Cloning and Characterization of the Gene for Rabbit C-Reactive Protein†

Shou-Ih Hu, Steven M. Miller, and David Samols*

Department of Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

Received May 1, 1986; Revised Manuscript Received June 26, 1986

ABSTRACT: C-reactive protein (CRP), an acute-phase plasma protein of hepatic origin in man and rabbit, is a cyclic pentamer composed of five identical nonglycosylated M_r 22 500 subunits. We have isolated both cDNA and genomic clones for rabbit CRP. These clones were used as probes to demonstrate that when CRP synthesis is increased following an acute inflammatory stimulus, there is a corresponding increase in the level of accumulated CRP mRNA. The rabbit CRP gene is 2.6 kilobases in length containing a single intron of 252 base pairs (bp) which interrupts the codon for amino acid 2 in the protein. The mRNA for CRP contains a 5'-nontranslated region of 113 bp and a 3'-nontranslated region of 1550 bp. Sequencing of the protein-coding region of the gene indicates that the primary translation product contains a 20 amino acid N-terminal signal peptide. The deduced amino acid sequence is in general agreement with the published sequence [Wang, C. M., Nguyen, N. Y., Yonaha, K., Robey, F., & Liu, T.-Y. (1982) *J. Biol. Chem. 257*, 13610-13615] except in the region between amino acids 63 and 73. In this region, the sequence of both cDNA and genomic clones indicates the presence of 28 amino acids not previously reported. This alteration may be the result of genetic heterogeneity or an error in the reported protein sequence.

Part of the systemic response to inflammation includes an increase in the plasma concentration of a number of plasma proteins of hepatic origin, which collectively are termed the acute-phase proteins (Kushner, 1982). C-reactive protein (CRP) was the first discovered and among the best characterized of this class (Tillett & Francis, 1930). In response to inflammatory stimuli, its serum concentration may increase by as much as 100-fold or more when compared to normal individuals (Kushner & Feldmann, 1978; Pepys, 1981). The CRP response in humans and rabbits is quite similar, and their CRPs share considerable amino acid homology (Oliveira et al., 1979; Wang et al., 1982). Both are reportedly composed of five identical, nonglycosylated M_r 22 500 subunits (Gotschlich & Edelman, 1965). These properties, among others, have made the rabbit an excellent animal model for studying the physiological and biochemical phenomena of the CRP response to inflammation.

Very little is known about the mechanism of CRP induction. The dramatic increase in serum concentration of CRP has been shown to be due to new synthesis of the protein by hepatocytes after an inducing signal from an unknown mediator (Kushner & Feldmann, 1978; Macintyre et al., 1983). Samols et al. (1985) further showed by in vitro translation assay that this increased CRP synthesis after an inflammatory stimulus was accompanied by an increase in the level of translatable rabbit hepatic CRP mRNA. In this report, as the first step toward further investigation of the mechanism regulating the transcription and secretion of CRP, we have isolated and characterized the gene for rabbit CRP.

Materials and Methods

The acute-phase response was induced in New Zealand white rabbits by intramuscular injection of 0.5-1.0 mL of turpentine in each thigh 24 h before sacrifice. Control rabbits received no inflammatory stimulus. Serum CRP concentration at the time of sacrifice was measured as described previously (Kushner & Somerville, 1971) and was 60-70 μg/mL for inflamed rabbits and 3 μg/mL for the control animals.

RNA Isolation. RNA was isolated from liver of inflamed or control rabbits after homogenization in guanidinium thiocyanate as described by Chirgwin et al. (1979). Poly(A)-containing RNA was isolated by chromatography on oligo(dT)-cellulose (Aviv & Leder, 1972). RNA was stored in water at −70 °C until use.

Northern Blot Analysis. For Northern blot analysis, 5 μg of poly(A)-containing RNA or 25 μg of total RNA was size fractionated on 0.8% formaldehyde-containing agarose gels, transferred to a nitrocellulose filter, prehybridized in 50% formamide, 2× Pipes [1× Pipes = 40 mM piperazine-*N*,*N*'-bis(2-ethanesulfonic acid) (Pipes) buffer, pH 6.4, 1 M ethylenediaminetetraacetic acid (EDTA), and 0.4 M NaCl], 0.5% sodium dodecyl sulfate (SDS), and 100 μg/mL salmon sperm DNA, and hybridized in the same buffer with 10^6 cpm/mL nick-translated DNA for 24 h at 42 °C. The filters were washed in 2× SSC (1× SSC = 0.15 M NaCl and 0.015 M sodium citrate) and 0.1% SDS followed by 0.2× SSC and 0.1% SDS at 65 °C. To strip off the original probe for rehybridization, the filter was washed in 1× Denhardt's (0.2% Ficoll, 0.2% polyvinylpyrrolidone, and 0.2% bovine serum albumin), 0.2% sodium pyrophosphate, 0.3 mM EDTA, and 5 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris·HCl), pH 8, at 80 °C for 2 h. The filter was exposed to X-ray film for 2 days to confirm that the original probe was removed.

