriptase. 20 μg of leghemoglobin mRNA was incubated in a 400 μl reaction mixture containing 50 mM Tris/HCl pH 8.5, 8 mM MgCl₂, 50 mM KCl, 4 mM Na₂P₂O₇, 0.2 mM DTT, 60 μCi ³H-dTTP, 0.2 mM dATP, dGTP, dTTP, dCTP, 10 μg actinomycin D, 7.5 μg oligo-dT and 180 units of reverse transcriptase. The reaction was for 90 min at 42°C. NaOH and EDTA were added to a final concentration of 0.3 M and 10 mM, respectively, and the mixture further incubated for 45 min at 70°C in order to hydrolyse RNA. The reaction was neutralized with HCl, and SDS added to a final concentration of 0.5%. The mixture was then subjected to gel filtration on a SP Sephadex C-50 column⁷ (10 × 1 cm) which previously had been equilibrated with 300 mM NaCl and 10 mM Na acetate pH 5.5. The cDNA, which elutes in the void volume, was precipitated with alcohol and collected by centrifugation. The yield of cDNA was about 50% of the original mRNA template.

Double-stranded DNA synthesis. 6 μg of cDNA was incubated in 600 μl containing 35 mM HEPES pH 6.9, 8 mM Tris pH 8.3, 20 mM KCl, 1.5 mM MgCl₂, 2 mM β-mercaptoethanol, 2 μCi α³²P dATP, 0.5 mM dATP, 0.5 mM dGTP, 0.5 mM dCTP, 0.5 mM dTTP, and 130 units of DNA polymerase²⁴. The mixture was incubated for 2 h at 15°C. After incubation SDS was added to a final concentration of 0.5% and

the mixture was chromatographed on a SP Sephadex C-50 column as above. The void volume containing the dsDNA was pooled and NaAc pH 4.5 and ZnCl_2 were added to a final concentration of 50 mM and 3 mM, respectively. The mixture was then incubated with 200 units of S1 nuclease for 1 h at 37°C, followed by phenol extraction and precipitation with ethanol. The yield of dsDNA was approximately 40% of the original cDNA template.

Addition of homopolymer tails. 200 ng of dsDNA was incubated at 37°C in 40 μl of a solution containing 140 mM K-cacodylate, 30 mM Tris pH 6.9, 1 mM CoCl_2 , 100 μM DTT, 1 μCi $\alpha^{32}\text{P}$ dATP, 45 μM dATP, and 50 units of terminal transferase⁸. The reaction was stopped when about 100 residues had been added. The tailed DNA was then isolated by chromatography on SP Sephadex C-50 as described above.

Construction of recombinant DNA plasmids. Recombinant DNA consisting of plasmid pBR322 and dsDNA derived from leghemoglobin mRNA was constructed by incorporation of the tailed eukaryotic DNA into the Bam HI site of the plasmid. 2 μg of pBR322 DNA was linearized by digestion with Bam HI enzyme and thymidylic acid residues were added by terminal transferase as described above. The reaction was stopped when on average 100 residues had been added. The tailed pBR322 DNA and eukaryotic DNA were mixed in a total volume of 300 μl consisting of 10 mM Tris pH 7.6, 100 mM NaCl, 1 mM EDTA. The mixture was incubated for 10 min at 65°C, followed by 2 h at 45°C after which an additional 600 μl of incubation buffer was added. The resulting mixture was used directly for transformation.

Transformation. For transformation the *E.coli* K12 strain HB101 was used⁹. It lacks restriction and modification functions and is *recA*⁻. Competent cells of HB101 were prepared according to the procedure of Mandel and Higa¹⁰ and transformation was carried out as described by Harpold *et al.*¹¹. After transformation the cells were spread directly on L-broth plates containing 30 $\mu\text{g}/\text{ml}$ ampicillin and incubated at 37°C. Ampicillin resistant colonies were screened on L-broth plates containing 10 $\mu\text{g}/\text{ml}$ tetracycline.

Colony screening for recombinant plasmids. Ampicillin resistant, tetracycline sensitive colonies were further screened by colony hybridization using purified ^{125}I -labelled leghemoglo-

bin mRNA. Colonies were transferred from L-broth plates onto nitrocellulose filters by contact and grown overnight at 37°C by layering the filters on top of an L-broth plate. After growth the filters were prepared for hybridization essentially as described by van der Berg *et al.*¹². Hybridization was performed in 50% formamide, 5 × SSC, using 2 × 10⁵ cpm of radioiodinated leghemoglobin mRNA per filter (16 cm²)¹³.

