Journal of Biomolecular NMR (2006) 36:137–146
DOI 10.1007/s10858-006-9075-0

Article

General method of preparation of uniformly ^{13}C, ^{15}N-labeled DNA fragments for NMR analysis of DNA structures

Brigitte René†, Grégoire Masliah†, Loussiné Zargarian, Olivier Mauffret & Serge Fermandjian*

Département de Biologie et Pharmacologie Structurales, UMR 8113 CNRS – LBPA Ecole Normale Supérieure de Cachan, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France

Received 18 April 2006; Accepted 27 July 2006

Key words: DNA, isotope labeling, NMR, PCR, structural analysis

Summary

^{13}C, ^{15}N labeling of biomolecules allows easier assignments of NMR resonances and provides a larger number of NMR parameters, which greatly improves the quality of DNA structures. However, there is no general DNA-labeling procedure, like those employed for proteins and RNAs. Here, we describe a general and widely applicable approach designed for preparation of isotopically labeled DNA fragments that can be used for NMR studies. The procedure is based on the PCR amplification of oligonucleotides in the presence of labeled deoxynucleotides triphosphates. It allows great flexibility thanks to insertion of a short DNA sequence (linker) between two repeats of DNA sequence to study. Size and sequence of the linker are designed as to create restriction sites at the junctions with DNA of interest. DNA duplex with desired sequence and size is released upon enzymatic digestion of the PCR product. The suitability of the procedure is validated through the preparation of two biological relevant DNA fragments.

Introduction

NMR spectroscopy continues to emerge as the most powerful tool to unravel 3D structures of large biomolecules and complexes in solution, close to their physiological conditions (Ferentz and Wagner, 2000). This has been greatly aided by the development of ultra high-field magnets, of new acquisition-processing modules and of isotope labeling methods. ^{15}N, ^{13}C labeling of proteins has been introduced many years ago and has dramatically changed the way in which their structures can be determined by NMR spectroscopy (Fesik and Zuiderweg, 1990; Clore and Gronenborn, 1994a, b). This is also true for the isotope labeling of RNAs (Batey et al., 1992, 1995; Nikonowicz and Pardi, 1992; Michnicka et al., 1993; Hines et al., 1994; Varani et al., 1996). On another hand, the labeling of DNAs remains difficult, largely due to the lack of simple and efficient methods supplying the required products in NMR amounts. Obviously, chemical procedures permit the greatest flexibility in labeling schemes, but their use is restricted by the need of expensive labeled chemicals and the complexity of the synthesis itself (Ono et al., 1995; Fernandez et al., 1998; Tjandra et al., 2000; Kojima et al., 2001). The enzymatic methods are based either on the DNA polymerase fill-in (Zimmer and Crothers, 1995; Smith et al., 1997; Masse et al., 1998) or on the PCR reactions (Louis et al., 1998; Chen et al., 1998; Werner et al., 2001)

†The first two authors contributed equally to this work.
*To whom correspondence should be addressed. E-mail: sfermand@igr.fr

and use ^{15}N and/or ^{13}C labeled deoxynucleotide triphosphates (dNTPs) as precursors.

Here, we describe a general approach in which an oligonucleotide containing two copies of the desired DNA sequence coupled through a short linker is submitted to PCR amplification in the presence of labeled dNTPs. The sequence of the linker is selected to form a restriction site at the linker-desired DNA sequence junction. Actually, the number of possible restriction sites grows from 40 without linker to 1600 with a linker so that one can easily find an appropriate restriction endonuclease that liberates the exact desired DNA sequence. The suitability of the method is validated through the preparation of two biologically relevant DNA duplexes uniformly labeled with ^{15}N and ^{13}C: (i) a 21 mer, termed P site, deriving from the strongest cleavage site (site 22) in pBR322 for the enzyme topoisomerase II in presence of an ellipticine derivative (Fossé et al., 1991); and (ii) a 21 mer, termed LTR site, reproducing the U5-LTR (long terminal repeat) extremity of the human immunodeficiency virus type 1 (HIV-1) cDNA that encompasses the viral integrase (IN) attachment and processing site (Katzman and Katz, 1999; Renisio et al., 2005).

NMR spectra are provided to illustrate both the purity of labeled DNA duplexes and the interest of using ^{13}C and ^{15}N isotope labeling for analysis of DNA structure.

Materials and methods

Primers

Unlabeled oligonucleotides were purchased from Eurogentec (Belgium).

P site: the two, 51-nucleotides-long, DNA sequences used for PCR were as follows: P51S:

5' ACAGCTTATCATCGATCAC**GTAC**TTG
ACGTACAGCTTATCATCGATCACGT 3',

P51I:

5' ACGTGATCGATGATAAGCT**GTAC**GTC
AAGTACGTGATCGATGATAAGCTGT 3'

Each strand contained two repeats of P site (underlined) connected by a linker (dotted). The stretch in bold letters indicates the *Rsa* I recognition sites.

Site LTR: the two, 51-nucleotides-long, DNA sequences used for PCR were as follows: U51S:

5' ACTGCTAGAGATTTTCCA**CACGTG**GA
TA**GTAC**TGCTAGAGATTTTCCACAC 3',

U51I

5' GTGTGGAAAATCTCTAGCA**GTAC**TAT
C**CACGTG**TGGAAAATCTCTAGCAGT 3'

Each strand contained two repeats of the HIV-1 U5-LTR extremity (underlined) connected by a linker (dotted). The stretch in bold letters indicates the *Bsa*A I and the *Rsa* I recognition sites.

