A PCR-based method for uniform $^{13}\text{C}/^{15}\text{N}$ labeling of long DNA oligomers

Xian Chen, S.V. Santhana Mariappan, John J. Kelley III, John H. Bushweller, E. Morton Bradbury, Goutam Gupta

Abstract A polymerase chain reaction (PCR)-based method is described for uniform $^{13}\text{C}/^{15}\text{N}$ labeling of DNA duplexes. In this method, multiple copies of a blunt-ended duplex are cloned into a plasmid with each copy containing the sequence of interest and the restriction HinII sequences at the 5′ and 3′ ends. PCR with uniformly $^{13}\text{C}/^{15}\text{N}$-labeled dNTP precursors results in a labeled DNA duplex containing multiple copies of the sequence of interest. Use of bi-directional primers, instead of self-priming [Louis et al. (1998) J. Biol. Chem. 273, 2374–2378], produces a DNA fragment of unique length. Twenty-four cycles of PCR of this purified product followed by restriction and purification gives (with 30% yield) the uniformly $^{13}\text{C}/^{15}\text{N}$-labeled duplex sequence for multi-nuclear magnetic resonance spectroscopy.

© 1998 Federation of European Biochemical Societies.

Key words: Uniform labeling of DNA duplexes; Polymerase chain reaction; Multi-nuclear magnetic resonance; High resolution structure

1. Introduction

Since the advent of well-developed and standard techniques for the synthesis of uniformly $^{13}\text{C}/^{15}\text{N}$-labeled proteins and RNA, multi-nuclear magnetic resonance (NMR) spectroscopy has been routinely applied for the determination of their structures in solution [1–3]. By comparison, the determination of DNA structures by multi-dimensional NMR is rather scarce primarily due to the lack of economic and efficient techniques for the synthesis of uniformly $^{13}\text{C}/^{15}\text{N}$-labeled DNA molecules. The availability of uniform $^{13}\text{C}/^{15}\text{N}$ labeling allows resolution and sequential assignment of spin systems belonging to individual nucleotides in a long DNA duplex which, in turn, result in the collection of a large set of distance constraints essential for determining the high resolution structure. This is particularly important in delineating the structure of DNA in its complex with a specific protein because NMR spectroscopy with uniform DNA labeling will not only enable us to identify specific DNA-protein contacts but will also help us to examine whether there is any conformational change in DNA upon protein binding.

In principle, labeled DNA molecules can be chemically or enzymatically synthesized using uniformly $^{13}\text{C}/^{15}\text{N}$-labeled nucleotides as precursors [4,5]. DNA labeling by chemical synthesis has not been popular because the preparation of uniformly $^{13}\text{C}/^{15}\text{N}$-labeled protected mononucleotides required for phosphoramidite chemistry is often costly and technically challenging. DNA labeling by enzymatic methods appears more appealing since the preparation of uniformly $^{13}\text{C}/^{15}\text{N}$-labeled deoxynucleoside triphosphates (dNTPs) is less laborious than making uniformly $^{13}\text{C}/^{15}\text{N}$-labeled protected mononucleotides [6]. However, the biggest obstacle in making labeled DNA of a defined length by an enzymatic method is to control the length of chain elongation. This has been overcome by suitable design of hairpin templates in which the single-stranded overhang in the stem controls the length of the newly synthesized DNA strand [6,7]. Alternatively, an RNA primer is used in the enzymatic synthesis of labeled DNA of defined length [8]. In these methods, uniformly $^{13}\text{C}/^{15}\text{N}$-labeled DNA duplexes are obtained by adding equimolar amounts of two complementary strands made from two different templates. Therefore, inherent in these methods is a step in which the newly synthesized labeled strand is separated from the unlabeled template. Here, we report a method based upon polymerase chain reaction (PCR) for large-scale synthesis of DNA duplexes which does not require separation of the labeled DNA from the unlabeled one. In addition, the product to template ratio is 2:1 for $n$ PCR cycles rather than 1:1 for the template-based method.

