Gene, 44 (1986) 177–183
Elsevier

GENE 1662

Cloning of random-sequence oligodeoxynucleotides

(Recombinant DNA; synthetic linkers; consensus sequences; genetic element; mutagenesis; *E. coli* DNA polymerase I; restriction enzymes; M13 phage vector)

Arnold R Oliphant, Alexander L. Nussbaum and Kevin Struhl *

Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115 (U.S.A.) Tel. (617)732-2104

(Received April 30th, 1986)
(Accepted May 8th, 1986)

SUMMARY

Methods are described for cloning random or highly degenerate nucleotide (nt) sequences. The procedures use synthetically derived mixtures of oligodeoxynucleotides (oligos) whose heterogeneous central portions are bounded at their 5′ and 3′ ends by sequences recognized by restriction endonucleases. Oligo collections of defined length and nt composition are synthesized by utilizing appropriate concentrations of all four nucleotide precursors during each addition step for the central region. Single-stranded oligos with appropriate 5′ and 3′ ends can be ligated directly, although inefficiently, into double-stranded (ds) DNA molecules with complementary 5′ and 3′ extensions produced by restriction endonuclease cleavage. A more general and efficient method is to convert the oligo into a ds form by incubating it with the Klenow (large) fragment of *Escherichia coli* DNA polymerase I. If the 3′ ends are palindromic, two oligo molecules will serve as mutual primers for polymerization. The resulting products are ds molecules containing two oligo units separated by the original 3′ restriction site and bounded at each end by the original 5′ restriction site. After appropriate restriction endonuclease cleavage, oligo units can be cloned by standard procedures. Analysis of 26 recombinant M13 phages indicates that the nt sequences of the cloned oligos are in good accord with what was expected on a random basis.

INTRODUCTION

We have been interested in determining the nt sequence requirements of various genetic elements. One obvious approach, saturation mutagenesis, is simply to alter each bp within a region known to be required for a particular function, and to analyze the phenotypic effects. As an alternative approach, we have developed a 'random selection' method. The basic idea is to generate a collection of recombinant DNA molecules in which random or highly degenerate nt sequences replace a genetic element of interest,

* To whom correspondence and reprint requests should be addressed.

Abbreviations: bp, base pair(s); dNTP, any of the four deoxynucleoside triphosphates; ds, double stranded; DTT, dithiothreitol; HPLC, high-performance liquid chromotography; nt, nucleotide(s); oligo, oligodeoxynucleotide; PA, polyacrylamide; PolI, *E. coli* DNA polymerase I, PolIk, Klenow (large) fragment of PolI; ss, single stranded; 3X buffer, see MATERIALS AND METHODS, section **b**.

and then to select by genetic or biochemical means particular sequences that confer an equivalent function. A comparison of nt sequences that satisfy a particular selection should result in a consensus that defines the genetic element (Oliphant and Struhl, 1986).

Although random nt sequences can be generated by utilizing appropriate mixtures of nt precursors at each step of the chemical synthesis (Zon et al., 1985), methods for cloning such oligos have not been described. Current methodologies for cloning oligos can be divided into two basic classes (reviewed by Zoller and Smith, 1982). In one approach, the oligo is hybridized to an ss template (usually in the form of an M13-hybrid molecule) and extended with PolI. The product of this reaction is a heteroduplex, which, upon repair and/or replication in *E. coli*, can be converted into a homoduplex containing the mutation. In the other approach, two complementary oligos with cohesive ends are synthesized, and the hybridized product is ligated into ds DNA that has been cleaved with appropriate restriction endonucleases. However, these standard methods are unsuitable for cloning random DNA because the extreme heterogeneity of the oligo mixture precludes the availability of a complementary template or oligo. To solve this problem, we describe procedures for (i) direct cloning of ss oligos into ds DNA molecules, and (ii) converting oligos into ds DNA that can be easily inserted into standard vectors.