Synthesis of uniformly ^{15}N, ^{13}C-labeled 21 mers DNA Oligonucleotides by PCR

The two steps of PCR were performed in Hybaid thermal cyclers (Thermo Electron Corporation, Waltham, MA) featured with heated lid for oil free operations.

The 50 µl PCR step 1 reactions consisted in the following: 5 µl of 10 × Pfu buffer (Promega, Madison, WI); 72 nM of each primer (S and I); 2.5 U of Taq DNA polymerase (5 U/µl; New England Biolabs, Beverly, MA); 0.2 U of Pfu DNA polymerase (3 U/µl; Promega, Madison, WI); 0.2 mM of uniformly ^{13}C, ^{15}N-labeled dNTPs (Spectra Stable Isotopes, Columbia, MD or Silantes GmbH, Munich, Germany). PCR reactions were carried out on Thermo Hybaid PCR Sprint thermal cycler. Each reaction tube was pre-denatured at 95 °C for 4 min (“hot start”) before the thermal cycle, i.e. denaturation at 95 °C for 1 min, annealing at 55 °C for 2 min, and extension for 3 min at 72 °C. After 14 cycles, thermal cycling was followed by incubation at 72 °C for 7 min. The amplified mixture (600 µl) was used to further amplify 3 ml of step 2 reaction mixture (300 µl of 10 × Pfu polymerase buffer, 150 U of Taq DNA polymerase (5 U/µl); 12 U of Pfu DNA polymerase (3 U/µl); 0.2 mM of uniformly ^{13}C, ^{15}N-labeled dNTPs) in a 96 wells plate Thermo Hybaid PCR Express thermal cycler with a gradient block module and temperature control software (50 µl aliquots): one time 95 °C for 4 min, 55 °C for 2 min, 72 °C

for 2 min; ten times 95 °C for 1 min, 55 °C for 2 min, 72 °C for 2 min; ten times 95 °C for 1 min, 55 °C for 2 min, 72 °C for 2 min (with +0.05 min per cycle); fifteen times 95 °C for 1 min, 55 °C for 2 min, 72 °C for 2 min (with +0.03 min per cycle); one time 95 °C for 1 min, 55 °C for 2 min, and 72 °C for 7 min. The PCR products showed a large size distribution on a 0.8% agarose gel. PCR products (3 ml) were pooled, desalted by steric exclusion chromatography on Sephadex™ G25 prepacked columns (GE Healthcare, Life Sciences, UK) and digested overnight at 37 °C with 800 U of *Rsa* I (80 U/µl; Promega, Madison, WI) (P site) or with 800 U of *Rsa* I and 150 U of *Bsa*A I (5 U/µl, New England Biolabs, Beverly, MA) (LTR site) in a final volume of 3.5 ml containing 350 µl of buffer C 10 × (Promega, Madison, WI) and 35 µl of bovine serum albumin (10 mg/ml) (Promega, Madison, WI). The digestion products were checked by polyacrylamide gel electrophoresis (PAGE).

The two 21 mer double-stranded oligonucleotides were purified using 15% polyacrylamide gels in Tris Borate (TB) buffer (Tris HCl 100 mM; borate 83 mM, pH 8.0 at 25 °C). Bands were visualized by UV shadowing. Those corresponding to the size of 21 mers were cut from the gel, eluted overnight in water. Highly water-soluble acrylamide contaminants were removed by fast protein liquid chromatography (FPLC) over a HiTrap Q-HP anion exchange column (GE Healthcare, Life Sciences, UK) using a linear gradient from 100% 10 mM Tris, 20 mM NaCl, pH 7.5 to 100% 10 mM Tris, 800 mM NaCl, pH 7.5 over 30 min at 0.5ml/min. Fractions containing the 21 mers oligonucleotides were pooled. The purified oligonucleotides were desalted on a Sephadex G25 prepacked columns (GE Healthcare, Life Sciences, UK) equilibrated with water. The samples were dried in a Thermo Savant SpeedVac (Thermo Electron Corporation, Waltham, MA) then resuspended in 400 µl of a phosphate buffer containing 10% $^{2}H_2O$, 2 mM EDTA.

NMR Spectroscopy

The NMR spectra were recorded on a Bruker AVANCE™ 500 spectrometer (Bruker Biospin, Germany) using a ^{1}H-BB gradient probe (Bruker Biospin, Germany). NMR samples of labeled DNA were prepared to a final volume of 400 µl at a 0.2 mM concentration.

Results

Design of the primers

The method allows the preparation by PCR amplification of desired ^{15}N, ^{13}C labeled DNA sequences. The starting oligonucleotides consist in two copies of the desired sequence coupled in tandem repeats through a linker of selected size and sequence. The choice of the linker is essential as its incorporation between the desired DNA sequences must create an endonuclease cleavage site at the desired DNA sequence-linker junctions.

