Nucleic Acids Research, 1993, Vol. 21, No. 11 2747–2754

Synthetic gene for the hepatitis C virus nucleocapsid protein

Y.E.Khudyakov[1,2,*], H.A.Fields[1], M.O.Favorov[1,2], N.S.Khudyakova[1], M.-T.Bonafonte[1,3] and B.Holloway[1]

[1]Hepatitis Branch, National Center for Infectious Diseases, Centers for Disease Control, 1600 Clifton Rd, Atlanta, GA 30333, USA, [2]The D.I.Ivanovsky Institute of Virology, 16 Gamaleya St., Moscow 123098, Russia and [3]Molecular Biology Department, Biokit SA, Barcelona, Spain

Received December 7, 1992; Revised and Accepted April 30, 1993

EMBL accession no. X71611

ABSTRACT

A synthetic gene encoding the hepatitis C virus (HCV) nucleocapsid protein was constructed and expressed in *E.coli*. To synthesize this gene, we developed a new method that results in the enzymatic synthesis of long polydeoxyribonucleotides from oligodeoxyribonucleotides. The method, designated as the 'Exchangeable Template Reaction' (ETR), uses oligonucleotides as templates for DNA polymerase. A special mechanism was designed to exchange the templates during the polymerase reaction. The mechanism relies on the formation of a single-stranded 3′-protrusion at the 'growing point' of the elongating DNA such that it can be subsequently annealed, in a sequence-specific manner, with the next synthetic oligonucleotide. When annealed to the 3′-protrusion, the added oligonucleotide becomes a template for DNA polymerase, and the protruding 3′-end of the double-stranded DNA is used as the primer. The HCV nucleocapsid gene was assembled with DNA ligase from three fragments synthesized by ETR. The data verify that this method is efficient. The main advantage of ETR is the ability to combine more than two oligonucleotides in one tube together with polymerase and an enzymatic activity that produces a 3′-protrusion (e.g., *BstXI*) rather than the sequential addition of each component. The data demonstrate that as many as five oligonucleotides can be used simultaneously, resulting in a synthesized DNA fragment of designed sequence. The synthetic gene expressed in *E.coli* produced a 27kDa protein that specifically interacted with antibodies from sera obtained from HCV-infected individuals.

INTRODUCTION

Hepatitis C virus (HCV) is a recently identified agent responsible for most cases of post-transfusion non-A, non-B (NANB) hepatitis worldwide (1). The virus contains a positive strand RNA genome that comprises about 9,400 nucleotides (nt) and encodes for a polyprotein of more than 3,000 amino acids. The genetic organization of the HCV genome was recently elucidated (2–6). The N-terminal region of the HCV polyprotein is processed into structural proteins C (a nucleocapsid protein) and E1 and E2/NS1 (envelope proteins). Although the precise points for processing of the polyprotein have not been confirmed, the polyprotein contains four nonstructural proteins: NS2, NS3, NS4, and NS5. The HCV nucleocapsid protein C has been cloned and expressed in bacteria (7–9), and in eukaryotic cells (10). Because HCV circulates in low titers, a source of DNA using chemical-enzymatic synthesis would be of significant benefit.

The use of natural sources of DNA for expression of proteins has been greatly facilitated by PCR. Compared with conventional synthetic approaches for preparing DNA, PCR is much less complicated and labor intensive. Conventional DNA synthesis from oligonucleotides is fundamentally different from DNA synthesis that occurs within organisms. The natural process uses a pre-existing DNA template as a substrate for multiple enzymatic activities involved in DNA replication. However, to synthesize a gene *in vitro* of a desired sequence a pre-existing template does not exist, and one must rely on the use of relatively short single-stranded oligonucleotides (20 to 100 nucleotides). The use of several single-stranded oligonucleotides for the assembly of polynucleotides may result in many problems associated with complementary adverse interactions with each other. These factors are major contributors in restricting the efficiency and use of synthetic approaches for the assembly of DNA.

Two general methods exist for the assembly of oligonucleotides into long DNA fragments. The first method involves the synthesis of oligonucleotides comprising the entire sequence to be synthesized. After a set of these oligonucleotides is annealed, nicks are repaired with DNA ligase (11, 12). The fragment can then be cloned directly (13–20), or after amplification by PCR (21, 22). Although very sensitive to the secondary structure of each oligonucleotide, this method is used in many laboratories (23) despite its low efficiency. The second method uses polymerase to fill gaps in annealed pairs of oligonucleotides,

* To whom correspondence should be addressed at: Hepatitis Branch, National Center for infectious Diseases, Centers for Disease Control, 1600 Clifton Rd, Atlanta, GA 30333, USA

thereby converting single-stranded regions of oligonucleotides into double-stranded DNA (24–32). This approach is less dependent on the secondary structure of oligonucleotides than the first and appears to be more efficient. However, the size of the synthesized DNA fragments is limited by use of only two oligonucleotides for the synthesis. For assembly of longer DNA molecules the primary synthesized double-stranded segments must be digested with restriction endonucleases for further assembly by DNA ligation. Recently, PCR was used for the assembly of long DNA molecules from oligodeoxyribonucleotides (33, 34). These methods enable the synthesis of a gene from oligonucleotides that comprise only one of two strands. The efficiency of the methods is very high. However, repeated denaturation and annealing reactions created with each PCR cycle results in a relatively higher probability for alternative and adverse complementary interactions of oligonucleotides with itself and oligonucleotides with the synthesized DNA. Because of these side reactions, PCR cannot be routinely applied for the synthesis of DNA molecules longer than 1–2 kb. In this paper a method is described that imitates to some extent the natural template mechanism of DNA synthesis. With this method, efficient enzymatic synthesis (not assembly from pre-synthsized oligonucleotides) of long DNA fragments is achieved by using: (1) oligonucleotides as templates for a polymerase reaction, and (2) a unique cyclic mechanism for the introduction of new templates to continue the synthesis. Using this method, designated the 'Exchangeable Template Reaction' (ETR), a synthetic gene for the HCV nucleocapsid protein was constructed and subsequently expressed in *E.coli*.