MATERIALS AND METHODS

(a) Oligodeoxynucleotide synthesis

Oligos were synthesized on an Applied Biosystems DNA synthesizer (Model 380A) using the phosphite triester method (Matteucci and Caruthers, 1981; Beaucage and Caruthers, 1981). In some cases, especially those involving long regions of random DNA sequence, the procedure was modified by omitting the capping reaction after each step of random addition. This modification improves the yield considerably because oligos that fail to react at a given step are not irreversibly inactivated and thus can react at subsequent steps; it also results in oligos that are more heterogeneous in length as compared to conventional syntheses. In performing this modification, it was essential to carry out a mock reaction without the capping reagent for the normal amount of time. Simple elimination of the capping step resulted in oligos that were much shorter than expected.

After detachment and removal of all but the 5′-dimethoxytriphenylmethyl protecting groups, the oligos were separated from shorter congeners by HPLC on a Waters C-8 column using a 40-min linear gradient of 0.1 M triethylammonium bicarbonate, pH 7.1, from 0–25% acetonitrile. The peak containing the trityl chromogen (emerging near the top of the gradient) was desalted by flash evaporation in vacuo at temperatures below 30°C and completely deprotected by treatment with 80% aqueous acetic acid at room temperature for 20 min, followed by flash evaporation as above.

(b) Conversion of ss oligos to ds molecules

The oligo (1 to 5 μg) was diluted into a total volume of 10 μl of 3X buffer (30 mM Tris pH 7.5, 150 mM NaCl, 30 mM $MgCl_2$, 15 mM DTT, 0.1 mg/ml gelatin), hybridized for at least one hour at 37°C, and then allowed to cool slowly to room temperature. dNTPs (to a final concentration of 250 μM for each of the four) and [^{32}P]dATP (10 μCi) were then added, and the reaction mixture was diluted to a final volume of 30 μl. PolIk (5 units) was added, and after incubation at 37°C for at least one hour, the reaction products were extracted with phenol and precipitated with ethanol. The resuspended DNA was cleaved to completion with the restriction endonuclease recognizing the outside sites (originally the 5′ site), extracted with phenol, and concentrated by ethanol precipitation. After electrophoresis in a 12% PA gel, the desired ds molecules containing two oligo units were eluted in 0.5 M ammonium acetate, 1 mM EDTA for 4–24 h at 25°C, and then concentrated by ethanol precipitation. The purified DNA was cleaved with the restriction endonuclease recognizing the central site (the original 3′ site) to produce the final product, a ds version of the oligo mixture with 5′ and 3′ ends suitable for ligation into standard vector molecules. Standard ligation reactions using T4 DNA ligase were carried out at 15°C in 20 μl reactions containing 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 10 mM

DTT, 500 μM ATP. As the yield was somewhat variable, the amount of oligo to be added to a given amount of vector was determined empirically to optimize the ligation reaction. It should be noted that the 5′ end does not have to be cleavable by a restriction endonuclease because mutually primed synthesis produces blunt ends that are suitable for cloning. Indeed, even in situations involving enzymatic cleavage at the 5′ end, it is advantageous to minimize the length of the palindrome to disfavor hybridization of the 5′ ends that might block complete extension.

RESULTS

(a) Ligation of ss oligos into ds vectors

In principle, ss oligos can be ligated directly into ds vector DNAs if the former contain 5′ and 3′ ends with nt complementary to the 5′ and 3′ ss extensions produced by appropriate restriction endonuclease cleavage of the vector DNA. At both the 5′ and 3′ ends, the ds vector DNA and the ss oligo would anneal in a manner similar to that used for standard cloning experiments. The desired ligation product would be a ds molecule containing a gap corresponding to the central region of the oligo (Fig. 1b); presumably this gap would be correctly repaired upon introduction into *E. coli* cells.