Isolation of plasmid DNA. Small cultures (5 ml) of positive colonies were grown in L-broth medium containing 30 µg/ml ampicillin. At a density of 5 × 10⁸ cells/ml 200 µg/ml of chloramphenicol was added and the next day the bacteria were harvested and lysed according to the procedure of Post *et al.*¹⁴. The lysate was extracted with phenol and nucleic acid precipitated with alcohol.

Larger amounts of plasmid DNA were prepared from one liter cultures of transformed HB101 grown as described above. After lysis nucleic acid was isolated using the polyethylene glycol procedure described by Humphreys *et al.*¹⁵. Plasmid DNA was finally isolated by centrifugation in CsCl/ethidium bromide gradients.

Iodination of leghemoglobin mRNA. This was performed as described by Efstratiadis *et al.*¹⁶.

Gel electrophoresis. The size of the recombinant plasmid DNA was estimated by gel electrophoresis in 1.25% agarose gels using 36 mM Tris, 30 mM NaH₂PO₄ pH 7.8, and 1 mM EDTA as buffer. pBR322 and pBR322 carrying a 700 base pair insertion (a gift from Dr. S. Pedersen) were used as markers. DNA fragments generated by restriction endonuclease cleavage of recombinant DNA were sized on 0.5% agarose, 2% polyacrylamide slab gels using 80 mM Tris, 24 mM Na-acetate, 36 mM NaCl, 4.4 mM EDTA adjusted to pH 8.0 as buffer.

Hybridization. Liquid hybridization of recombinant plasmid DNA with purified ³²P-leghemoglobin cDNA and subsequent S1 digestion were essentially as described by Seidman *et al.*¹⁷.

Sequencing procedures. Endonuclease digested DNA was labeled terminally with ³²P by the exchange reaction catalysed by polynucleotide kinase¹⁸. After digestion with an alternative endonuclease the DNA fragment was purified by agarose gel electro-

phoresis. The DNA was recovered from the gel by the method of Tabak and Flavell¹⁹. The actual sequencing was performed as described by Maat and Smith²⁰ and references given in this paper.

Safety considerations. Cloning experiments and handling of recombinant DNA was carried out in P2/EK1 conditions as required by the NIH guidelines. This classification has been accepted by the Danish Committee for Recombinant DNA Research.

RESULTS AND DISCUSSION

Leghemoglobin mRNA from soybeans has so far been isolated in a purity of 30-40%³. Thus a supply of pure leghemoglobin mRNA in a reasonable quantity is still a problem. We have, therefore, prepared a plasmid in which a double-stranded DNA copy of leg-hemoglobin mRNA is inserted into plasmid pBR322. The 9S mRNA obtained by size-fractionation on sucrose gradients of polysomal polyA(+) RNA was used as template for reverse transcriptase. The cDNA was synthesized in the presence of actinomycin D and was isolated after alkaline digestion of template RNA by gel filtration on SP Sephadex C-50 columns. Second strand synthesis was catalysed by *E.coli* DNA polymerase. The double-stranded DNA formed in this reaction was treated with S1 nuclease directly after SP Sephadex C-50 gel filtration. The eukaryotic DNA was inserted into the Bam H1 site of pBR322 using the poly-dAT-joiner method. The integration of foreign DNA sequences into the Bam H1 site of pBR322 leads to inactivation of the tetracycline resistance gene while the ampicillin resistance gene is not affected. Transformation of strain HB101 with 2 µg of annealed DNA yielded about 500 ampicillin resistant colonies, 450 of which were tetracycline sensitive. These transformants were screened by colony hybridization using purified ¹²⁵I-labelled leghemoglobin mRNA as a probe. Over 125 colonies gave positive signals. The high proportion of colonies carrying recombinant leghemoglobin DNA sequences reflects the content of leghemoglobin mRNA in the original preparation used as a template for the *in vitro* synthesis of the structural gene sequence. Plasmid DNA isolated from 100 such colonies were analysed by gel electrophoresis in 1.25% agarose gels using pBR322 and pBR322 carrying a 700 base-pair insertion as markers.

Sixteen clones which carried insertions of 700 base pairs or more were selected for further study.

The DNA from these clones were then screened for their ability to protect purified ^{32}P -labelled leghemoglobin cDNA from S1 nuclease digestion after hybridization. DNA from four clones which gave the best protection (60% or more) of cDNA after S1 digestion were chosen for further studies.

Plasmid DNA from these clones were digested with a variety of restriction endonucleases. During the sizing of the fragments so obtained, it became apparent that the plasmid of one clone, 2431, carried the longest DNA insertion. This plasmid, in the following called p2431, was subjected to restriction endonuclease cleavage analysis.