Two-dimensional hetero-nuclear magnetic resonance spectra are shown to illustrate the utility of uniform $^{13}\text{C}/^{15}\text{N}$ labeling of DNA for resolving resonance overlap. Detailed studies on the high resolution structure of this duplex based upon three-dimensional nuclear Overhauser effect spectroscopy (NOESY) data will be presented elsewhere. The high resolution structure of the yeast replication origin is of biological significance since the structure determines the mechanism of the initiation of replication in which specific protein factors are involved. It has been postulated that the replication origin sequence is intrinsically bent which may have important implications in the binding of proteins to this replication origin [9,10]. Our structural studies by multi-dimensional NOESY experiments will enable us to confirm whether this yeast replication origin sequence is, indeed, bent.

2. Materials and methods

2.1. Materials

$^{13}\text{C}/^{15}\text{N}$-labeled dNTPs with isotopic enrichment greater than 99% were purchased from Martek Bioscience (Columbia, MA). Taq DNA polymerase, Qiaquick PCR purification kit and QIAquick gel extraction kit were purchased from Qiagen (Chatsworth, CA). Twelve strip 0.2 ml PCR reaction tubes were obtained from Fisher Scientific (Pittsburgh, PA). The restriction endonuclease HinII was purchased from New England Biolabs (Beverly, MA). Escherichia coli XL2-blue competent...
petent cells were purchased from Stratagene (La Jolla, CA). X-gal and IPTG were purchased from Sigma. Sephacryl S-200 and Q-Sepharose were obtained from Pharmacia (Alameda, CA). Ultrafiltration devices and Centriprep-3 (molecular weight cut-off of 3 kDa) were obtained from Amicon (Beverly, MA).

2.2. Preparation of the PCR template of a recombinant DNA plasmid containing the TS (steps 1–2)

The target sequence (TS) was synthesized with a blunt-end Hinfl site at both 5′ and 3′ end (A in Fig. 1). After phosphorylation, monomers of the TS were ligated to produce multiple copies of the TS plus the Hinfl sites at the two ends. The ligation product (B in Fig. 1) was sized on an 0.8% agarose gel. Bands corresponding to the largest number of the tandem repeats were cut from the gel and purified by QiAquick gel extraction kit.

The purified oligomer was then inserted at the Hinfl site of the pUC19. After transformation into E. coli XL2-blue cells, the cells were plated on solid LB media containing 2% X-gal and 1 mM IPTG on the surface. The recombinant clones containing inserts of the template with labeled dNTPs as precursors. The reaction mixture from the individual recombinant plasmid. The recombinant, pUC19-TS (C in Fig. 1), with a maximum repeat number of TS (n=19) was selected for the PCR template and further purified on a 0.5% agarose gel.

2.3. PCR amplification and 13C/15N labeling of the target sequence (steps 3–4)

Optimized PCR primers were selected by a computer program (DNA Star) that gave melting temperatures of 54°C for the primer attached duplexes. The upper primer, 5′-AACGAGCGGCAGT-GAAT-3′, is located 50 bp away from the Hinfl site and the lower primer, 5′-ACCATGATTACGCCAAGC-3′, is located 34 bp away from the Hinfl site (see Fig. 1).

The plasmid (pUC19-TS) DNA (100 ng per 100 μl) was first used as the template with labeled dNTPs as precursors. The reaction mixture was loaded on a 1.0% agarose gel. The labeled product was cut off from the gel and purified by QiAquick gel extraction kit. This labeled product contained primer annealing sites and multiple copies of the TS that was used later for massive PCR reactions (D in Fig. 1).

The 100 μl PCR reactions consisted of the following:

1. 1.0 μl of 10× PCR buffer (1× in 100 μl)
2. 2. μl of 25 mM MgCl₂ (2.0 mM)
3. 3. μl of 20 μM upper primer (5.9 μg/ml, 1 μM)
4. 4. μl of 20 μM lower primer (5.9 μg/ml, 1 μM)
5. 5. 100 ng of 13C/15N labeled template (1 μg/ml)
6. 6. 0.5 unit of Taq DNA polymerase (5 units/ml)
7. 7. 200 μM (20 nmol, 6.6 μg) of each 13C/15N-labeled dNTPs.

100 μl PCR reactions [11] were carried out on a 96-well Perkin Elmer GeneAmp PCR system 9600. Each reaction tube was pre-denatured at 96°C for 30 s before the thermal cycle, i.e. denaturation at 96°C for 15 s, annealing at 50°C for 30 s, and extension for 1 min at 72°C. As shown in Fig. 2A, the extension reaction of the upper primer will stop at the position of the 3′ end of the lower primer while the extension of the lower primer will be terminated at the 5′ end of the upper primer. The PCR products are about 400 bp in size and contain the TS.