To test this possibility, we synthesized an oligo containing 5 nt of the *Eco*RI site (AATTC) on the 5′ end, 14 nt of random sequence in the center, and 5 nt of the *Sac*I site (GAGCT) on the 3′ end (Fig. 1a). Although only the central 4 nt of each restriction site are necessary for the annealing of cohesive ends, inclusion of the extra nt allows for the regeneration of the site. The oligo was phosphorylated at its 5′-end with T4 polynucleotide kinase and ligated to M13mp18 DNA (Messing, 1983) that was previously cleaved with *Eco*RI + *Sac*I. The resulting products (Fig. 1b) were introduced into *E. coli* strain JM101 and analyzed by the 'blue-white' screen (Messing, 1983). Successful cloning of the oligo will produce white plaques in contrast to the M13mp18 vector which produces blue plaques.

In several experiments, standard ligation conditions (15 °C) using 0.1 μg of M13mp18 vector DNA

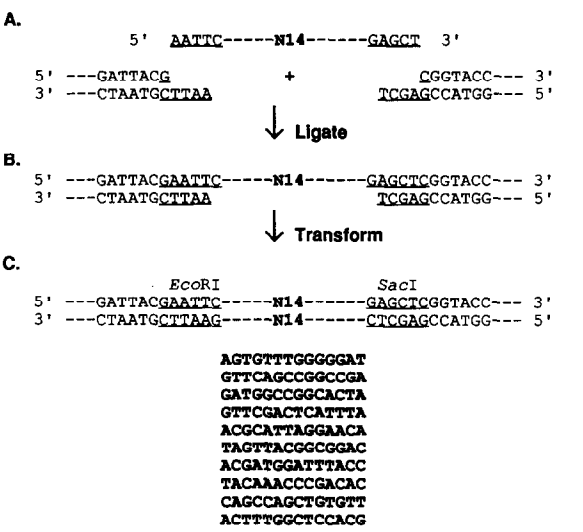

Fig. 1. Direct ligation of single-stranded oligodeoxynucleotides. The nt sequences near the restriction endonuclease cleavage sites of the following molecules are shown: **A**, ss oligo containing 14 nt of an equimolar mixture of all four nt precursors (N) bounded at the 5′ end by 5 nt of the *Eco*RI site and at the 3′ end by 5 nt of the *Sac*I site (top line) to be ligated with *Eco*RI + *Sac*I cleaved M13mp18 ds DNA (bottom line); **B**, the gapped molecule resulting from the ligation; **C**, the top (5′–3′) strand of ten transformants containing a cloned oligo. The nt corresponding to restriction endonuclease recognition sites are underlined. See RESULTS, section **a**.

and 3 ng of phosphorylated oligo (a 15-fold molar excess) produced approx. 25 white plaques. Sequence analysis (Sanger et al., 1980) of ss DNA from 12 such plaques revealed that ten contained an oligo that was inserted between the regenerated *Eco*RI and *Sac*I sites (Fig. 1c). The remaining two plaques represented aberrant joining of the vector *Eco*RI and *Sac*I sites. As expected, ss ligation requires phosphorylation of the oligo. When the oligo was not phosphorylated, four white plaques were observed, all of which were due to aberrant joining.

Although these results indicate that ss oligos can be cloned directly, the frequency of the desired products is low. Parallel ligations using ds insert DNAs generally produced about 100-fold more white plaques. Various attempts were made to increase the frequency of ss ligation. However, uniformly low frequencies were observed over a wide range of oligo concentrations (up to 10000-fold molar excess) and over a wide range of temperatures (0 to 37 °C). Low temperatures caused minor reductions in the ef-

ficiency of the reaction. Although the reason for the inefficiency of this ligation reaction is unclear, it can not be due simply to the annealing because identical cohesive ends are involved. One possibility is that an optimal substrate for T4 DNA ligase requires an adequate stretch of duplex DNA on both sides of the nick. In the reaction described here, one side of the nick contains only 4 bp before the gap caused by the oligo. In contrast, the standard joining of cohesive ends generates annealed DNA segments that are completely duplex except for the staggered nicks.

Thus, ss oligos can be ligated directly in standard ds DNA molecules. The procedure is technically simple, and it permits the directed insertion of a single oligo; multiple insertion events are impossible. However, although this method is useful for constructing specific DNA molecules, the low efficiency of ligation makes it difficult to generate large numbers of molecules each containing a single oligo insertion. In addition, the method is restricted to target DNAs that can be cleaved with pairs of enzymes that generate both 5′ and 3′ ss extensions.