MATERIALS AND METHODS

Sequence design

The sequence for synthesis of the gene encoding for the nucleocapsid protein of HCV was obtained from published data (35). The secondary structure of mRNA around the initiator codon was calculated according to published methods (36–38). The sequence of the mRNA 5'-untranslated region was designed by following published recommendations (36, 39). The primary structure of oligonucleotides was taken from published data (35) and modified as necessary to introduce *BstXI*-sites, or other sites for restriction endonucleases. In addition, the secondary structure of the oligonucleotides was predicted as described (37, 38), and complementary interactions between oligonucleotides for each set of oligonucleotides were taken into consideration. Whenever the 3'-terminal region of the oligonucleotides was potentially involved in undesirable interactions, the primary structure of the oligonucleotides was changed without influencing the protein sequence.

Oligodeoxyribonucleotide synthesis

Oligodeoxyribonucleotides were synthesized with an automatic synthesizer (Applied Biosystem Model 480A) and purified by electrophoresis in 10% PAGE containing 7 M urea in TBE buffer (0.045M Tris-borate, 0.001M EDTA, pH8.3). Oligonucleotides were recovered from the gel by electroelution using the model 230A HPEC system (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol.

Exchangeable template reaction conditions

ETR was performed in a 50 μl solution of 10 mM Tris–HCl, pH 7.9, containing 1 mM $MgCl_2$, 50 mM NaCl, and 1 mM DTT (*NEB2* buffer, New England BioLabs, Beverly, MA), 0.25 mM each dATP, dGTP, dTTP, and dCTP (Pharmacia, Piscataway, NJ); 5 units of native *Taq* DNA polymerase (Perkin Elmer Cetus, Norwalk, CT); 30 units of *BstXI* (New England BioLabs, Beverly, MA); and 0.5–100 pmol of each oligonucleotide, at 37°C for 0.5–5 hours, or overnight. In order to provide analysis of the reaction course, one of the oligodeoxynucleotides without a *BstXI* site was radiolabeled with [gamma-^{32}P]ATP in buffer: 50 mM Tris–HCl, pH7.6, containing 10 mM $MgCl_2$, 5 mM DTT with 10 μCi [gamma-^{32}P]ATP (5000 Ci/mmole, New England Nuclear) and 10–20 pmol of oligonucleotide. After completion the products of the ETR were analyzed by electrophoresis in 8% PAGE in TBE.

Assembly of the gene

ETR fragments were recovered from PAGE by incubating pieces of the gel in 100 μl of solution containing 0.15 M NaCl and 1 mM EDTA for 30 min. at 65°C. After incubation, the DNA was precipitated with 250 μl of ethanol. The pellet was dissolved in 20 μl of water, and 1 μl of the DNA solution was used to amplify the ETR fragments by PCR. The terminal oligonucleotides used for synthesis of the fragments were used as primers for PCR. Briefly, 20–50 pmol of each primer was added to the reaction mixture, followed by 30 cycles as follows: 94°C for 45 sec, 65° for 20 sec, and 72°C for 1 min. The amplified ETR fragments were treated with the appropriate restriction endonuclease with the recognition sites located at the termini of each fragment, and then ligated in 10 μl of a solution containing all three fragments, 50 mM Tris–HCl, pH 7.5, 10mM $MgCl_2$, 1 mM DTT, 1 mM ATP, and 10 units of DNA ligase (Pharmacia, Piscataway, NJ) for 6 h. One μl of the ligase reaction mixture was used to amplify the fragment by PCR to provide the full-length DNA using PCR conditions described above and using the two terminal oligonucleotides as primers. Amplified full-length DNA was recovered from agarose gel by a DEAE procedure (40) and treated with restriction endonucleases to confirm the structure of the synthesized gene.

Plasmid construction

A vector, designated *pTS7*, was constructed to prepare a plasmid for expression of the HCV nucleocapsid gene. Two oligonucleotides containing the gene 10 promoter from T7 phage and encoding for the 5'-untranslated region of the mRNA (Figure 1A) were annealed at 20°C and inserted into *pBR322* between the *EcoRI* and *BamHI* sites. A *BamHI-ScaI*-fragment of the plasmid containing this synthetic sequence and the *N*-terminal part of the *amp* gene were introduced into the plasmid *pSP65* between the *BamHI* site localized in the multiple cloning region and the *ScaI*-site of the *amp* gene. The resulting plasmid *pTS7* and the synthetic gene assembled from ETR fragments were treated with *NdeI* and *HindIII*, combined together, and ligated. *E.coli* HB101 competent cells (Invitrogene, San Diego, CA) were transformed with the ligation mixture. A plasmid containing the synthetic gene under the control of the T7 phage promoter was designated *pTSC6178-7*. The sequence of the cloned HCV synthetic gene was verified using the polymerase chain terminator method (41).

Analysis of expression

In order to express the protein encoded by the synthetic gene, *E.coli* BL21(DE3) (42) competent cells were prepared (43) and

transformed with *pTSC6178-7*. Cells were grown in LB medium until an optical density at 600 nm was equal to 0.6 after which the T7 promoter was activated by the addition of IPTG at a final concentration of 1 mM. After 4–6 h of growth at 37°C, the cells were harvested and a lysate was prepared (44). Aliquots of the lysate were analyzed by Western blot (45). Nitrocellulose filters containing immobilized proteins were incubated at 20°C for 2 h with human sera diluted 50 times in 50 mM tris HCl, pH7.5, containing 0.5% Triton X-100, 1% gelatin, and 1% bovine serum albumin (NET). The filters were washed with NET three times, and then incubated for 1 h with affinity chromatography purified anti-human IgG coupled to horseradish peroxidase (TAGO, Burlingame, CA) diluted 1:5000 in NET. After washing, diaminobenzidine (Sigma, St.Lous, MO) and hydrogen peroxide were used to develop the reaction.

Sera

Sera were obtained from a collection reposited at the D.I.Ivanovsky Institute of Virology, Moscow, Russia. All sera were initially tested by commercially available kits (*ABBOTT* Laboratories, Abbott Park, IL) for markers of hepatitis B, or hepatitis D infection, and for the presence of anti-HCV activity.