The complete sequence and restriction map of pBR322 has recently been published²². This allows rather accurate sizing of the restriction endonuclease fragments obtained from digestions of p2431. Fig. 1 shows the result of one such analysis and Table 1 is a summary of the results so obtained. On the basis of these data a cleavage map for the inserted leghemoglobin DNA could be constructed as shown in Fig. 2.

In the inserted DNA there are no cleavage sites for EcoRI, SalI, Bam HI, AvaI, PvuI, KpnI, HaeII, and BglII, SmaI, XhoI, XbaI, and SacI. The inserted DNA contains single cleavage sites for HindIII, PstI, HindII, and TaqI, and four cleavage sites for AluI. Using the Southern hybridization technique²¹ the fragments produced by cleavage with HindIII and PstI were shown to hybridize with purified ^{32}P -leghemoglobin cDNA, indicating that the inserted DNA in p2431 consists of leghemoglobin DNA sequences.

The amino-acid sequence of the major forms of leghemoglobin has recently been elucidated⁴. The relative positions of the restriction sites are in accordance with the amino acid sequence. The HindIII cleavage site corresponds to the sequence -Lys-Leu- (amino acids 64 and 65), the PstI cleavage site to the sequence -Pro-Ala-Ala- (amino acids 38, 39, 40), the HindII cleavage site to the sequence -Gly-Gln-Leu- (amino acids 74, 75, 76) and the Taq site to the sequence -Ser-Ile- (amino acids 32, 33).

In addition the orientation of the inserted DNA could be

Figure 1. DNA fragments produced by digestion of p2431 with various endonucleases. The products were stained with ethidium bromide after gel electrophoresis.

1. HindIII/SalI double digest, 2. PstI/SalI double digest.
3. TaqI digest. 4. Marker fragments derived from TaqI digestion of pBR322. The number of base pairs in each marker fragment is indicated.

deduced with the result that the 3'-end of leghemoglobin for RNA is closer to the EcoRI site of pBR322 than the 5'-terminal

TABLE 1

Enzyme	Number of base pairs				
PstI	3470	+	1590		
HindIII	4270	+	790		
TaqI	535	+	435		
HindII	1230	+	550		
AluI	520	+	520	+	190
PstI/SalI	1590	+	455		
PstI/HindIII	3470	+	790	+	75
PstI/HindII	1230	+	445	+	105
PstI/AluI	520	+	500	+	190
HindIII/SalI	790	+	530		
AluI/HindII	520	+	520	+	180
TaqI/HindIII	435	+	435	+	95

Table 1. The size of restriction endonuclease fragments containing inserted sequences obtained from p2431. The accuracy of the size determination is better than 5%.

part.

DNA sequencing data for the cloned fragment confirm that it contains leghemoglobin structural sequences. For sequence analysis the recombinant plasmid DNA was digested with HindIII which produced two fragments. The 5'-end of the two DNA fragments were labelled with ^{32}P using T4-polynucleotide kinase and $\gamma\text{-}^{32}\text{P}\text{-ATP}$. After labelling the fragments were redigested with SalI. Consistent with the map shown in Fig. 1, three labelled fragments were obtained by gel electrophoresis. The 530 base-pair fragment containing the 5'-end of the structural sequence was subjected to DNA sequence analysis as described in Materials and Methods.

The sequence determined from part of the HindIII SalI fragment is completely consistent with that predicted for the coding region from amino acid 52 to 58 and further establish the polarity of the inserted structural leghemoglobin sequence.

The drawback of the poly-dAT-joiner method is that the in-

Figure 2. The structure of p2431

Open boxes, pBR322 DNA. Closed boxes, leghemoglobin DNA. AT, poly-dAT tails. The cleavage sites of the various restriction endonuclease enzymes are indicated above the boxes. Distances are given in number of base pairs.

a. Linear cleavage map of p2431

The orientation of the inserted leghemoglobin DNA is indicated below the box. The total insert of about 700 base pairs include both leghemoglobin DNA and poly-dAT tails.

b. Detailed structure of the inserted leghemoglobin DNA. For Alu the cleavage sites indicated represent a minimum number.

c. The DNA sequence of part of the inserted sequence compared to the amino-acid sequence from residue 52 to 58 in leghemoglobin from soybeans.

serted sequence cannot be cleanly exised from the plasmid DNA by restriction endonucleases. Repeated attempts to exise this sequence for size determination by partial denaturation of the polydAT and digestion with S1²³ have so far failed. However,

preliminary DNA-sequence analysis suggest that the inserted structural sequence is at least 500 base pairs long implying that the entire structural gene may have been cloned.