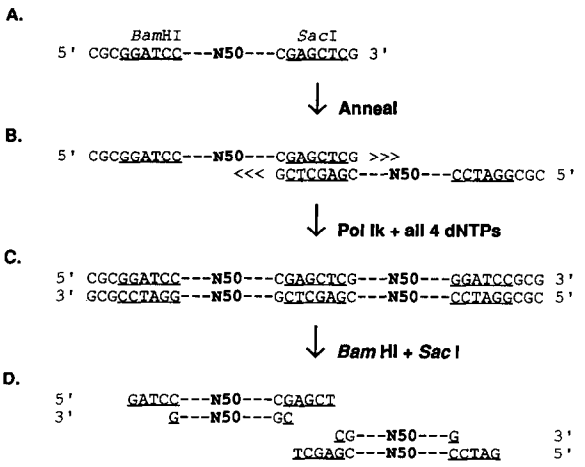

Fig. 2. Production of random DNA. The structures of the following molecules are shown: **A**, ss 50-mer of an equimolar mixture of all four nt precursors bounded by *Bam*HI (5′ end) and *Sac*I sites (3′ end); **B**, hybrid of two oligos annealed by virtue of their self-complementary 3′ ends; **C**, ds DNA resulting from mutually primed synthesis; **D**, pair of ds DNAs suitable for cloning into *Bam*HI + *Sac*I cleaved M13mp19 DNA. The procedural details are discussed in MATERIALS AND METHODS, section **b**. The nt within the restriction sites are underlined.

(b) Cloning random-sequence oligos by mutually primed synthesis

A more general and efficient method of cloning random-sequence oligos is to convert them to the ds form by using a palindromic restriction site at the 3′ end (Fig. 2). In this way, two oligo molecules should anneal such that they serve as mutual primers for extension with PolIk. The product of this reaction should be a ds molecule containing two oligonucleotide units that are separated by the original 3′ restriction site and are flanked by the original 5′ restriction sites.

This idea was tested by synthesizing an oligo containing 50 nt of random sequence activated by *Bam*HI and *Sac*I sites (Fig. 2a). One additional nt was added on each side of the *Sac*I site to create an 8-bp palindrome which would facilitate hybridization. Three nt were added 5′ of the *Bam*HI site to ensure greater cleavage efficiency. After allowing the oligo mixture to self-hybridize (Fig. 2b), the products were treated with PolIk in the presence of high concentrations of all four dNTP and [α-^{32}P]dATP to trace the products of the reaction. As predicted (Fig. 2c), the major product of this reaction was 126 bp in length, and subsequent treatment with *Bam*HI, *Sac*I, and *Bam*HI + *Sac*I generated bands corresponding respectively to 114, 61, and 53 nt (Fig. 3). For cloning purposes, after *Bam*HI cleavage of the ds oligo, the relevant product was purified on a PA gel in order to remove all remaining ss molecules. This material was then cleaved with *Sac*I to produce ds oligos with cohesive ends suitable for standard ligation reactions. Varying amounts of the final product (from 0.1 to 100 ng) were combined with 0.1 μg of *Bam*HI + *Sac*I-cleaved M13mp19 vector DNA, treated with T4 DNA ligase in a standard reaction, and analyzed after transfection of the JM101 host.

The cloning efficiency strongly depends on the amount of oligo present in the ligation reaction. Optimal conditions were achieved at a 3–10 fold molar excess of oligo over the vector. At lower concentrations, the number of plaques was roughly proportional to the amount of oligo. At higher concentrations, the number of plaques decreased.