RESULTS

Expression system design

The nucleocapsid protein of HCV was expressed in *E.coli* as an authentic non-hybrid protein. One of the main problems for the efficient expression of non-hybrid proteins in bacteria is a poor understanding of the mechanism of how the mRNA structure may influence the efficiency of translational initiation (46). When designing our HCV protein C expression system, we used the sequence of the strong promoter of the bacteriophage T7 gene 10 (47) to provide efficient transcription (Figure 1A). Our strategy was to prepare a synthetic promoter rather than use the plasmid T7 promoter, because a synthetic sequence would allow for the creation of any desirable 5′-untranslated region of the mRNA. The secondary structure of the mRNA around the initiator codon of the HCV nucleocapsid protein was designed as a hairpin with an ATG codon exposed in the loop and the Shine-Dalgarno (SD) sequence located in the double-stranded region. This hairpin structure with a G = −2.8 kcal/mol may be easily destroyed by the interaction of the Shine-Dalgarno sequence (SD) with the 3′-end of the ribosomal 16S rRNA, resulting in the ATG start codon becoming exposed in the center of a large single-stranded region. A structure of this type has been hypothesized to be a very efficient ribosome binding site (36, 39). Another feature of this translation initiation region is an A-rich sequence downstream from the hairpin. This sequence has a low potential for secondary structure formation and separates the initiation region from the rest of the mRNA, thus decreasing the influence of the downstream sequence on the secondary structure within the translation initiation zone. Presence of the A-rich sequence makes the prediction of the hairpin more probable. Within the sequence of the coding region for the HCV protein C, a set of unique restriction endonuclease recognition sites was introduced (Figure 1B). Since this gene was to be expressed in bacterial and mammalian cells, codon usage was of no concern. The sequence for the synthesis of oligonucleotides was taken from published data, position 330–917 nt (35), and modified for insertion of restriction endonuclease sites, or modified to avoid adverse complementary interactions.

The mechanism of the exchangeable template reaction

Oligonucleotides for the synthesis of the HCV protein C gene were designed specially for ETR, a method for the synthesis of long polynucleotide DNA fragments using short synthetic oligonucleotides as templates for the DNA polymerase reaction. The method is based on a sequential cyclic mechanism involving three components: (1) DNA polymerase to synthesize double-stranded DNA, (2) enzymatic activity to create a 3′-terminal single-stranded region, and (3) specifically designed synthetic oligodeoxynucleotides used as templates for the polymerase reaction. The essential step is the enzymatic creation of a 3′-terminal single-stranded region at a 'growing point' of the synthesizing polynucleotide chain that will anneal with the next oligonucleotide (i.e., template) and thus allow for the continuation of the polymerase reaction. Restriction endonucleases, or other nucleases capable of creating a 3′-protrusion, such as T7 or lambda exonucleases, or other enzymes capable of melting double-stranded DNA at the termini, such as *recA*, may be used as the enzymatic activity to facilitate the introduction of new templates at the 3′-end of the growing DNA chain. Each cycle begins with the complementary binding of the 3′-terminal region of a synthetic oligonucleotide (step 1, Fig.2). After annealing,

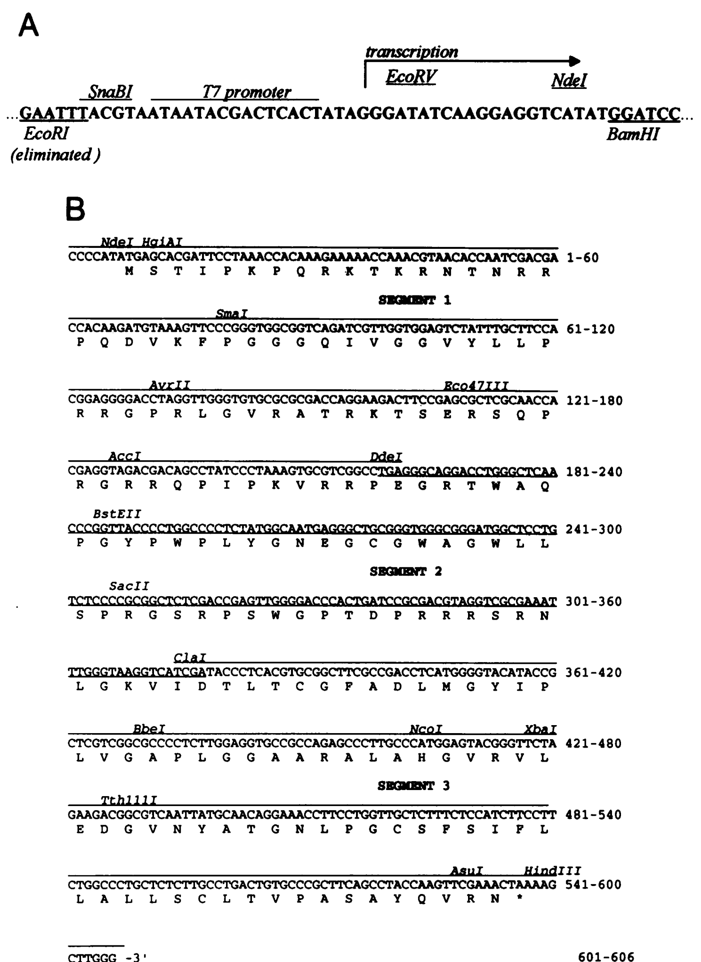

Figure 1. A) Nucleotide sequence of the T7 promoter and 5′-untranslated region of the mRNA encoding for the HCV nucleocapsid protein in the plasmid *pTSC6178-7*. **B**) Primary structure of the synthetic DNA fragment containing the coding region for the HCV nucleocapsid protein. Three segments of the synthetic genes synthesized by ETR are indicated.

EXCHANGEABLE TEMPLATE REACTION (ETR)

single stranded DNA or restricted double stranded DNA
5'...CCANNNNNNTGG...3'
3'...GGTNNNNNNACC...5'
BstXI site
anneal synthetic oligomer (NNNN_____BstXI)
STEP 1
polymerase
STEP 2
BstXI
STEP 3
CCANNNNN 3'
GTTN 5'
anneal next synthetic oligomer (NNNN_____BstXI)
STEP 1
CCANNNNN 3'
GTTNNNNN
polymerase
STEP 2

Figure 2. Cyclic mechanism of the exchangeable template reaction.