Library Construction. A cDNA library was prepared in pBR322 from poly(A)-containing RNA isolated from the liver of an inflamed rabbit primed with calf thymus random primers and cloned by the G-C tailing method, using the procedures described by Maniatis et al. (1982). A rabbit genomic library constructed in the λ cloning vector Charon 4A was a gift from Dr. R. Hardison (Department of Biochemistry and Biophysics, Pennsylvania State University, University Park, PA).

Screening of cDNA and Genomic Library. The cDNA library was plated on *Escherichia coli* strain HB101. The rabbit genomic library was plated on *E. coli* strain DP50. Screening of duplicate filters was carried out according to the standard procedures (Maniatis et al., 1982). For screening

† This work was supported by Grant DCB84-17616 from the National Science Foundation and Grant AG02467 from the National Institutes of Health.

0006-2960/86/0425-7834$01.50/0

the cDNA library, the hybridization was performed in 50% formamide and 6× SSC at 37 °C using a nick-translated human CRP cDNA clone as a probe. This human cDNA clone was the generous gift of Dr. H. Colten (Department of Pediatrics, Washington University, St. Louis, MO). For screening the genomic library, hybridization was performed in 50% formamide and 5× SSC at 42 °C in the presence of 10% dextran sulfate using nick-translated rabbit cDNA 3c as a probe. The filter was washed at 37 °C in 2× SSC followed by a final wash in 0.1× SSC for cDNA screening and a final wash in 0.3× SSC at 50 °C for genomic screening.

Restriction Mapping and Southern Blot Analysis of λ Clones. The positions of restriction enzyme sites within the genomic clones λCRP I and λCRP II were determined by complete digestions singly or doubly with the enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, and *Xho*I. The lengths of the digestion products were determined following electrophoresis on 0.9% agarose gels using *Hin*dIII-digested λ DNA as a molecular weight standard. For Southern blot analysis, DNAs were transferred from the gel to a nitrocellulose filter prehybridized and hybridized with the nick-translated cDNA probes in 50% formamide, 2× SSC, 1× Denhardt's, 50 mM potassium phosphate buffer, pH 6.5, and 100 μg/mL salmon sperm DNA at 42 °C for 16 h. The final wash was done in 0.2× SSC at 60 °C.

Sequence Analysis. cDNA was sequenced by using the chemical degradation procedure of Maxam and Gilbert (1977). For this purpose, cDNA inserts were subcloned into pUC9, 5′ end labeled at restriction enzyme sites in the polylinker (*Bam*HI or *Hin*dIII), and sequenced on both strands. Portions of the genomic clone were sequenced by the chain termination method of Sanger et al. (1977). Individual restriction fragments as indicated in Figure 2B were subcloned into M13mp18 or M13mp19 for this purpose. Single-stranded DNA was isolated from the subclone-containing phage. In most cases, a commercially available M13 sequencing primer, a 17-mer, was used in the reactions. To help sequence through the large *Hin*dIII fragment, a synthetic 18-mer, GAGGATCTCGTTAAATTG, was used as a primer.

S1 Nuclease Experiment. The procedure for S1 nuclease mapping modified that of Berk and Sharp (1977). Probes for these experiments were prepared as follows. For mapping the 5′ end of the transcript, plasmid p19RS5′, which is the 5′ *Eco*RI/*Sph*I fragment of the 2.7-kilobase (kb) *Eco*RI fragment (see Figure 2) cloned into plasmid pUC19, was digested with restriction enzyme *Nco*I, treated with alkaline phosphatase, and 5′ end labeled with ^{32}P using T4 polynucleotide kinase and [γ-^{32}P]ATP. The resulting DNA was digested with *Eco*RI. The fragment containing the 5′-nontranslated region was purified from an agarose gel and used as a probe for the S1 analysis. For mapping the 3′ end of the transcript, plasmid pCGR3, which is the 4.3-kb *Eco*RI fragment (see Figure 2) downstream from the bulk of the CRP gene cloned into plasmid pUC9, was digested with *Eco*RI. The fragment was 3′ end labeled with ^{32}P using T4 DNA polymerase and [α-^{32}P]CTP in the presence of unlabeled dATP, dGTP, and TTP.

Primer Extension Experiment. The primer extension assay was based on the procedure of Luse et al. (1981). The primer used in this experiment was produced by digestion of the 5′ end labeled *Nco*I/*Eco*RI fragment used in the S1 nuclease experiment with restriction enzyme *Hin*fI. The size of the primer was about 40 base pairs (bp) in length.

Miscellaneous. Preparation of phage DNA and plasmid DNA and subclonings were performed according to Maniatis et al. (1982). The procedure for transformation of plasmid DNA into *Escherichia coli* was that of Hanahan (1983). Restriction enzymes, T4 polynucleotide kinase, T4 DNA polymerase, T4 DNA ligase, DNA polymerase holoenzyme, DNA polymerase Klenow fragment, and calf intestinal alkaline phosphatase were purchased from Boehringer Mannheim. ^{32}P-Labeled compounds were purchased from New England Nuclear.