Two 21 bp long duplexes were prepared to illustrate the efficiency of the method (Figure 1). The first one termed P site carries a strong cleavage site for topoisomerase II in presence of an ellipticine derivative (Figure 2a) (Fossé et al., 1991). The 3′ end to 5′ end linkage of one P site to another P site fortuitously creates a restriction enzyme recognition site (*Rsa* I GT/AC), but we used a linker to demonstrate the feasibility of the method. This linker was selected through several attempts to determine the best sequence and the minimal size (nine bp) for a *Rsa* I cleavage > 80%. The designed primers were: 5′(ACAGCTTATCATCGAT-

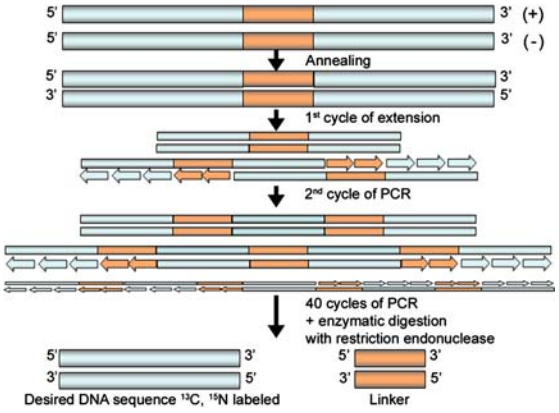

Figure 1. Method outline. The starting oligonucleotide duplex bears two direct repeats of the desired DNA sequence (blue) connected through a selected linker sequence (smaller, orange). The ends of the latter create a restriction endonuclease site with the ends of the desired DNA sequence. The PCR amplification was followed by a restriction endonuclease digestion that released the desired DNA sequence and the linker.

CACGT *ACTTGACGT* ACAGCTTATCATCG ATCACGT)3′, 3′(TGTCGAATAGTAGCTAG TG CA *TGAACTGCA* TGTCGAATAGTAG-CTAGTGCA)5′ (the nucleotides belonging to the *Rsa* I sites are underlined, those belonging to the linker are in italic).

The second target sequence (LTR site) reproduces the U5-LTR extremity of the HIV-1 cDNA (Figure 2b). Several types of restriction sites can be created upon combination of LTR site ends with a selected linker: (i) the last three nucleotides AGT-3′ may form a*Sca* I site (AGT↓ACT) while the last two nucleotides GT-3′ may form a *Rsa* I site (GT↓AC); (ii) the first three nucleotides 5′-GTG may form a *Pml* I site (CAC↓GTG) or more generally a *Bsa*A I site (YAC↓GTR). 5′-ACTNCAC-3′ (with N, sequence of several residues) appeared as a suitable linker for the above four restrictions sites. The size and composition of N leading to the most efficient cleavage was assessed in various test systems, and finally, that is the sequence 5′-ACTATCCAC-3′ (ATC = N) which was selected as linker. By testing the performances of restriction endonucleases in various experimental conditions, there are *Rsa* I and *Bsa*A I which appeared as the most convenient to perform the double digestion of DNA products. Thanks to this choice we diminished the loss of DNA product occurring during DNA purification by alcohol precipitation, in fact after the first incubation performed before redissolution in the appropriated second restriction enzyme buffer. Therefore, designed primers consisted in the linker flanked by two copies of the desired DNA sequence: 5′(GTGTGGAAAATCTC TAGCAGT *ACTATCCAC* GTGTGGAAAATC TCTAGCAGT)3′, 3′(CACACCTTTTAGAG ATCGTCA *TGATAGGTG* CACACCTTTTAG AGATCGTCA)5′ (the nucleotides belonging to the restriction sites are underlined, those belonging to the linker are in italic).

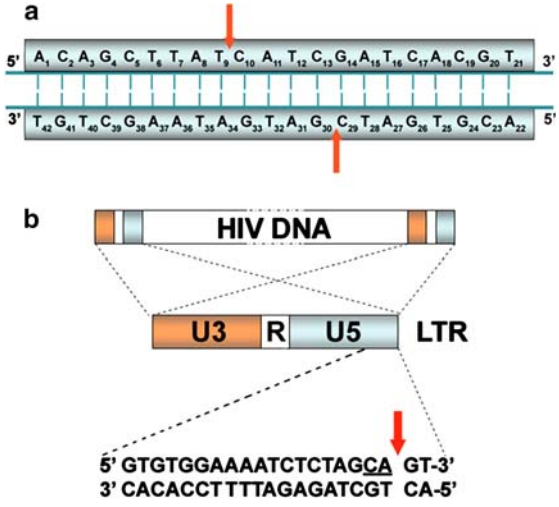

Figure 2. (a) *P site sequence*. This 21 mer DNA derives from the strongest cleavage site (site 22) in pBR322 for the enzyme topoisomerase II in presence of an ellipticine derivative (Fossé et al., 1991). The red arrows indicated the bonds cleaved by topoisomerase II. (b) *LTR site sequence*. This 21 mer DNA (bottom of the Figure) reproduces the U5-LTR (long terminal repeat) extremity of the HIV-1 cDNA. A scheme of the HIV-1 viral cDNA is given at the top of the Figure LTRs at both ends of the viral cDNA consist of three consecutive elements, U3–R–U5, repeated in the same orientation (middle of the Figure). 3′-processing catalyses the resection (red vertical arrows) of the viral DNA ends immediately 3′ from the conserved CA dinucleotide (underlined), thereby generating two dinucleotides (5′-GT) (one from each viral 3′-end) and reactive 3′-hydroxyl DNA ends.