The semi-purified sample was loaded on a Sephacryl S-200 (2.5×50 cm) column with a DNA exclusion limit of 30 bp. This allows separation of the TS from other digested DNA fragments. The column was then washed with 1 mM phosphate buffer (pH 7.0) with a 0–1.5 M NaCl gradient at a flow rate of 1 ml/min. Five ml fractions were collected in each tube and the target DNA sequence eluted at approximately 160 mM NaCl.

The semi-purified sample was loaded on a Sephacryl S-200 (2.5×50 cm) column with a DNA exclusion limit of 30 bp. This allows separation of the TS from other digested DNA fragments. The column was subsequently eluted with 1 mM phosphate buffer (pH 7.0) containing 0.1 mM EDTA at a flow rate of 1.0 ml/min collecting 3 ml fractions. A aliquot from each tube was examined by PAGE to locate the target oligonucleotide which eluted at approximately 160 mM NaCl. Fractions containing TS were pooled and concentrated to 5 ml by ultrafiltration.

2.4. Isolation of the labeled TS by Hinfl digestion and purification of the labeled target sequence (shown in shades)

mM EDTA were used to pre-wash the column to remove non-DNA impurities. The column was then washed with 1 mM phosphate buffer (pH 7.0) with a 0–1.5 M NaCl gradient at a flow rate of 1 ml/min. Five ml fractions were collected in each tube and the target DNA sequence eluted at approximately 160 mM NaCl.

The column was subsequently eluted with 1 mM phosphate buffer (pH 7.0) containing 0.1 mM EDTA at a flow rate of 1.0 ml/min collecting 3 ml fractions. A aliquot from each tube was examined by PAGE to locate the target oligonucleotide which eluted at approximately 160 mM NaCl. Fractions containing TS were pooled and concentrated to 5 ml by ultrafiltration.

2.5. Degree of labeling

Our PCR product contains 19 copies of the target sequence. The cut-off sizes of flanking sequences from both primer ends are 50 bp and 34 bp, respectively. The flanking regions are also labeled with 13C/15N, i.e. (50+34)×(18×2)=48 bp. The ratio of the labeled target sequence to the labeled flanking region is calculated as: number of
monomer TS, \( n \), times the length of TS divided by 48. Therefore, the efficiency of labeling the TS increases with larger numbers of tandem repeats. For \( n = 19 \), percentage of labeled target sequence in undigested PCR product will be 87.

2.6. NMR spectroscopy

NMR experiments were performed on a Bruker DRX-500 spectrometer at LANL. In order to circumvent the problem of low DNA concentration and poor signal-to-noise ratio, we performed \(^{13}\text{C}/^{1}\text{H} \) experiments on a 720 MHz spectrometer at Florida State University (FSU). A susceptibility matched NMR tube that holds 180 \( \mu \)l of the sample was used for all measurements. 1 mM DNA sample in buffered solution of 10 mM sodium phosphate (pH 7.0) and 150 mM NaCl was used for all NMR measurements. \(^{1}\text{H} \) chemical shifts were measured with reference to 3,3,3-trimethylsilylethylpropionate as an internal standard. \(^{15}\text{N} \) chemical shifts were measured with reference to \( \text{KNO}_3 \) and converted into the primary reference of nitromethane in accordance with the protocol described by Witanowski et al. [12]. The \(^{15}\text{N} \) excitation was centered at 276.06 ppm from saturated \( \text{KNO}_3 \). The \(^{13}\text{C} \) chemical shifts were measured from tetramethylsilane (TMS) as a secondary reference. The \(^{13}\text{C} \) excitation was centered at 100 ppm from TMS. 100 ppm for \(^{13}\text{C} \) occurs at the decoupler offset of 0 Hz in \( \text{H}_2\text{O} \) or \( \text{D}_2\text{O} \). Two-dimensional heteronuclear \((^{15}\text{N}-^{1}\text{H} \) and \(^{13}\text{C}-^{1}\text{H} \) single and multiple quantum coherence (HSQC and HMQC) experiments were performed on the labeled 16-mer duplex [13,14]. In all the experiments, water suppression was achieved via the WATERGATE pulse scheme [15]. The NMR data were processed on a Silicon Graphics workstation (Indigo 2) with Felix software (Version 2.3, Biosym Tech.).