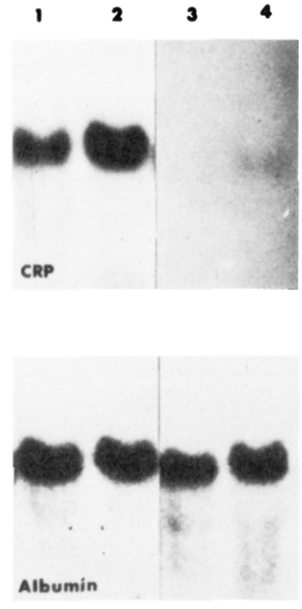

FIGURE 1: Northern blot analysis of RNA from inflamed or uninflamed rabbit liver. (Top) Five micrograms of poly(A)-containing RNA from inflamed (lanes 1 and 2) or uninflamed (lanes 3 and 4) rabbit livers was separated by electrophoresis on a 0.8% denaturing agarose gel, transferred to a nitrocellulose filter, and hybridized with ^{32}P-labeled rabbit CRP cDNA clone 3c. (Bottom) The same filter rehybridized with a ^{32}P-labeled rat albumin probe.

Results and Discussion

Isolation of Rabbit CRP cDNA and Northern Blot Analysis of CRP mRNA. Using a human CRP cDNA as a probe, we screened a rabbit cDNA library and identified two plasmids which were potentially derived from rabbit CRP mRNA. These two cDNAs, called 2a and 3c, were used as probes in Northern blot experiments with liver RNA from inflamed and control rabbits. Both probes gave the same results, an example of which is shown in the top panel of Figure 1. An RNA species of about 2.4 kb in size was identified in the RNA from the inflamed rabbit liver and was only weakly if at all present in the RNA from an uninflamed control rabbit liver. This is consistent with the expected abundance of CRP mRNA, and it confirms our previous finding (Samols et al., 1985) that the increased synthesis of CRP after an inflammatory stimulus is due to increased accumulation of translatable CRP mRNA and not due to modification of preexisting CRP mRNA. The closely related human CRP mRNA is 2.2 kb in size (Tucci et al., 1983).

As a control, the hybridized radioactivity on the filter was removed and the filter rehybridized with a rat albumin cDNA probe which cross-hybridizes with rabbit albumin mRNA. As shown in the bottom panel of Figure 1, an abundant 2.1 kb RNA species hybridized to the probe about equally well in all lanes, indicating that about equal amounts of RNA were applied in each lane. Therefore, the absence of an RNA species of 2.4 kb which hybridizes specifically with the presumed CRP cDNAs in the control rabbit liver was due to the absence of mRNA for CRP and was not due to underloading RNA in that lane of the gel.

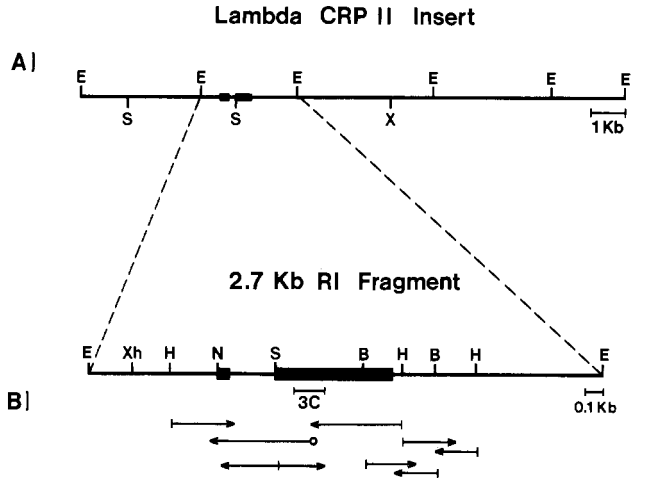

FIGURE 2: Restriction map of the rabbit CRP genomic DNA clone λCRP II and sequencing strategy. (A) Restriction map of λCRP II. The size of the λCRP II insert is 17 kb. It contains five *Eco*RI fragments of sizes (from left to right) 3.6, 2.7, 4.3, 4.1, and 2.1 kb. The protein coding region of the CRP gene as indicated by the black boxes is within the 2.7 kb *Eco*RI fragment. The relative position of cDNA clone is 3c is shown. Restriction endonuclease sites are (B) *Bam*HI, (H) *Hin*dIII, (N) *Nco*I, (S) *Sph*I, (X) *Xam*I, and (Xh) *Xho*I. (B) Sequencing strategy. Fragments were cloned into M13-derived vectors mp18 or/and mp19 and sequenced by the dideoxy chain termination procedure. Arrows indicate the direction and extent of sequencing. The position of a synthetic 18-mer from which sequencing was extended is indicated as an open circle.