PCR amplification and restriction enzyme digestion

The DNA synthesis is based on PCR amplification of the primers in the presence of ^{13}C, ^{15}N-labeled dNTPs, followed by a treatment of DNA products with *ad hoc* restriction endonucleases that release the desired DNA fragments (Figure 1).

An improvement of the PCR elongation step can be gained from the replacement of the commonly used Taq DNA polymerase or Pfu DNA polymerase by a blend of both. Amplification of DNA fragments larger than 1 kb by Taq DNA polymerase is problematic for two reasons: (i) the lack of proofreading capacity of the enzyme that produces misincorporations reducing the efficiency of amplification and introduces mutations; (ii) the single base 3′ overhang that may occur in the extension products, which is not efficiently removed by the 3′-5′ exonuclease activity (Clark, 1988). On the other hand, the use of the proofreading thermostable Pfu DNA polymerase, avoids misincorporated nucleotides and 3′ overhangs (Cheng et al., 1994), but the yield of the amplification product typically drops proportionally to the increase of DNA length. Addition of a small quantity of Pfu DNA polymerase to Taq DNA polymerase has been shown to significantly improve the PCR amplification of DNA fragments up to 35 kb (Cheng et al., 1994). An efficient DNA polymerase blend is obtained by combining

2.5 units of Taq DNA polymerase and 0.2 units of Pfu DNA polymerase for a 50 µl reaction. In such conditions, PCR amplification provides a better yield and a higher specificity. Products are collected in sufficient amounts to allow downstream applications.

Note that for each template, annealing and extension temperatures were optimized in one experiment by using thermal cycler gradient block module and temperature control software. We checked that the annealing temperature was optimal with the gradient function of our thermal cycler, which enables us to obtain a different temperature for each column of the 96 well heating block (12 columns). Each DNA sample was then checked for the size and dosed at 260 nm to estimate the quantity. The Thermo Hybaid PCR Express thermal cycler has a software function ''time/temperature – increments/decrements'', these features enable the time interval and/or the temperature of a specified programmed step to be increased or decreased with successive temperature cycle. The use of the time increment function is recommended for reactions with a great number of cycles to increment the extension time interval to compensate for deterioration of enzyme activity in later cycles. So, elongation time for long templates (in step 2) was optimized using the time increment function of the thermal cycler. Fifty cycles of annealing and extension resulted in hundreds of tandem copies of the target and the linker duplex (Figure 3a).

Optimization of purification procedure

In our first assays the PCR products were immediately submitted to restriction endonuclease digestion. We noticed that the presence of impurities and salts altered the enzyme efficiency. PCR products were therefore desalted by elution on Sephadex G25 prior to restriction digest. Moreover, those products obtained with the blend of Taq/Pfu DNA polymerases (Figure 3c) permitted a better digestion by restriction endonucleases compared with those obtained with the Taq DNA polymerase alone (Figure 3b).

The desired P site (21 mer) was separated from the linker (9 mer) and the longer products (30 and 39 mers) as well as other components through electrophoresis on polyacrylamide gel (PAGE) (Figure 3b and c). The undigested material migrating as 30 and 39 mers is due to the 3′ overhang that obliterates the *Rsa* I recognition site

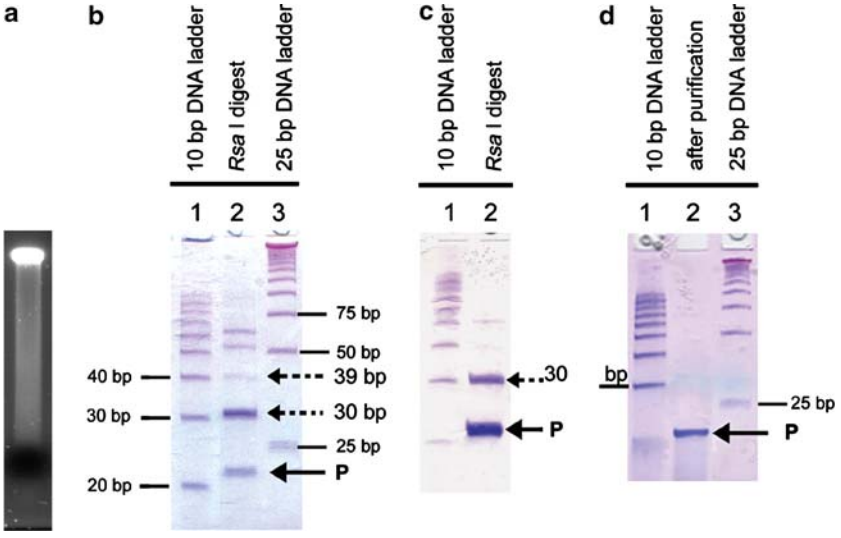

Figure 3. *Production of labeled P site*: (a) 0.8% agarose gel electrophoresis of 5 µl of reaction mixture, after 40 cycles of PCR amplification. (b) 15% polyacrylamide gel electrophoresis of a 10 bp double-stranded marker (lane 1), a 25 bp double-stranded marker (lane 3), 10 µl from the *Rsa* I digests after amplification with the Taq DNA polymerase alone (lane 2). The linker (9 bp) was running out of the gel. (c) 15% polyacrylamide gel electrophoresis of a 10 bp double-stranded marker (lane 1), and 10 µl from the *Rsa I* digests after PCR amplification with the blend Taq/Pfu (2.5 U:0.2 U) DNA polymerases (lane 2). The linker (9 bp) was running out of the gel. (d) 15% polyacrylamide gel electrophoresis of a 10 bp double-stranded marker (lane 1), a 25 bp double-stranded marker (lane 3), and 2 µl of the purified target DNA sequence (lane 2).

and results in oligonucleotides that cannot be cleaved to monomers (Louis et al., 1998). However, the use of polymerases blend dramatically decreases the amount of undigested material (Figure 3c).