A

```
  BglII                      Oligo A
5'-CCCAGATCTCAATCTCGGGAATCTCAATGTTAGTATTCCTTGGACTCATAAGGTGGGAA

  BstXI                      Oligo B
5'-CCCCCACCACTCTGGATTAAAGATAGGTACTGTAGAGGAAAAAAGCGCCGTAAAGTTTCCCACCTTAT

  ApaI                       Oligo C
5'-CCCGGGCCCACAAATTGTTGACACCTATTAATAATGTCCTCTTGTAAATGAATCTTAGGAAAGGAAGGAGTTT
   GCCACT
```

B

```
  Nde I                      Oligo 1
5'-CCCCATATGAGCACGATTCCTAAACCACAAAGAAAAACCAAACGTAACACCAATCGACGACCACAAGATGTAA
   AGT

  BstXI                      Oligo 2
5'- CCCCCACCTCCGTGGAAGCAAATAGACTCCACCAACGATCTGACCGCCACCCGGGAACTTTACATCTTG

  BstXI                      Oligo 3
5'-CCCCCATCTTCCTGGTCGCGCGCACACCCAACCTAGGTCCCCTCC

  BstXI                      Oligo 4
5'-CCCCCAACCTCGTGGTTGCGAGCGCTCGGAAGTCTTC

  DdeI                       Oligo 5
5'-CCCCCTCAGGCCGACGCACTTTAGGGATAGGCTGTCGTCTACCTC

  DdeI                       Oligo 6
5'-CCCCCTGAGGGCAGGACCTGGGCTCAACCCGGTTACCCCTGGCCCCTCTATGGCAATGAGGGCTGCGGGT
   GGGCG

  BstXI                      Oligo 7
5'-CCCCCAGATCAGTGGGTCCCCAACTCGGTCGAGAGCCGCGGGGAGACAGGAGCCATCCCGCCCACCCGCA
   G

  ClaI                       Oligo 8
5'-CCCATCGATGACCTTACCCAAATTTCGCGACCTACGTCGCGGATCA

  ClaI                       Oligo 9
5'-CCCATCGATACCCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC

  BstXI                      Oligo 10
5'-CCCCCAACTCCATGGGCAAGGGCTCTGGCGGCACCTCCAAGAGGGGCGCCGACGAGCGGTAT

  BstXI                      Oligo 11
5'-CCCCCAGGAAGATGGAGAAAGAGCAACCAGGAAGGTTTCCTGTTGCATAATTGACGCCGTCTTCTAGAACCC
   GTACTCC

  HindIII                    Oligo 12
5'-CCCAAGCTTTTAGTTTCGAACTTGGTAGGCTGAAGCGGGCACAGTCAGGCAAGAGAGCAGGGCCAGAAGGA
   AG
```

Figure 3. Sequences of synthetic oligonucleotides used in ETR.

```
Oligo A
    BglII
5'-CCCAGATACT---------ATAAGGTGGGAA          Oligo B
                      TATTCCACCCTT------GGTCTCACCACCCCC-5'
                                             BstXI
                     | polymerase
                     v
Oligo A
5'-CCCAGATACT---------ATAAGGTGGGAA------ccagagtggtggggg-3'
3'-gggtctatga---------TATTCCACCCTT------GGTCTCACCACCCCC-5'
                                             Oligo B
                     | BstXI
                     v
5'-CCCAGATACT---------ATAAGGTGGGAA------ccagagtg
3'-gggtctatga---------TATTCCACCCTT------GGTC
                     | oligo C
                     v
5'-CCCAGATACT---------ATAAGGTGGGAA------ccagagtg
3'-gggtctatga---------TATTCCACCCTT------GGTC||||
                                            TCAC---CCCGGGCCC-5'
                     | polymerase           Oligo C   ApaI
                     v
  BglII
5'-CCCAGATACT---------ATAAGGTGGGAA------ccagagtg---gggcccggg-3'
3'-gggtctatga---------TATTCCACCCTT------GGTCTCAC---CCCGGGCCC-5'
                                                     ApaI
```

Figure 4. Mechanism of ETR synthesis of a short HBV-specific sequence using oligonucleotides A, B, and C.

a DNA polymerase reaction creates a second strand of DNA using a short synthetic oligonucleotide as a template (step 2, Fig. 2). After the polymerization is complete, the double-stranded DNA has been extended by the length of the synthetic oligonucleotide. Another cycle of DNA synthesis is initiated by an enzymatic reaction that removes several nucleotides from the 5'-terminus, leaving a 3'-protrusion that can be used to anneal another short synthetic oligonucleotide (step 3, figure 2). One example of an enzymatic reaction capable of removing several nucleotides from the 5'-terminus is the restriction endonuclease *BstXI*. This restriction endonuclease is compatible with ETR for the following reasons: (1) a 3'-protrusion is produced, (2) the single-stranded 3'-protrusion lacks sequence restrictions, and (3) after cleavage the restriction site cannot be restored by the interaction of the next synthetic oligonucleotide. The specificity of oligonucleotide addition for each cycle is encoded in each 3'-terminal sequence. At the 3'-terminus of the growing DNA molecule, which occurs following enzymatic reactions by those enzymes mentioned above, a specific sequence of nucleotides anneals with a complementary sequence of nucleotides from the synthetic oligonucleotide in the mixture. Thus, the ETR synthesizes long DNA fragments in one step by simply combining an entire set of oligodeoxynucleotides in one reaction tube containing all of the required enzymatic activities, and incubating the mixture at the appropriate temperature.