3. Results

3.1. The PCR method

After \( n \) thermal cycles of PCR, \( 2^n \) copies of the target sequence are produced. This DNA amplification procedure is much more efficient than the template-driven method in which the target sequence is produced at a 1:1 ratio with the template DNA. The labeled dNTPs are incorporated as efficiently as the naturally occurring dNTPs by PCR since the enrichment of the stable \(^{13}\text{C}/^{15}\text{N} \) isotopes has no effect on PCR. Taq polymerase transfers a non-template nucleotide adduct to the 3’ end of each strand of the PCR products due to its terminal deoxynucleotidyl transferase activity. However, this problem is eliminated by introducing a restriction site at both ends of the target sequence. In addition, the length of the labeled DNA of a particular sequence is controlled by the restriction site.

The length of the PCR product extended from upper and lower primers is dictated by the following mechanism. In the first cycle, single-stranded DNA is derived from the strands of the parent plasmid with a nick at the 5' hydroxyl end of each primer. In the following PCR cycles, each primer is annealed at its complementary strand and the extension is stopped at the nick position, i.e. as shown in Fig. 2A, the extension reaction of the upper primer stops at the 3' end of the lower primer while the extension of the lower primer is terminated at the 5' end of the upper primer. After \( n \) cycles of PCR, \( 2^n \) copies of ‘d’ are produced from the pUC19-TS template containing 19 copies of the TS.

Initially, we have produced the labeled PCR template ‘d’ on a large scale. All the nucleotides incorporated in this template ‘d’ are labeled except the primer ends. The PCR amplification of the DNA oligomer from this template gives labeled prod-

Fig. 2. Gel electrophoresis of PCR product and its HinII digests. A: 1% agarose gel electrophoresis of a portion (5 ml out of 100 ml) of PCR products with different concentrations of \(^{13}\text{C}/^{15}\text{N}-\text{labeled dNTPs when pUC19-TS is used as a template (2.5 ng). Lane 1, 5 mM; lane 2, 10 mM; lane 3, 20 mM; lane 4 shows various double-stranded markers. B: 4–20% gradient polyacrylamide gel electrophoresis of a portion (10 ml out of 0.5 ml) of the purified labeled PCR products after HinII digestion. F1: flanking fragment from upper primer to HinII site; F2: flanking fragment from lower primer to HinII site; TS: labeled target sequence (16-mer). Lanes 1 and 2: double-stranded DNA markers; lane 3: partially digested PCR products (10 min of HinII digestion at 37°C); lanes 4, 5, and 6: complete HinII digestion of the PCR products.
ucts with 99% enrichment of $^{13}\text{C}/^{15}\text{N}$ isotopes (i.e. the same as those of the dNTPs). HincII digestion of the PCR product and subsequent purification result in the 16-mer duplex for NMR studies.

Here, we have obtained only 19 copies of the insert in the plasmid. The number of copies can be increased to 256 using the procedure of Robinett et al. [16]; they have shown that 256 direct repeats of the lac operator can be inserted in a plasmid. Incorporation of such a high copy number of the target sequence will further enhance the efficiency of our PCR method. While this manuscript was in review, we came across the only other PCR-based method of Louis et al. [17]. In their method, the chain extension is based upon self-priming which leads to the synthesis of multiple copies of the sequence of interest (with the flanking restriction sites). In our method, we precisely control the number of copies of the sequence of interest, i.e. fragment D in Fig. 1. The actual sequence and the length of the sequence of interest determine the maximum number of copies that can be inserted in the plasmid. Incorporation of the technique reported by Robinett et al. [16] will allow us to insert about 100 copies of a sequence of 30–40 nucleotides. The larger the number of copies, the higher the efficiency of this method for double-stranded and single-stranded DNA.

Although we have restricted our discussion to the synthesis of uniformly labeled DNA duplexes, our method can easily be extended to the synthesis of uniformly labeled single-stranded DNA. The modification of the method would involve a thiol group attachment at the 5' end of one primer that is extended to produce the strand of interest. After isolation, fragment D (Fig. 1) would be treated with exonuclease. The thiol-modified strand would be protected whereas the other strand would be digested. The thiol-modified single strand is then annealed with GTC_GAC which would produce HincII digestion sites (marked _) on the single strand at regular intervals. Therefore, HincII treatment of the thiol-modified strand and subsequent purification create the single-stranded sequence of interest plus three nucleotide fragments at each end. It is not clear how the method of Louis et al. [17] based on self-priming can be extended to the synthesis of labeled single strands.