The products were separated from each other by preparative PAGE using a Tris Borate (TB) buffer. The desired DNA sequence was recovered from the gel through elution in water without "crush and soak" to reduce the amount of polyacrylamide fragments swept along into extracts. Typically 90% of the desired DNA sequence was recovered by this method. Yet, the method left highly water-soluble acrylamide oligomers mixed to DNA products in the samples. To clear them from such contaminants, that give rise to strong proton signals in the aromatic region of NMR spectra, these were eluted on a HiTrap Q-HP anion exchange column and then desalted on Sephadex G25. Solutions were dried on SpeedVac and resuspended in a phosphate buffer containing 10% $^{2}H_2O$, 2 mM EDTA. The procedure provided sufficient material to prepare a 0.2 mM NMR purified sample (Figure 3d) in 400 µl solution.

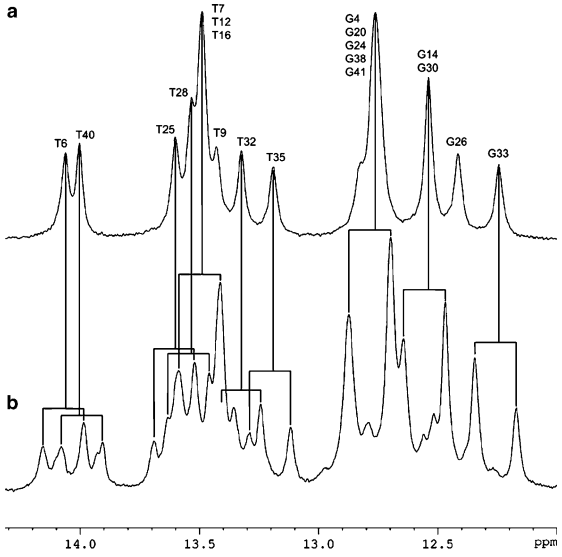

Figure 4. The imino protons 1D-NMR region of the P site duplex in 90% H_2O/10% $^{2}H_2O$, unlabeled (a) and uniformly ^{13}C, ^{15}N labeled (>99%) (b). Both spectra were acquired at 20°C using the Watergate pulse sequence as to suppress water resonance (Piotto et al., 1992). Comparison attests for the fidelity of the PCR amplification. Observation of ^{1}J ^{1}H–^{15}N of ~ -90 Hz along the spectrum illustrates the uniform ^{15}N labeling.

NMR experiments

NMR spectra permit to check the quality, quantity, integrity, and purity of the DNA products and therefore to validate the method. The imino protons region of DNAs is favorable to such an analysis.

An in-depth check of products necessitates the assignments of NMR signals associated to enriched atoms. 1D NMR spectra of the imino protons region of unlabeled and labeled P site are shown in Figure 4a and b, respectively. Assignments are indicated in the spectrum of the unlabeled compound (Figure 4a). The spectrum of the labeled compounds (Figure 4b) is enriched by the appearance of ^{1}H–^{15}N splittings. Protons of the labeled and unlabeled imino groups display same chemical shifts indicating that the compounds have the same nucleotide composition and sequence. In our experimental conditions no extra peaks are visible. 2D ^{1}H–^{15}N HSQC-sensitivity enhanced spectra of the labeled P site were obtained with excellent signal/to noise ratio after 3 h at 500 MHz, using a classical sample volume of 400 µl. The corresponding imino region is presented in Figure 5. Here also the chemical shifts are consistent with the expected P site sequence.

The $^{2}J_{NN}$ COSY experiment presented in Figure 6 records the ^{15}N–^{15}N coupling constants occurring in base pairs through hydrogen bondings. Detection of such couplings (3–7 Hz) that rely on weak spin–spin interactions of ^{15}N atoms is highly demanding. Values are proportional to the strength of the hydrogen bond insuring the base pairing. Yet, all the expected $^{2}J_{NN}$ couplings were measured, without exception, reflecting the stability of the P site duplex in our experimental conditions.

To further assess the efficiency of our method, we recorded also a ^{1}H, ^{13}C-HMQC spectrum of the labeled P site. The C2/C6/C8-H2/H6/H8 region is presented in Figure 7. The chemical shifts are the same as measured in the NOESY spectrum of the unlabeled P site (data not shown). The only visible difference concerns the H8-C8 cross peak of the 5′ terminal residues A_1 and A_{22} that appears at proton chemical shift of 8.47 ppm in the spectrum of labeled P site and 8.30 ppm in the spectrum of the unlabeled version. This is imputed to the phos-