Model oligonucleotides for ETR

The first ETR synthesized product was a DNA fragment encoding for the hepatitis B virus (HBV) terminal protein. Although not directly related to this study, the sequence was specifically modified to test a functional motif. This fragment was created from three oligonucleotides designed specifically for ETR (Fig.3A and 4). The mechanism of ETR using these oligonucleotides is shown in Figure 4. All three oligonucleotides were combined in one tube with a DNA polymerase and *BstXI* in the presence of dNTP's. Oligonucleotide A (oligo A) was radiolabeled at the 5'-end. Oligo A was used at a concentration of 1 pmol; oligo B, at 10 pmol; and oligo C, at 100 pmol. The reaction was incubated at 20°C or 37°C for 5 h. Two different types of polymerases were tested in the reaction: *E.coli* DNA polymerase I (large fragment), and *Taq* polymerase (recombinant and natural). Although all of the polymerases were able to fill in the gap during the first step of the reaction (Figure 4), allowing *BstXI* to cleave the synthesized duplex in all cases, only natural *Taq* polymerase produced a fragment of the appropriate size for a 'trimer' (Figure 5). Apparently, the recombinant *Taq*

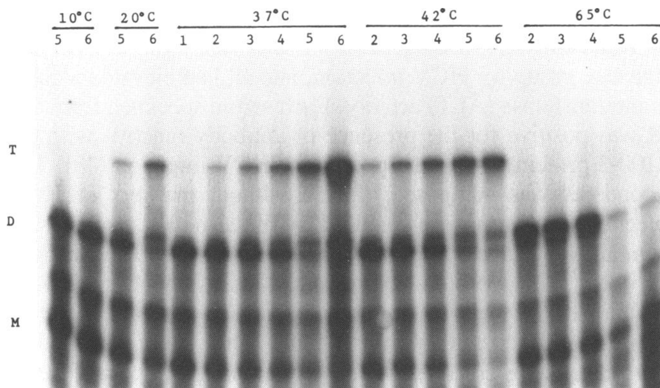

Figure 5. Effect of temperature and time of incubation of ETR reaction mixture containing oligonucleotides A (radiolabeled), B, and C on the efficiency of DNA synthesis. Autoradiogram of an 8% PAGE with 8 M urea in TBE. Products of the ETR after 5 min incubation (1), 30 min (2), 1 h (3), 2 h (4), 5 h (5), and 16 h (6) are marked M for A ('monomer'), D for A+B ('dimer'), and T for A+B+C ('trimer').

polymerase was missing an essential function important for this reaction. In subsequent experiments only natural *Taq* polymerase was used. In experiments using radiolabeled oligo A, a full-sized fragment was identified after electrophoresis at denaturating conditions. However, when radiolabeled oligo C was used for analysis of the reaction, only oligo C was observed by autoradiography. This result was reproducible and suggested that only oligo A initiated polymerase synthesis of a full size DNA strand using oligo B and C as templates. The full-size double-stranded DNA product of ETR contains a non-interrupted strand synthesized by the polymerase reaction primed with oligo A, and a second strand with nicks between oligonucleotides participating in the reaction as templates. These nicks can be repaired with DNA ligase. Even if not repaired, these nicks do not prevent cloning of the fragment with plasmids, amplification by PCR, or cleavage by restriction endonucleases.

Synthesis of a DNA fragment encoding the HCV nucleocapsid protein

The sequence of the HCV nucleocapsid protein was divided into three fragments. Each fragment was synthesized by ETR (Figure 6 and 7). The first fragment was synthesized using five oligodeoxynucleotides, the second fragment using three, and the third fragment using four (Figure 3B and 6). All reactions were carried out as described above. Fragments obtained using four and five oligonucleotides consisted of 228 and 216 bp, respectively. The yield of full-size fragments was approximately 5−10% after a 14 h incubation period at 37°C (Figure 7). Different buffers were tested for the synthesis. Buffer *NEB3* (New England Biolabs, Beverly, MA; 50 mM Tris−HCl, pH 7.9, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT) was optimal for the *BstXI* reaction, and *Taq* buffer (10 mM Tris−HCl, pH 7.8, 50 mM KCl, 1 mM DTT, 1,5 mM, or 3 mM, or 5 mM $MgCl_2$) was optimal for *Taq* DNA polymerase. The best results for ETR, however, were obtained with buffers *NEB2* (see Materials and Methods, *Exchangeable Template reaction conditions*) and *NEB4* (New England Biolabs, Beverly, MA; 20 mM Tris-acetate, pH 7.9, 10 mM Magnesium acetate, 50 mM Potassium acetate, 1 mM DTT). Both enzymes have high optimal temperatures, but because of the very short single-stranded protrusion formed by

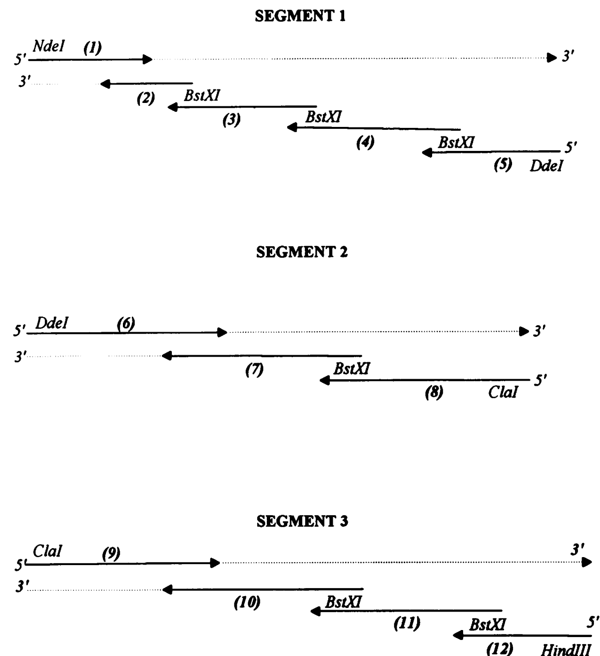

Figure 6. Strategy for ETR synthesis of the segments of the HCV nucleocapsid protein gene.