The sequence of cDNA clone 3c was determined and is shown in Figure 3 along with its deduced amino acid sequence. Part of the nucleotide sequence of 3c, from nucleotides 2–94, matches the published protein sequence (Wang et al., 1982) from amino acids 33–63. This confirms that 3c is indeed a CRP cDNA. However, the sequence of 3c from nucleotides 95–168 did not match the published protein sequence following amino acid residue 63, even though it contains a single open reading frame which is continuous with that from nucleotides 2–94. 3c was then subcloned into fragments, one corresponding to the region homologous to the protein sequence of the cDNA and the other corresponding to the nonhomologous region of the clone. Both subclones, when used as probes in Northern blot experiments, hybridized to a 2.4 kb mRNA (data not shown) only in RNA from inflamed rabbit liver. We conclude that both portions of cDNA clone 3c are derived from CRP mRNA.

Isolation of the CRP Gene. To isolate the gene for CRP, we used cDNA clone 3c as a probe to screen a rabbit genomic libary. About 10^6 phage plaques were screened, and 7 plaques were identified. Digestion of DNA prepared from these phage with restriction enzyme *Eco*RI showed that they represented two overlapping clones, named λCRP I and λCRP II. These two clones were further restriction mapped. The restriction map of λCRP II is shown in Figure 2A. λCRP I has the same central three *Eco*RI fragments, but the lengths of 5′ and 3′ *Eco*RI fragments differ from λCRP II. Southern hybridization of DNAs from λCRP I and λCRP II showed that both cDNA clones hybridized within a 2.7 kb *Eco*RI fragment and that clone 2a was positioned 3′ relative to 3c. The nonhomologous subclone of cDNA 3c hybridizes to the same restriction fragment as does 3c.

To assess the size of the CRP gene, the five *Eco*RI fragments of λCRP II were subcloned into plasmid pUC9. Northern blot analyses were done by using these subclones as probes. Only the 2.7 kb *Eco*RI fragment and a 3.5 kb *Eco*RI/*Xma*I subfragment of the 4.3 kb *Eco*RI fragment immediately downstream from the 2.7 kb fragment showed significant hybridization to CRP mRNA. These results suggested that the CRP gene was within the 6.2 kb *Eco*RI to *Xma*I fragment.

Sequence of the CRP Gene. The protein coding region of the CRP genomic fragment and also sequences upstream from the gene were sequenced by the dideoxynucleotide chain termination procedure. Figure 3 shows the resulting nucleotide sequence and the amino acid sequence deduced from it. Only the amino acid sequence in-frame with the published rabbit CRP protein sequence is shown. In vitro translation studies had shown that CRP contains a signal peptide of about 2000 daltons. Our sequencing data showed that an in-frame ATG codon is 60 nucleotides 5′ to the GCA, the first codon of mature protein, and is likely to be the translation initiation codon. This ATG is in a nucleotide environment, ACCATGG, which has been shown to be the optimal sequence for initiation of translation by eukaryotic ribosomes (Kozak, 1986). This 20 amino acid stretch is rich in hydrophobic amino acids as expected for a signal peptide. Several nucleotides downstream from the transcription initiation site, which was identified by experiments described below, there is another possible translation initiation codon. However, three codons downstream from this ATG there is a stop codon present which would prevent it from being translated into a full-length product. Furthermore, nucleotides around this ATG do not have the conserved sequence thought to be required for a functional initiation codon.

The protein coding region of the gene is interrupted by a single intron of 252 bp in length. The interruption is within the codon for amino acid residue 2 of the mature protein. An in-frame stop codon is present within the intron which, unless removed, would cause premature termination of translation. The intron has the consensus splicing donor sequences AG at its 5′ end and splicing acceptor sequence GT at its 3′ end.

Mapping the 5′ and 3′ Termini of the CRP Gene. The 5′ terminus of the CRP gene was determined by both S1 nuclease protection and primer extension experiments, and the 3′ terminus of the CRP gene was determined by an S1 nuclease protection experiment. The strategies and results of these experiments are shown in Figures 4–6.

For S1 mapping the 5′ terminus of the gene, the probe used consisted of the *Eco*RI/*Nco*I subfragment of the 2.7 kb *Eco*RI fragment. As shown in Figure 4, a unique fragment of 110 bp was protected from S1 nuclease digestion. The intensity of the band is much stronger when employing RNA from the inflamed rabbit liver (lane 2) than it is from the control counterpart (lane 1), consistent with the results of Northern blot analysis but also demonstrating the increased sensitivity of this technique for measuring low levels of CRP mRNA.