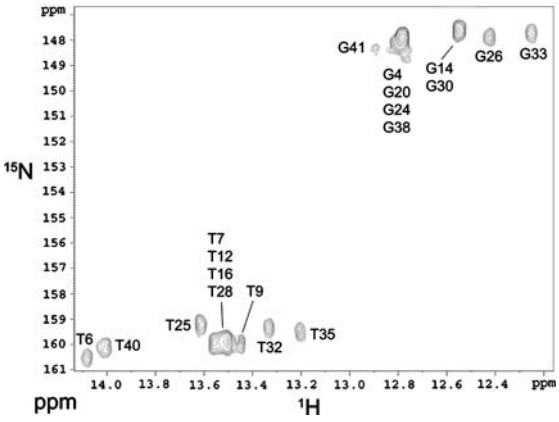

Figure 5. *Fingerprint* ^{1}H–^{15}N *HSQC regions* for uniformly ^{13}C, ^{15}N labeled P site duplex recorded at 20 °C on a 0.2 mM sample. Assignments of imino protons were achieved on a NOESY spectrum of unlabeled P site (not shown). Recording required 3 h with 200*(t_1) × 1024*(t_2) complex points and acquisition times of 79 ms and 51 ms in the ^{15}N and ^{1}H dimensions, respectively, and the ^{15}N carrier at 50 ppm. Nucleotides cluster together by residue-type according to their N1 (guanine) or N3 (thymine) chemical shift.

phoryl group introduced by the PCR method at the 5′ ends of oligonucleotides, while these remain free in the chemically prepared oligonucleotides.

It can be also noticed that in the ^{13}C dimension of the ^{1}H, ^{13}C-HMQC spectrum, the signals corresponding to pyrimidines are broader than those of purines. This broadening accounts for an additional ^{13}C–^{13}C coupling ($^{1}J_{CC}$) of about 67 Hz between the adjacent C5 and C6 atoms of the pyrimidine $\overset{\backprime}{C}_5{=}C_6{-}H$ motif (Wijmenga and van Buuren, 1998). In the thymidine motifs, the C6 signal displays an extra ^{13}C broadening, likely resulting from an additional ^{13}C–^{13}C coupling ($^{2}J_{CC}$) with the methyl groups carried by the C5 atom.

Discussion

Our primary goal was to develop a general method for preparation of uniformly labeled oligonucleotides useful for NMR studies. The method is based on the polymerization of labeled dNTPs by PCR amplification. The starting DNA consists in two copies of the desired DNA sequence connected through a linker of selected size and sequence. The choice of the linker is of importance as it helps to create well-defined restriction sites with the ends of

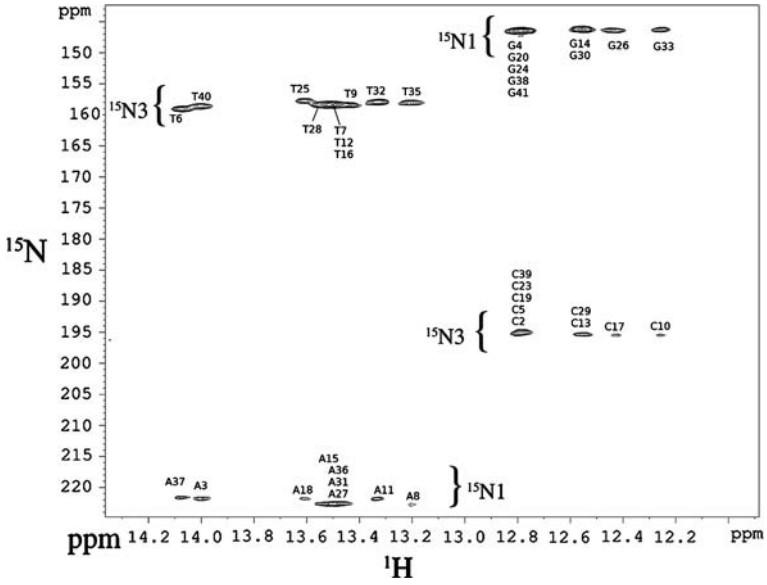

Figure 6. *Fingerprint quantitative* J_{NN} *HNN-COSY region* (Dingley and Grzesiek, 1998) for the uniformly (>99%) ^{13}C, ^{15}N labeled P site duplex. Positive contours (δ ^{15}N 145–160 ppm) correspond to "diagonal peaks" resulting from ^{1}J $^{1}H3$–$^{15}N3$ ($^{1}H1$–$^{15}N1$) in thymine (guanine), and negative contours (δ ^{15}N 190–225 ppm) correspond to cross-peaks resulting from interstrand scalar $^{15}N3$–$^{15}N1$ ($^{15}N1$–$^{15}N3$) magnetization transfer between thymine and adenine (guanine and cytosine), inside a base pair. Resonances are labeled with currently available assignment information. Spectra required 31 hours of measurement, with 100*(t_1) × 2048*(t_2) complex points, acquisition times of 19.6 and 102 ms in the ^{15}N and ^{1}H dimensions, respectively, and the ^{15}N carrier at 180 ppm. Assignments of residues are indicated in the spectrum.