3.2. NMR spectroscopy

Fig. 3A shows the imino-imino 500 MHz NOESY connectivities of the duplex, d(G1-A2-C3-A4-T5-T6-A7-T8-G9-T10-T11-T12-A13-G14-T15-C16)d(G17-A18-C19-T20-A21-A22-A23-C24-A25-T26-A17-A28-T29-G30-T31-C32) recorded in water at 5°C. A: 150 ms $^1$H-$^1$H NOESY spectrum showing the imino-imino sequential connectivities. B: Selected region of the $^{15}\text{N}$-$^1\text{H}$ HSQC spectrum showing the cytosine amino resonances. The $^{15}\text{N}$ excitation was centered at 276.06 ppm from saturated KNO$_3$. $^{13}\text{N}$ of 90 Hz was employed in all the heteronuclear correlation experiments. The acquisition parameters for the HSQC experiment: sweep width and the number of data points in the proton dimension = 25 ppm and 2000; sweep width in the $^{15}\text{N}$ dimension = 15 ppm; t1 increments = 64; number of scans = 32; relaxation delay = 1.3 s. Panel B has the same D1 values as panel A.

Fig. 3. NMR spectra of the $^{13}\text{C}/^{15}\text{N}$-labeled duplex, d(G1-A2-C3-A4-T5-T6-A7-T8-G9-T10-T11-T12-A13-G14-T15-C16)d(G17-A18-C19-T20-A21-A22-A23-C24-A25-T26-A17-A28-T29-G30-T31-C32) recorded in water at 5°C. A: 150 ms $^1$H-$^1$H NOESY spectrum showing the imino-imino sequential connectivities. B: Selected region of the $^{15}\text{N}$-$^1\text{H}$ HSQC spectrum showing the cytosine amino resonances. The $^{15}\text{N}$ excitation was centered at 276.06 ppm from saturated KNO$_3$. $^{13}\text{N}$ of 90 Hz was employed in all the heteronuclear correlation experiments. The acquisition parameters for the HSQC experiment: sweep width and the number of data points in the proton dimension = 25 ppm and 2000; sweep width in the $^{15}\text{N}$ dimension = 15 ppm; t1 increments = 64; number of scans = 32; relaxation delay = 1.3 s. Panel B has the same D1 values as panel A.
The overlap of the non-exchangeable base protons (H8/H6/H2) is removed by 13C-1H HMQC experiments at 720 MHz. Fig. 4A,B shows such correlations involving C8/C6-H8/H6 and C2-H2. Note that the H8 protons of adenines and guanines are clustered as two different groups and so are the H6s of cytosines and thymines. The H2s of adenines are located at 20 ppm higher field than the corresponding H8s.

4. Discussion

The power of NMR spectroscopy with labeled DNA is clearly demonstrated in Fig. 4A,B which shows how resonance overlap in the 1H dimension is removed by 13C labeling. As shown in Fig. 4, resolution of the base (H8/H6/H2) protons belonging to 32 different nucleotides will allow identifications of their interactions with other base and sugar protons. This rich data set produces a large number of interproton distance constraints that are needed to obtain a high resolution structure of a DNA duplex of this length. For the ARS-A sequence, DNA labeling is particularly useful in determining the interactions involving H2s of adenines which are helpful in identifying intrinsic DNA bending. The bending of ARS-A may be important in defining the nucleosome boundary in the yeast replication origin since the ARS elements (i.e. B1, B2, B3, and A) should all be free of nucleosomes before the start of replication.

Acknowledgements: This work was supported by the Human Genome Project of the Office of Health and Environmental Research from the Department of Energy. We thank Dr. C. Unkefer and Stable Isotope Resource for their support during various stages of this work. We thank Dr. N. Murrali of the National High Magnetic Field Laboratory at FSU for providing access to the 720 MHz spectrometer. We thank Ms. S. Thompson for synthesis and purification of some of the DNA oligomers and Ms. Yuh Jen Ueng for her help in DNA sequencing. This work was partially supported by Los Alamos Grant 8E10/ XADH.

References