Soybean leghemoglobin consists of three major forms, leghemoglobin a, leghemoglobin c₁, and leghemoglobin c₂. There are six amino-acids differences between leghemoglobin a and leghemoglobin c₁⁴. The two c forms differ in one amino acid only. Most of the differences observed can be accounted for by point mutations. At present the restriction cleavage pattern and the DNA sequence data are insufficient to allocate the inserted structural sequence in clone 2431 to any of the three possibilities. However, one other clone isolated has an additional AluI cleavage site towards the 3'-end. Thus it is probable that both leghemoglobin a and leghemoglobin c sequences have been cloned. Further DNA-sequence analysis will allow the correct distinction between the two possibilities. The cloned leghemoglobin DNA is now used as a hybridization probe in order to study the organization of the leghemoglobin structural gene sequences in soybean genomic DNA.

ACKNOWLEDGEMENTS

This study was supported by The Danish Research Council, The Novo Foundation, and The Carlsberg Foundation. Ms. Y.Gregersen, Mr. M.Meibom-Larsen, and Ms. A.Marcker are acknowledged for skilful technical assistance.

REFERENCES

1. Sidloi-Lumbroso, R. and Schulman, H.M. (1977) *Biochim.Biophys.Acta* 476, 295-302.
2. Sidloi-Lumbroso, R., Kleiman, L. and Schulman, H.M. (1978) *Nature* 273, 558-560.
3. Baulcombe, D. and Verma, D.P.S. (1978) *Nucleic Acids Res.* 5, 4141-4153.
4. Sievers, G., Huhtala, M.-L., and Ellfolk, N. (1978) *Acta Chem.Scand.* B32 380-386.
5. Hallick, R.B., Chelm, B.K., Gray, P.W., and Orozco, E.M. (1977) *Nucleic Acids Res.* 4, 3055-3064.
6. Roberts, B.E. and Paterson, B.M. (1973) *Proc.Nat.Acad.Sci. U.S.A.* 70, 2330-2334.

7. Friedman, B.Y. and Rosbash, M. (1977) *Nucleic Acids Res.* 4, 3455-3471.
8. Roychoudhury, R., Jay, E., and Wu, R. (1976) *Nucleic Acids Res.* 3, 101-116.
9. Boyer, H.W. and Roulland-Dussoix, D. (1969) *J.Mol.Biol.* 41, 459-472.
10. Mandel, M. and Higa, A. (1970) *J.Mol.Biol.* 53, 159-162.
11. Harpold, M.M., Dobner, P.R., Evans, R.M., and Bancroft, F.C. (1978) *Nucleic Acids Res.* 5, 2039-2053.
12. van den Berg, J., van Ooyen, A., Mantel, N., Schamböck, A., Grosveld, G., Flavell, R.A., and Weissmann, C. (1978) *Nature* 276, 37-44.
13. Grunstein, M. and Hogness, D.S. (1975) *Proc.Nat.Acad.Sci, U.S.A.* 72, 3961-3965.
14. Post, L.E., Arfsten, A.E., Reusser, F., and Nomura, M. (1978) *Cell* 15, 215-229.
15. Humphreys, G.O., Willshaw, G.A., and Anderson, E.S. (1975) *Biochim.Biophys.Acta* 383, 457-463.
16. Efstratiadis, A., Maniatis, T., Kafatos, F.C., Jeffrey, A., and Vournakis, J.N. (1975) *Cell* 4, 367-378.
17. Seidman, J.G., Edgell, M.H., and Leder, P. (1978) *Nature* 271, 582-585.
18. Berkner, K.L. and Folk, W.R. (1977) *J.Biol.Chem.* 252, 3176-3184.
19. Tabak, H.F. and Flavell, R.A. (1978) *Nucleic Acids Res.* 5, 2321-2332.
20. Maat, J. and Smith, A.J.H. (1978) *Nucleic Acids Res.* 5, 4537-4545.
21. Southern, E.M. (1975) *J.Mol.Biol.* 98, 503-517.
22. Sutchcliffe, J.G. (1978) *Nucleic Acids Res.* 5, 2721-2728.
23. Hofstetter, H., Schambock, A., van den Berg, J., and Weissmann, C. (1976) *Biochim.Biophys.Acta* 454, 587-591.
24. Wickens, M. P., Buell, G. N. and Schimke, R. T. (1978) *J. Biol. Chem.* 253, 2483-2495.