To rule out the possibility that the S1-protected fragment was due to the presence of an intron at this position, a primer extension experiment was performed. In this experiment, the primer consisted of 5′ end labeled *Nco*I/*Hin*fI fragment 40 nucleotides in length. As shown in Figure 5, the primer-extended product (lane 3) is the same size as the S1-protected fragment (lane 1). This is expected if the 5′ end defined by the S1 protection experiment is the 5′ end of the transcript. A Maxam–Gilbert sequencing ladder of the same DNA fragment run side by side with an S1 nuclease experiment (Figure 4) positioned the transcription start site as 113 ± 2 nucleotides upstream from the ATG translation initiation codon (see Figure 3). Twenty and fifty-nine nucleotides, respectively, upstream from this presumed cap site are potential RNA polymerase II recognition signal "TATA" and "CAT" boxes. We conclude that there is a unique transcription start

```
TTCATTTTTCCTGTCACAAAAAGCTGGCAATTCGGTAGTCACAGGAGTTTGTAATAAAGTAACCCACATTG
ATTTCTCTGTTCCAGAATGATTTTGCTTCCCTTCCTCCCAGAGCCCTGACACTTTCCCTAAGAGGCGGTGT
AGAAAATCATTTTACAAAGTGGCACAGATTCCCCATCTACTGTGGATATAAATCCAGGCAGGATGGAGTGG
CGCTGAGGCAAGGGATGCAGGACTTCTAGTACCTGAGCCTTTCAGCCAGAGACGTTTTCTCCAAAGGAGTG
                                                -20
GGATTCTGAGCCTGCCTGGTAGCACTGGTGGCAGGGAGTGACC ATG GAG AAG CTG CTG TGG TGT
                                               MET Glu Lys Leu Leu Trp Cys
                                                  -1   1
TTC CTG ACT TTG GTC AGC TTC TCT AAT ATG TCT GAC CAG GCA G GTAAGGTGCCTAC
Phe Leu Thr Leu Val Ser Phe Ser Asn MET Ser Asp Gln Ala G
TCCAGGGTACGCCAGAGATTTGAACTGAGAGATAGGAATAGGGCCTGAGATTCTACAGACTGGAGCCAGAG
GAACATACATCACAGAGATAACATGCGTGTGCGTGTGTGTGTGTGTGTGTGTGTGTCTACCTGGG
TTAGTGTATATCTCAATCCCCAAATCTGTATGTGAATGAGGGGTAGATCTTTATGAACTAGAAATTTCTCA
                          2                                    10
TCACTACTTGGTTTTCTGTTCTACAG GC ATG CAC AAG AAG GCC TTT GTG TTC CCC AAA GAG
                           ly MET His Lys Lys Ala Phe Val Phe Pro Lys Glu
                                20                                   30
TCA GAT AAT TCC TAC GTG TCC CTC AAC GCA CAG TTA AAG AAG CCA CTC AAA GCC
Ser Asp Asn Ser Tyr Val Ser Leu Asn Ala Gln Leu Lys Lys Pro Leu Lys Ala
                                                Thr*                 Leu*
                                   40
TTC ACT GTG TGC CTC TAC TTC TAC ACT GAT CTG TCC ATG ACT CGT GGG TAC AGT
Phe Thr Val Cys Leu Tyr Phe Tyr Thr Asp Leu Ser MET Thr Arg Gly Tyr Ser
  . ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
50                                                60
ATT TTC TCC TAT GCC ACC AGG AGA CAA TTT AAC GAG ATC CTC CTG TTT TGG TCC
Ile Phe Ser Tyr Ala Thr Arg Arg Gln Phe Asn Glu Ile Leu Leu Phe Trp Ser
... ... ... ... ... ... .A. ... ... ... ... ... ... ... ... ... ... ...
                            Lys
         70                                       80
AAG GAC ATA GGA TAT AGT TTT TCA GTG GGT GGA GAT GAA ATA ATA TTC AAG GTT
Lys Asp Ile Gly Tyr Ser Phe Ser Val Gly Gly Asp Glu Ile Ile Phe Lys Val
... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
                     90                                      100
TCT GAC ATC CCT GTG GAT CCA ACT CAC CTC TGT GCA AGC TGG GAG TCC AGC ACA
Ser Asp Ile Pro Val Asp Pro Thr His Leu Cys Ala Ser Trp Glu Ser Ser Thr
... ... ..
                               110                                  120
GGC ATT GCA GAG CTC TGG GTA GAT GGG AAG CCC ATG GTG AGG AAG AGT CTG AAG
Gly Ile Ala Glu Leu Trp Val Asp Gly Lys Pro MET Val Arg Lys Ser Leu Lys
                                   130
AAG GGC TAC ATT TTG GGG CCA GAG GCA AGC ATT ATT CTG GGG CAG GAT CAG GAT
Lys Gly Tyr Ile Leu Gly Pro Glu Ala Ser Ile Ile Leu Gly Gln Asp Gln Asp
140                                               150
TCG TTT GGT GGA AGC TTT GAG AAA CAA CAG TCT TTG GTG GGA GAC ATT GGA AAT
Ser Phe Gly Gly Ser Phe Glu Lys Gln Gln Ser Leu Val Gly Asp Ile Gly Asn
                                  Trp*                                Asp*
         160                                      170
GTG AAC ATG TGG GAC TAT GCA CTT TCA CCA GAA GAG ATT AAT ACC ATC TAT GCT
Val Asn MET Trp Asp Tyr Ala Leu Ser Pro Glu Glu Ile Asn Thr Ile Tyr Ala
                                                                 Val*
                     180                                     190
GGT GGG ACC TTT AGT CCC AAT GTC CTA GAC TGG CGC GAG CTG ACA TAT CAA GTA
Gly Gly Thr Phe Ser Pro Asn Val Leu Asp Trp Arg Glu Leu Thr Tyr Gln Val

                               200
CGT GGT GAA GTA CAC GTG AAG CCC CAG CTA TGG CCC TGA GCCTCTGCCAAGGATCCTG
Arg Gly Glu Val His Val Lys Pro Gln Leu Trp Pro Ter
AAAGGTGCTTCTTAGGTTACAAGTCACAAGGCCCCATACTTCTGGCTGTGGACCTTTACCCCTACATATAC
TGAATGCCTGCTACATAAACATAGCTTCCTAGCTTTGCCTTCTTCAACACCAGAGAATACAAATTAAATAT
CTGAGGATCTTGTGG
```