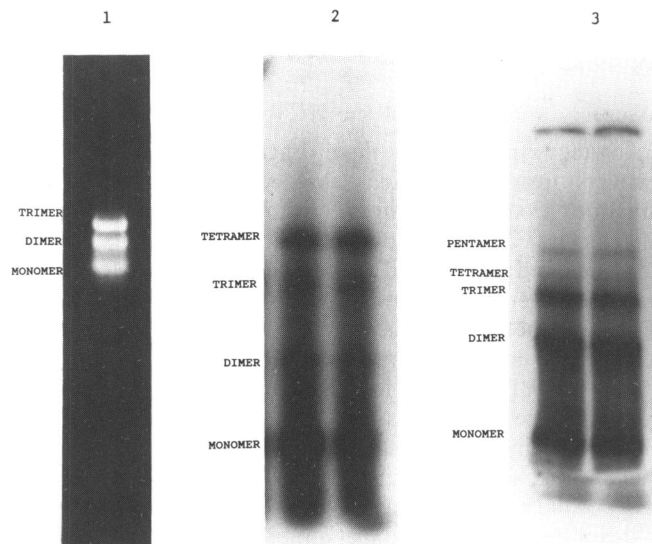

Figure 7. Products of the ETR demonstrating different segments of the HCV nucleocapsid gene. Agarose gel stained with ethidium bromide is shown for the segment that was synthesized using 3 oligonucleotides (1). Autoradiograms of 8% PAGE with 8M urea in TBE are shown for the segments that were synthesized using 4 oligonucleotides (2) and 5 oligonucleotides (3). The top band of the autoradiogram for No.3 marks the top of the gel.

BstXI, ETR fails to work at elevated temperatures. As shown previously for the HBV DNA fragment, 37°C was optimal for buffers *NEB2* and *NEB4* for both enzymatic activities. For the synthesis of the first segment of the HCV nucleocapsid gene, the optimal relative concentrations of the oligodeoxynucleotides were 1:4:20:40:60. When the relative concentrations were changed to 1:1:20:40:60, the rate of ETR changed as well. At

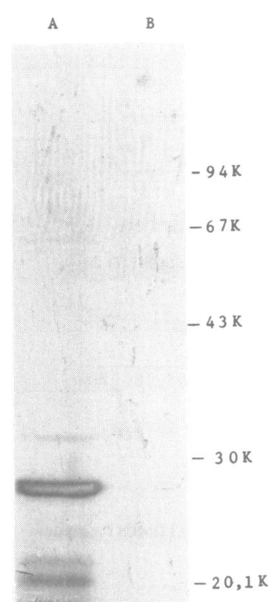

Figure 8. Western blot analysis of the HCV core protein encoded by *pTSC6178-7*.

a 1:4 relative concentration of oligos 1 and 2, a full-size fragment was observed after a 3 h incubation period at 37°C in *NEB2*. At the 1:1 relative concentrations of oligos 1 and 2, a fragment was synthesized in detectable amounts after only 30 min (data not shown).

Assembly and cloning of the HCV gene

Following synthesis, all three fragments were purified by recovery from PAGE, and subsequently amplified by PCR. Amplified products were digested with the appropriate restriction endonuclease and joined by DNA ligase. The whole gene was then amplified by PCR and analyzed by restriction site mapping. Following verification by restriction analysis, the fragment containing the synthetic gene was inserted into the expression vector *pTS7* under the control of the T7 promoter. This vector was specifically designed for the expression of the nucleocapsid synthetic gene (see expression system design and Material and Methods). Among the first five clones chosen for analysis, one contained a plasmid, designated *pTSC6178-7*, that contained an insert of the appropriate size. Analysis of the primary structure of this insert demonstrated that all three fragments of the ETR-synthesized DNA had the correct sequence. Thus, the gene for the nucleocapsid protein of the hepatitis C virus was assembled in the correct form by the new method and cloned.

Expression of the synthetic gene in *E.coli*

Plasmid *pTSC6178-7* was unstable in transformed bacteria growing with or without the IPTG inducer. Modified derivatives of *pTSC6178-7* were often found in bacteria after 2–3 passages on solid LB agar or in liquid medium. Because of this problem, the cells were transformed each time the HCV protein was to be expressed, or transformed cells were stored frozen and used as aliquots. In cells transformed with plasmid *pTSC6178-7*, a 27kDa protein identified by Western blot analysis effectively bound an antibody from sera containing serologic markers specific for HCV infection (Figure 8). Sera without markers of HCV infection did not immunoreact with this protein. To confirm the specific reactivity of the protein with anti-HCV antibody, we collected two sets of 24 sera. Both sets were from donors in a high risk group for HCV infection, and all had elevated alanine aminotransferase (ALT) activity. Each serum specimen from one set was positive for the presence of antibody reactive with the C100-3 protein, a marker specific for HCV infection (48). The second set was composed of sera without markers of HCV infection. Both sets were analyzed by Western blot. Among sera from HCV-positive donors, 20 sera contained antibodies reactive with the 27kDa protein expressed in cells transformed with *pTSC6178-7*. Two sera from the second set were also reactive with this protein. Although anti-C100-3 has been identified in most cases of HCV infection, this marker may appear late in infection or not at all (49, 50). Because both sets of sera came from high risk-group donors, it is likely that some of the HCV-positive sera would be negative for anti-C100-3 activity. Thus, the synthetic gene assembled from three fragments synthesized by ETR is functionally active, and the product of its expression is specific for HCV. The data suggest that this protein can be used for the development of a diagnostic test for the detection of antibody specific for the HCV nucleocapsid protein.

DISCUSSION

Using the *BstXI* restriction endonuclease to create the 3′-protrusion in the ETR, we synthesized four DNA fragments of 161 bp, 187 bp, 216 bp, and 228 bp. The maximum number of ETR cycles tested was four using 5 oligodeoxynucleotides combined in one mixture. Our results demonstrate that ETR is an efficient method for the synthesis of DNA from oligonucleotides. The most restrictive step in the reaction is BstXI cleavage, which appears to be relatively slow at 37°C, especially when terminally located. Another problem is the very short 3′-terminal single stranded region formed by *BstXI*. The temperature limitation for ETR is apparently related to the BstXI reaction, which results in a short region (4 nt) available for complementary binding to the next oligonucleotide. Methods that would increase the length of the 3′-protrusion undoubtedly would increase the efficiency of the reaction. The melting temperature for complexes of four-nucleotide-long complementary regions is much less than 37°C, the optimal temperature for ETR. However, since the reaction produces a product at 37°C, *Taq* polymerase, even at suboptimal conditions, apparently stabilizes these transient complementary complexes of oligonucleotides. This finding suggests that the introduction of other methods to obtain longer terminal single-stranded regions at the 'growing point' of the double-stranded DNA will improve the ETR by allowing for a more reliable process of template exchange. Theoretically, a 5′-exonuclease specific for double-stranded DNA can be used in the ETR. Exonucleases specific only to the terminus of the growing DNA would eliminate the sequence limitation caused by restriction endonucleases. There are no strict requirements on the structure of the synthetic oligonucleotides used in this technique except that the secondary structure must allow the 3′-terminus, the 'working region' of the molecule, to be available for annealing. The rest of the nucleotide sequence may have any secondary or tertiary structure without adversely influencing the efficiency of the reaction, since the macrostructure of the growing DNA molecule is usually destroyed by the DNA polymerase reaction during the polymerization process.