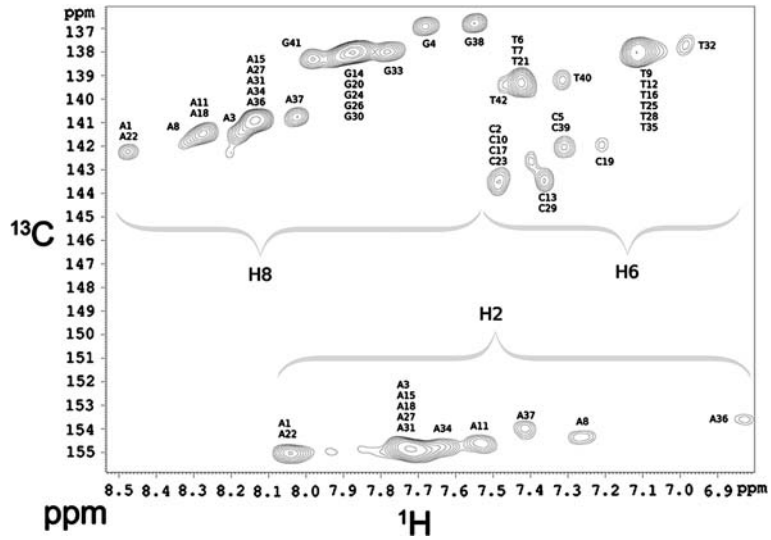

Figure 7. Selected region of the ^{1}H–^{13}C HMQC spectrum of the P site duplex recorded at 30 °C. The region shows the correlations between H2/H6/H8 and C2/C6/C8 of the aromatic bases. The spectrum required 12 h of measurement, with data set of 200 real × 2048 complex points, acquisition time of 20 and 205 ms in the ^{13}C and ^{1}H dimensions, respectively. One hundred and twenty-eight scans per transient were performed, and spectral widths were set to 10 and 80 ppm in ^{1}H and ^{13}C dimensions, respectively.

the desired DNA sequence. Such setting needs to take into account the binding capacities of restriction end.

The proposed method differs from the ESRA one reported by Louis et al. (1998) in which the desired DNA sequences are linked together in tandem repeats through 3′ 5′ connections. The restriction endonuclease site is therefore created at the junction of the desired DNA sequences from residues that belong to both the 5′ and 3′ extremities. In that case the number of possible combinations to assemble a restriction site is very low. For instance for a hexanucleotide recognition sequence, the above method allows only 40 combinations of two triplets on the 4096 possible ones (Tables 1–3), thus strongly limiting the choice of the oligonucleotides which can be prepared, including those biologically relevant. The LTR site (Figure 2b) reproducing the U5-LTR 3′ extremity of the HIV-1 cDNA and that contains the integrase attachment and processing sites provides a good illustration (Zargarian et al., 2003; Renisio et al., 2005). Any alteration within the six outer most residues either by addition, substitution or deletion of residues abolishes the recognition event (Zargarian et al., 2003), while no known restriction site can be created through 3′ 5′ coupling of the LTR sites. This makes obligatory the use of a linker sequence to generate a restriction site.

The sequence of the linker is selected to create an endonucleolytic cleavage site at each of its two extremities, which is exactly at the linker-desired DNA junctions (Figure 1). We have seen that in that case the triplet combinations corresponding to the hexanucleotide recognition sequence (Tables 1–3) leads to 1600 combinations on 4096, that is 40% of all the possible combinations (when the ESRA method allowed only 1%). If a recognition sequence cannot be obtained with a hexanucleotide linker, the use of a heptanucleotide is generally sufficient to solve the problem (Tables 2 and 3). Globally, the size of the linker is of importance. As a general rule efficient cleavage requires at least six base pairs surrounding the cleavage site on both sides and a maximum that must be greatly smaller than the size of the DNA target sequence (to permit a better separation and purification of products). Moreover, the linker must have a minimal tendency to form internal loop.

In conclusion, we propose a general PCR-based method for the production of ^{13}C, ^{15}N labeled DNA duplexes in sufficient amounts to carry out NMR studies. NMR studies using labeled DNAs are still very rare due to pitfalls and difficulties linked to current synthesis methodologies. While the increase of the number and variety of NMR parameters (including residual dipolar

Table 1. Palindromic tetra and Hexa-nucleotide recognition Sequences

AATT	ACGT	AGCT	ATAT	CATG	CCGG	CGCG	CTAG	GATC	GCGC	GGCC	GTAC	TATA	TCGA	TGCA	TTAA
□□↓□□		*Alu* I, **CviJ I**				*Bst*U I		*Dpn* I		*Hae* III, *Pho* I, **CviJ** I	*Rsa* I			*Hpy* CH4 V	
A □ □ ↓ □ □ T			*Ssp* I						*Afe* I	*Stu* I	*Sca* I			*Bfr*B I	
C □ □ ↓ □ □ G		*Pml* I, ***Bsa*A I**	*Pvu*II, ***Msp*A1 I**		*Sma*I	*Msp*A**1 I**									*Hpa* II, ***Hinc* II**
G□ □ ↓ □ □C		*Zra* I	*Ecl* 136 II	*Eco*R V		*Nae* I			*Sfo* I			*Bst*Z17 I	***Hinc* II**		
T □ □ ↓ □ □ A		*Sna*B I, ***Bsa*A I**							*Fsp* I	*Msc* I		*Psi* I			*Dra* I

Sequences at the top of each column are written 5′ to 3′ according to convention. Open squares at the left of each row are place holders for nucleotide within a restriction endonuclease recognition sequence; arrows indicate the point of cleavage. Sequences of complementary strands and their cleavage sites are implied. Enzymes written in normal characters recognize only one sequence while enzymes written in bold characters recognize multiple sequences.