FIGURE 3: Nucleotide sequences of the rabbit CRP genomic fragment and cDNA clone 3c. The sequence presented was derived from genomic clone λCRPII. The nucleotide sequence of cDNA clone 3c is shown as dots unless otherwise specified. The CAT box, TATA box, and transcription initiation site are underlined. The deduced amino acid sequence is shown beneath the nucleotide sequence and numbered starting with amino-terminal alanine of the mature protein. The 28 amino acid stretch which replaces the published protein sequence between residues 63 and 73 (Phe-Phe-Val-Lys-Glu-Val-Ile-Val-Tyr) is underlined. Other amino acid residues that differ from the published protein sequence are indicated by asterisks.

site for the CRP gene and that this start site is utilized in both inflamed and uninflamed liver.

Using the 3′ end labeled 4.3 kb *Eco*RI fragment, which is immediately 3′ to the 2.7 kb *Eco*RI fragment, as a probe, the 3′ terminus of the gene was also determined. As shown in Figure 6, a unique fragment of 480 nucleotides was protected from S1 digestion. Again, there was more protection when liver RNA from an inflamed animal was used. Since there are about 1070 bp of the 3′-nontranslated region within the 2.7 kb *Eco*RI fragment, the total length of this region is 1550 bp, consistent with the 2.4 kb size of CRP mRNA and 2.6 kb for the size of the gene. The structure of the rabbit CRP gene is shown in Figure 7.

Amino Acid Sequence of Rabbit CRP. After further examination of the sequence at the protein level, one point is worth noting. The amino acid sequence derived from the nucleotide sequence agrees with the published protein sequence for the most part except for the stretch from amino acid residues 63–73. In fact, this stretch of nanopeptide does not match any of our deduced amino acid sequence and cannot

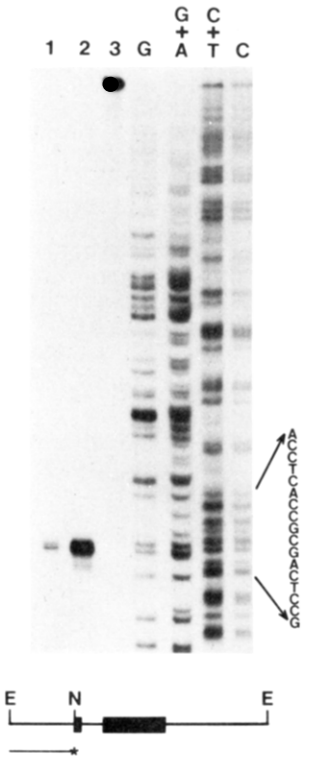

FIGURE 4: Mapping of the 5′ terminus of the rabbit CRP gene by S1 nuclease protection. The probe as indicated at the bottom of the panel was hybridized with RNA from uninflamed (lane 1) or inflamed (lane 2) rabbit liver and digested with S1 nuclease. Lane 3 is the undigested probe. A Maxam–Gilbert sequencing ladder of the same end-labeled DNA was run adjacent to the S1 nuclease reactions to position the 5′ terminus of the gene.

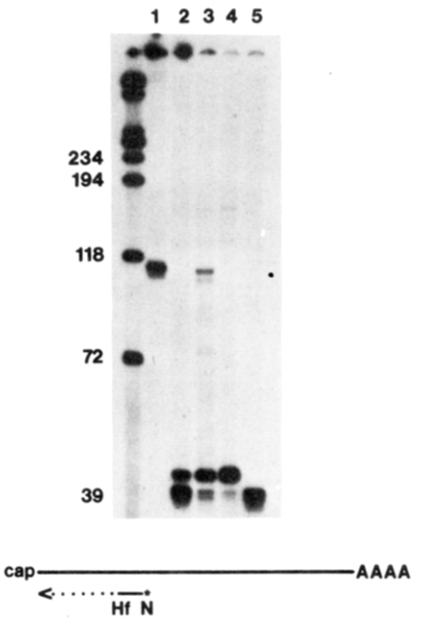

FIGURE 5: Mapping of the 5′ terminus of the rabbit CRP gene by primer extension analysis. ^{32}P-Labeled 40 bp *Nco*I/*Hin*fI primer as indicated at the bottom of the panel was annealed to the RNA from uninflamed (lane 2), inflamed (lane 3), or *E. coli* tRNA (lane 4), and the primer was extended with reverse transcriptase. Lane 1 is the S1 nuclease protected fragment from lane 2 of Figure 4. Lane 5 is primer by itself.