Recently, PCR has been applied to the synthesis of long DNA molecules using synthetic oligonucleotides (21, 22, 33, 34). These

applications are based on the extreme efficiency of PCR to amplify DNA molecules synthesized in small quantity during assembly from oligonucleotides. However, there is a theoretical limitation to the routine application of PCR for the synthesis of molecules longer than 1–2 kb. It is well-known that long DNA molecules are often difficult to amplify by PCR. The difficulty is likely caused by an increased possibility for long DNA molecules to misplace annealing of primers, or to the formation of strong conformation after the denaturation of the double-stranded structure. The reaction cocktail for the synthesis of DNA from oligonucleotides by PCR is more complex than the usual cocktail for PCR since it contains a number of short oligonucleotides in addition to the terminal primers used for amplification of a full-length product. With increasing growth of the DNA, the probability of incorrect annealing of these oligonucleotides is also increased. An additional parameter influencing the efficiency of synthesis of long DNA molecules using PCR is the increased probability of oligonucleotides specific to internal sequences of growing DNA to anneal to complementary sites. In this case, synthesis of DNA will be terminated at the place where an internal oligonucleotide is annealed thereby decreasing the whole efficiency of this method to synthesize full-length molecules. The presence of multiple species of DNA of different lengths, an expected consequence of using PCR to synthesize DNA, also may complicate the synthetic process by allowing the exposure of different DNA conformations of the same sequence in different DNA molecules, which may interfere with the primer annealing process. A method has been developed to use PCR for DNA synthesis from oligonucleotides that is almost free from side reactions caused by the simultaneous presence of multiple species of oligonucleotides and the growing DNA (34). The method is based on the stepwise use of only two primers for DNA elongation by PCR at each step. Each subsequent step of DNA elongation is initiated by the product synthesized in the previous PCR. This method is a very efficient application of PCR to the synthesis of DNA from synthetic oligonucleotides. However, this approach is particularly inconvenient requiring a number of rounds of PCR equal to the number of pairs of oligonucleotides used for DNA synthesis. Moreover, repetitive employment of PCR for DNA synthesis increases the probability of obtaining a final product with some modifications of the primary structure. Thus, the main disadvantage of the PCR application to the synthesis of long DNA molecules from relatively short oligonucleotides is the denaturation step intrinsic for PCR that destroys double-stranded DNA exposing the DNA in a single-stranded form for subsequent interaction with primers. This factor influences the efficiency of DNA synthesis from oligonucleotides to a much greater extent than the efficiency of PCR amplification of pre-existing DNA fragments. ETR, employing two enzymatic activities and oligonucleotides for the DNA synthesis, has been specifically designed to avoid denaturation of double-stranded DNA and at the same time, to allow annealing of oligonucleotides to the growing DNA molecule for the continuation of DNA synthesis. This multi-enzymatic mechanism was found to be sufficiently efficient to synthesize in one reaction vessel a DNA fragment composed of 5 oligonucleotides. As mentioned above, with a longer 3′-protrusion a significant improvement in ETR efficiency would be expected. Experiments for identification of ETR conditions allowing for the creation of a 3′-protrusion of any desirable size are in progress.

Three of the fragments synthesized by the ETR were used for the assembly of a functionally active gene encoding for the HCV nucleocapsid protein capable of specifically binding an antibody from the sera of patients infected with HCV. By analyzing 48 sera from high-risk-group donors, we found a high correlation between antibody to the C-100-3 protein, a small part of a non-structural HCV protein (48), and the nucleocapsid protein of the virus. Only four sera positive for anti-C100-3 activity were negative by Western blot analysis using the expressed nucleocapsid protein. This observation suggests that the nucleocapsid protein should represent a very important and useful component in various HCV diagnostic systems. Among 24 sera negative for anti-C100-3 activity, two were positive for anti-nucleocapsid activity by Western blot analysis. Thus, the gene encoding the nucleocapsid protein produced by ETR may help to identify more HCV-positive sera than those found positive using test systems that detect anti-C100-3 alone.