Table 2. Specificities greater than 6 bases

GTTT↓AAAC	*Pme* I
GCCC↓GGGC	*Sfr* I
ATTT↓AAAT	*Swa* I

Arrows indicate the point of cleavage. Sequences of complementary strands and their cleavage sites are implied.

Table 3. Interrupted palindromes

CACNN↓NNGTG	*Ahd* I
GATNN↓NNATC	*Bsa*B I
GCN↓NGC	*Cac*8 I
GTN↓NAC	*Hpy*8 I
GGN↓NCC	*Nla* IV
GAANN↓NNTTC	*Xmn* I

N = A or C or G or T. Arrows indicate the point of cleavage. Sequences of complementary strands and their cleavage sites are implied.

couplings) provided by isotope labeling will help us to unravel the structures of longer DNA and the mechanisms governing the DNA–protein recognition. We show that insertion between desired DNA sequences of a linker that can be removed by enzymatic digestion increases considerably the potential of the PCR approach. The ability to choose the size and the sequence of the linker offers a maximum flexibility and there is not anymore constraint for the choice of the DNA duplexes to be prepared. The use of a blend of Taq and Pfu (2.5:0.2) polymerases constitutes another improvement. Obviously, when only one labeled strand is desired in order to study hairpins and duplexes gathering labeled and unlabeled strands it is better to resort to the method of Zimmer and Crothers (1995). Moreover, site-specific labeling can be obtained through chemical synthesis (Ono et al., 1995; Fernandez et al., 1998; Tjandra et al., 2000; Kojima et al., 2001).

Acknowledgement

This work was financially supported by a grant from the French Ministry of Research (to G.M.).

References

Batey, R.T., Battiste, J.L., Williamson, J.R. (1995) *Methods Enzymol.*, **261**, 300–322.

Batey, R.T., Inada, M. and Kujawinski, E. et al. (1992) *Nucleic Acids Res*., **17**, 4515–4523.
Chen, X., Mariappan, S.V. and Kelley, J.J. III et al. (1998) *FEBS Lett*., **3**, 372–376.
Cheng, S., Chang, S.Y. and Gravitt, P. et al. (1994) *Nature*, **6482**, 684–685.
Clark, J.M. (1988) *Nucl. Acids Res*., **20**, 9677–9686.
Clore, G.M. and Gronenborn, A.M. (1994a) *Methods Enzymol*., **239**, 349–363.
Clore, G.M. and Gronenborn, A.M. (1994b) *Prog. Biophys. Mol. Biol*., **2**, 153–184.
Dingley, A.J. and Grzesiek, S. (1998) *J. Am. Chem. Soc*., **33**, 8293–8297.
Ferentz, A.E. and Wagner, G. (2000) *Q. Rev. Biophys*., **1**, 29–65.
Fernandez, C., Szyperski, T. and Ono, A. et al. (1998) *J. Biomol. NMR*, **1**, 25–37.
Fesik, S.W. and Zuiderweg, E.R. (1990) *Q. Rev. Biophys*., **2**, 97–131.
Fossé, P., René, B. and Le Bret, M. et al. (1991) *Nucl. Acids Res*., **11**, 2861–2868.
Hines, J.V., Landry, S.M. and Varani, G. et al. (1994) *J. Am. Chem. Soc*., **13**, 5823–5831.
Katzman, M. and Katz, R.A. (1999) *Adv. Virus Res*., **52**, 371–395.
Kojima, C., Ono, A. and Ono, A. et al. (2001) *Methods Enzymol*., **338**, 261–283.
Louis, J.M., Martin, R.G. and Clore, G.M. et al. (1998) *J. Biol. Chem*., **4**, 2374–2378.
Masse, J.E., Bortmann, P. and Dieckmann, T. et al. (1998) *Nucl. Acids Res*., **11**, 2618–2624.
Michnicka, M.J., Harper, J.W. and King, G.C. (1993) *Biochemistry*, **2**, 395–400.
Nikonowicz, E.P. and Pardi, A. (1992) *Nature*, **6356**, 184–186.
Ono, A., Tate, S. and Kainosho, M. (1995) *Tanpakushitsu Kakusan Koso*, **10**, 1509–1517.
Piotto, M., Saudek, V. and Sklenar, V. (1992) *J. Biomol. NMR*, **6**, 661–665.
Renisio, J.G., Cosquer, S. and Cherrak, I. et al. (2005) *Nucl. Acids Res*., **6**, 1970–1981.
Smith, D.E., Su, J.Y. and Jucker, F.M. (1997) *J. Biomol. NMR*, **3**, 245–253.
Tjandra, N., Marquardt, J. and Clore, G.M. (2000) *J. Magn. Reson*, **2**, 393–396.
Varani, G., Aboul-ela, F. and Allain, F.H.T. (1996) *Progress Nucl. Mag. Res. Spectr*., **1–2**, 51–127.
Werner, M.H., Gupta, V. and Lambert, L.J. et al. (2001) *Methods Enzymol*., **338**, 283–304.
Wijmenga, S.S. and van Buuren, B.N.M. (1998) *Progress Nucl. Mag. Res. Spectr*., **4**, 287–387.
Zargarian, L., Benleumi, M.S. and Renisio, J.G. et al. (2003) *J. Biol. Chem*., **22**, 19966–19973.
Zimmer, D.P. and Crothers, D.M. (1995) *Proc. Natl. Acad Sci. USA*, **8**, 3091–3095.