be found in the other two reading frames of the nucleotide sequence. Instead, an additional 28 amino acid stretch replaces the published amino acid residues from 63 to 73.

The exact explanation for the discrepancy is unclear. Genomic Southern analysis indicates that there is only one copy

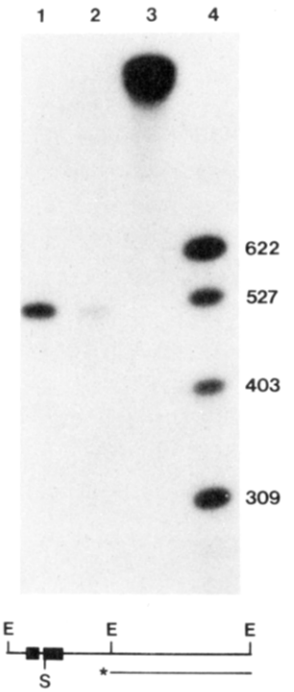

FIGURE 6: Mapping of the 3′ terminus of the rabbit CRP gene. The probe used is the 3′ end labeled 4.3 kb *Eco*RI fragment as indicated at the bottom of the panel. Lane 1, RNA from inflamed rabbit liver; lane 2, RNA from uninflamed rabbit liver; lane 3, undigested probe.

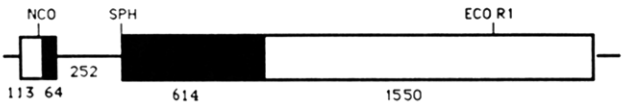

FIGURE 7: Structure of the rabbit CRP gene. The black boxes are the protein coding region of the exons. The open boxes are the 5′- and 3′-nontranslated regions of the gene. The intron is shown as a line. The size in base pairs of individual regions of the gene is indicated.

of the CRP gene in the rabbit genome. Thus, the possibility that there are two populations of CRP protein derived from two genes is unlikely. Also, as shown in Figure 3, 3c is colinear relative to the genomic sequence and is positioned at the junction of the amino acid sequence mismatch. We expect the abundance of the mature CRP mRNA relative to the primary transcript to be high. Since the 2.4 kb mature CRP mRNA was the only species detected in the Northern blot experiments using 3c or the subcloned nonhomologous portion of 3c as a probe, it is unlikely that cDNA clone 3c is derived from the primary transcript of CRP. The possibility that there is an unusual splicing of primary transcript in this region seems remote. Taken together, we believe that mature CRP protein consists of 205 amino acids instead of 186 amino acids as previously reported.

A similar observation was noted when the amino acid sequence deduced from a human genomic clone (Lei et al., 1985; Woo et al., 1985) was compared with the published human CRP protein sequence (Oliveira et al., 1979). In that case, an additional 19 amino acids were present between amino acid residues 61 and 62.

There is apparently some microheterogeneity in the nucleotide sequence of the CRP gene. As shown in Figure 3, at nucleotide 805, the genomic DNA has a G and cDNA 3c has an A which changes an arginine into a lysine. There are five amino acid residue mismatches between the deduced sequence and the protein sequence. Three of them are the result of single base changes.

Comparison between Human and Rabbit CRP Genes. The structure of the human CRP gene is similar to that of the

rabbit (for comparison, see Figure 7). The human CRP gene has 104 nucleotides of 5′-nontranslated region and a single intron of 278 nucleotides which interrupts the protein in the same position as that in the rabbit gene. As predicted from the nucleotide sequence, human CRP has a signal peptide of 18 amino acids. The 3′-nontranslated region of the human CRP gene is 1.2 kilobase pairs in length.