REFERENCES

1. Plagemann, P. G. W. (1991) Arch. Virol. 120, 165–180.
2. Miller, R. H., Purcell, R. H. (1990) Proc. Natl. Acad. Sci. USA 87, 2057–2061.
3. Takeuchi, K., Kubo, Y., Boonmar, Y., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, E., Miyamura, T. (1990) J. Gen. Virol. 71, 3027–3033.
4. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P., Weiner, A. J., Bradley, D. W., Kuo, G., Houghton, M. (1991) Proc. Natl. Acad. Sci. USA 88, 2451–2455.
5. Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Shimotohno, K. (1991) Proc. Natl. Acad. Sci. USA 88, 5547–5551.
6. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I., Okayama, H. (1991) J. Virol. 65, 1105–1113.
7. Muraiso, K., Hijikata, M., Ohkoshi, S., Cho, M.-J., Kikuchi, M., Kato, N., Shimotohno, K. (1990) Biochem. Biophys. Res. Commun. 172, 511–516.
8. Muraiso, K., Hijikata, M., Kato, N., Shimotohno, K., Okazaki, N., Ohkoshi, S., Uura, M., Kaneko, S., Kobayashi, K., Omata, M. (1991) Japan J. Cancer Res. 82, 879–882.
9. Takahashi, K., Okamoto, H., Kishimoto, S., Munekata, E., Tachibana, K., Akahane, Y., Yoshizawa, H., Mishiro, S. (1992) J. Gen. Virol. 73, 667–672.
10. Harada, S., Watanabe, Y., Takeuchi, K., Suzuki, T., Katayama, T., Takabe, Y., Saito, I., Miyamura, T. (1991) J. Virol. 65, 3015–3021.
11. Agarwal, K. L., Buchi, H., Caruthers, M. H., Gupta, N., Khorana, H. G., Kleppe, K., Kumar, A., Ohtsuka, E., Raj Bhandary U. L., van de Sande, J. H., Sgaramella, V., Weber, H., Yamada, T. (1970) Nature 227, 27–34.
12. Khorana, H. G. (1979) Science 203, 614–625.
13. Edge, M. D., Green, R. A., Heathcliffe, G. R., Meacock, P. A., Schuch, W., Scanlon, D. P., Atkinson, T. C., Newton, R. C., Markhom, A. F. (1981) Nature 292, 756–762.
14. Mandecki, W., Mollison, K. W., Bolling, T. J., Powell, B. C., Carter, G. W., Fox, J. L. (1985) Proc. Natl. Acad. Sci. USA 82, 3543–3547.
15. Sproat, B. S., Gait, M. J. (1985) Nucleic Acids Res.13, 2959–2977.
16. Ferretti, L., Karnik, S. S., Khorana, H. G., Nassal, M., Oprian, D. D. (1986) Proc. Natl. Acad. Sci. USA 83, 599–603.
17. Denefle, P., Kovari, K.S., Guitton, J.-D., Cartwright, T., Mayaux, J.-F. (1987) Gene 56, 61–70.
18. Hostomsky, Z., Smrt, J., Arnold, L., Tocik, Z., Paces, V. (1987) Nucleic Acids Res. 15, 4849–4856.
19. Bell, L. D., Smith, J. C., Derbyshire, R., Finlay, M., Johnson, I., Gilbert, R., Slocombe, P., Cook, E., Richards, H., Clissold, P., Meredith, D., Powell-Jones, C. H., Dawson, K. M., Carter, B. L., McCullagh, K. G. (1988) Gene 63, 155–163.
20. Wosnick, M. A., Barnett, R. W., Carlson, J. E. (1989) Gene 76, 153–160.
21. Jayaraman, K., Shah, J. (1987) Nucleic Acids Res. 17, 4403.
22. Jayaraman, K., Fingar, S. A., Shah, J., Fyles, J. (1991) Proc. Natl. Acad. Sci. USA 88, 4084–4088.
23. Brousseau, R., Sung, W., Wu, R., Narang, S. A. (1987) In Narang, S.A.(ed.), Synthesis and applications of DNA and RNA. Academic Press, New York, pp.95–114.

24. Rossi, J. J., Kierzek, R., Huang, T., Walker, P. A., Itakura, K. (1982) J. Biol. Chem. 257, 9226–9229.
25. Scarpulla, R. C., Narang, S. A., Wu, R. (1982) Anal. Biochem. 121, 356–365.
26. Rink, H., Liersch, M., Sieber, P., Meyer, F. (1984) Nucleic Acids Res. 12, 6369–6387.
27. Derbyshire, K. M., Salvo, J. J., Grindley, N. D. F. (1986) Gene 46, 145–152.
28. Adams, S. E., Johnson, I.D., Braddock, M., Kingsman, A. J., Kingsman, S. M., Edwards,R.M. (1988) Nucleic Acids Res. 16, 4287–4298.
29. Uhlmann, E. (1988) Gene 71, 29–40.
30. Weiner, M. P., Sheraga, H. A. (1989) Nucleic Acids Res. 17, 7113.
31. Chen, H.-B., Weng, J.-M., Jiang, K., Bao, L.-S. (1990) Nucleic Acids Res. 18, 871–878.
32. Ciccarelli, R. B., Loomis, L. A., McCoon, P. E., Holzschu, D. L. (1990) Nucleic Acids Res. 18, 1243–1248.
33. Jayaraman, K., Puccini, C. J. (1992) BioTechniques 12, 392–398.
34. Majumder, K. (1992) Gene 110, 89–94.
35. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., Shimotohno, K. (1990) Proc. Natl. Acad. Sci. USA 87, 9524–9528.
36. Khudyakov, Y. E. (1985) Mol. Biol. (Russia) 19, 702–716.
37. Frier, S. H., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., Terner, D. H. (1986) Proc. Natl. Acad. Sci. USA 83, 9373–9377.
38. Jaeger, J. A., Turner, D. H., Zuker, M. (1989) Proc. Natl. Acad. Sci. USA 86, 7706–7710.
39. Khudyakov, Y. E., Kalinina, T. I., Neplyueva, V. S., Smirnov, V. D. (1987) Mol. Biol. (Russia) 21, 1504–1512.
40. Dretzen, G., Bellard, M., Sassone-Corsi, P., Chambon, P. (1981) Anal. Biochem. 112, 295–298.
41. Sanger, F., Nicklen, S., Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
42. Studier, F. W., Moffott, B. A. (1986) J. Mol. Biol. 189, 113–130.
43. Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
44. Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) Molecular Cloning: A laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor, Vol. 3, pp. 18.40–18.41.
45. Harlow, E., Lane, D. (1988) Antibodies. A laboratory manual. Cold Spring Harbor Lab., pp. 471–510.
46. Gold, L., Stormo, G. (1987) In Neidhardt, F.C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., Umbarger, H. E. (eds), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. American Society for Microbiology, Washington, DC, Vol.2, pp. 1302–1307.
47. Tabor, S., Richarson, C. C. (1985) Proc. Natl. Acad. Sci. USA 82, 1074–1078.
48. van der Poel, C. L., Cuypers, H. T. M., Reesink, H. W., Weiner, A.J., Quan, S., di Nello, R., van Boven, J.J.P., Winkel, I., Mulder-Folkerts, D., Exel-Oehlers, P. J., Schaasberg, W., Leentvaar-Kuypers, A., Polito, A., Houghton, M., Lelie, P.N. (1991) Lancet 337, 317–319.
49. van der Poel, C. L., Reesink, H. W., Schaasberg, W., Leentvaar-Kuypers, A., Bakker, E., Exel-Oehlers, P.J., Lelie, P. N. (1990) Lancet 335, 558–560.
50. Ebeling, F., Naukkarinen, R., Likola, J. (1990) Lancet 335, 982–983.