The nucleotide sequence of the 5′-nontranslated region and the 150 bp upstream from the transcription initiation site show a considerable amount (60–70%) of homology. The protein coding regions of the genes are also conserved, as expected (80% homology). However, the sequenced portions of the 3′-nontranslated region of the genes show very little homology. Unexpectedly, there is substantial sequence homology in the intron when the rabbit and human CRP genes are compared. Most notable is a run of $(GT)_{15}$ in the middle of the intron. In addition, there are two other regions of the intron which show significant homology; nucleotides 1–66 of the human intron and nucleotides 1–68 of the rabbit intron are 70% homologous, and nucleotides 241–278 of the human intron and nucleotides 215–252 of the rabbit intron are 80% homologous. Intron sequences between human and rabbit are generally not conserved even though several cases of intron homology between species have been reported. For example, human and rat α-tubulin genes have related introns (Hall & Cowan, 1985) as do the human and goat α-globin genes (Shapiro et al., 1983). In the former case, only the first intron shows conservation, each having a run of alternating GT. Whether this intron sequence conservation is simply due to evolutionary conservation of the CRP gene or has some functional significance remains to be determined. GT repeats have been found in the introns of human α-tubulin (Hall & Cowan, 1985) and genomes of yeast, *Xenopus*, mouse, and human (Miesfeld et al., 1981; Hamada & Kakunaga, 1983; Rogers, 1983). Several proposals have been made regarding the function of oligo(GT). Most interesting to us is that they might have enhancer functions by virtue of their ability to form left-handed DNA helices of Z-DNA.

Not much is known about the molecular mechanisms mediating induction of acute-phase reactants such as CRP. This paper confirms that the increased synthesis of CRP during an inflammatory response is accompanied by an increase in the accumulated level of CRP mRNA in the liver. Whether the regulation of CRP mRNA accumulation is at the level of mRNA transcription or mRNA stabilization is still not known. With the information presented in this paper, we will be able to further pursue the mechanism of CRP induction and to define the requirements for expression of the gene at the nucleotide sequence level.

Added in Proof

After submission of the manuscript, a sequence derived from cDNA clones from rabbit CRP was published (Syin et al., 1986). The cDNA sequence presented is in close agreement with that shown in Figure 2.

Acknowledgments

This work is part of an active collaboration with the laboratories of Drs. Irving Kushner and Steven Macintyre, Department of Medicine, Case Western Reserve University at Cleveland Metropolitan General Hospital. We thank them for many useful discussions and their critical reviews of the manuscript.

Registry No. DNA (rabbit liver C-reactive protein gene), 104848-53-5; C-reactive protein (rabbit liver precursor reduced), 104848-55-7; C-reactive protein (rabbit liver reduced), 104848-56-8.

References

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Berk, A. J., & Sharp, P. A. (1977) *Cell (Cambridge, Mass.) 12*, 721–732.

Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Gotschlich, E. C., & Edelman, G. M. (1965) *Proc. Natl. Acad. Sci. U.S.A. 54*, 558–566.

Hall, L. J., & Cowan, N. J. (1985) *Nucleic Acids Res. 13*, 207–223.

Hamada, H., & Kakunaga, T. (1983) *Nature (London) 298*, 396–399.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–589.

Kozak, M. (1986) *Cell (Cambridge, Mass.) 44*, 283–292.

Kushner, I. (1982) *Ann. N.Y. Acad. Sci. 389*, 39–48.

Kushner, I., & Somerville, J. A. (1971) *Arthritis Rhuem. 14*, 560–570.

Kushner, I., & Feldmann, G. (1978) *J. Exp. Med. 52*, 561–571.

Lei, K.-J., Liu, T., Zon, G., Soravia, E., Liu, T.-Y., & Goldman, N. D. (1985) *J. Biol. Chem. 260*, 13377–13383.

Luse, D. S., Haynes, J. R., Van Leeuwen, D., Schon, E. A., Cleary, M. L., Shapiro, S. G., Lingrel, J. B., & Rodder, R. G. (1981) *Nucleic Acids Res. 9*, 4339–4354.

Macintyre, S. S., Schultz, D., & Kushner, I. (1983) *Biochem. J. 210*, 707–715.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–566.

Miesfeld, R., Krystal, M., & Arnheim, N. (1981) *Nucleic Acids Res. 9*, 5931–5947.

Oliveira, E. B., Gotschlich, E. C., & Liu, T.-Y. (1979) *J. Biol. Chem. 254*, 489–502.

Pepys, M. B. (1981) *Lancet i*, 653–657.

Rogers, J. (1983) *Nature (London) 305*, 101–102.

Samols, D., Macintyre, S. S., & Kushner, I. (1985) *Biochem. J. 227*, 759–765.

Sanger, F., Nicklen, M., & Goulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5476.

Shapiro, S. G., Schon, E. A., Townes, T. M., & Lingrel, J. B. (1983) *J. Mol. Biol. 169*, 31–52.

Syin, C., Gotschlich, E. C., & Liu, T.-Y. (1986) *J. Biol. Chem. 261*, 5473–5479.

Tillett, W. S., & Francis, T. (1930) *J. Exp. Med. 52*, 561–571.

Tucci, A., Goldberger, G., Whitehead, A. S., Kay, R. M., Woods, D. E., & Colten, H. R. (1983) *J. Immunol. 131*, 2416–2419.

Wang, C. M., Nguyen, N. Y., Yonaha, K., Robey, F., & Liu, T.-Y. (1982) *J. Biol. Chem. 257*, 13610–13615.

Woo, P., Korenberg, J. R., & Whitehead, A. S. (1985) *J. Biol. Chem. 260*, 13384–13388.