Sequence analysis indicates that as expected, all of the phages contained a single oligo that was inserted between the vector *Bam*HI and *Sac*I sites. In other experiments, phages containing three oligo insertions

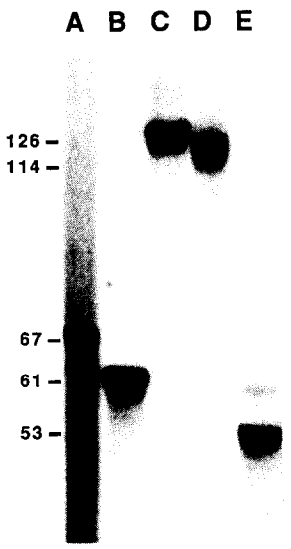

Fig. 3. Analysis of the mutually primed synthesis reaction. DNAs were electrophoretically separated in a 10% PA–7 M urea gel (Sanger et al., 1980). Lane **A** corresponds to the untreated oligo of Fig. 2a (labeled at its 5′ end with T4 polynucleotide kinase). Lanes **B–E** correspond to the products of the mutually primed synthesis (using [α-^{32}P]dATP in the reaction) that have been cleaved with *Sac*I (lane **B**), uncleaved (lane **C**), cleaved with *Bam*HI (lane **D**), and cleaved with *Sac*I + *Bam*HI (lane **E**). See RESULTS, section **b**.

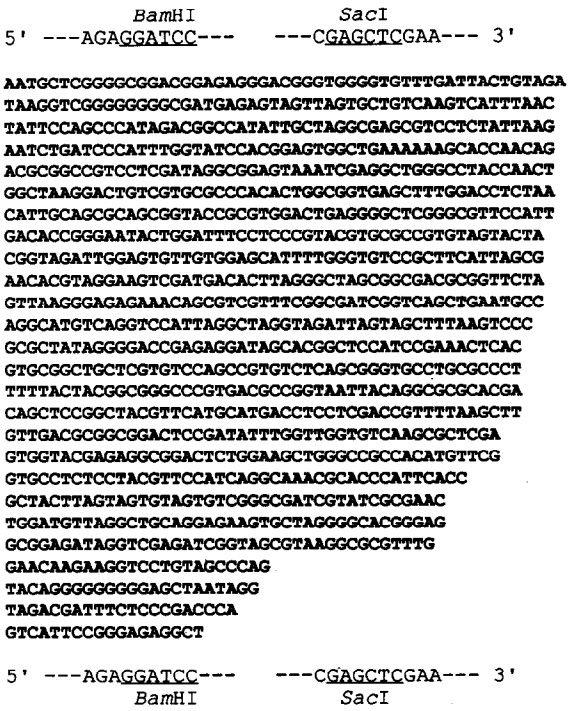

Fig. 4. Sequences of cloned random DNA. The sequences of 26 random insertions into *Bam*HI + *Sac*-cleaved M13mp19 DNA were determined as described by Sanger et al. (1980). Only the strand corresponding to the synthesized oligo is shown.

were occasionally observed especially when the plaques were derived from reactions containing a large excess of oligo. The nt sequences of the 26 insertions are shown in Fig. 4. The variable lengths of the inserts is due to the lack of capping during the synthesis of the oligo because such variation has not been observed in more standard syntheses. The frequencies of individual dinucleotides (listed from 3′ to 5′, the direction of chemical synthesis) are presented in Fig. 5. The frequency of G residues (0.323) is significantly greater than expected (0.25). However, as this deviation has not been observed in other experiments (Zon et al., 1985; our unpublished results), it is probably not inherent in the synthetic chemistry but rather due to minor differences in the ratios of the nt in the mixture used to generate the central N50 region. There are two possibly significant deviations from what is expected given the nt frequencies: 3′-GT-5′ occurs less often and 3′-AT-5′ occurs more often. However, the deviations are not great enough to draw definitive conclusions.

	Number of nt					
	5′–C	G	T	A	Sum	nt freq.
3′–C	60	91	61	56	268	0.245
G	92	116	61	85	354	0.323
T	60	78	57	47	242	0.222
A	53	73	64	40	230	0.210

Total of 1094 nt

Fig. 5. Frequency of specific nucleotides and dinucleotides. Each entry in the matrix (3′ nt listed in rows, and 5′ nt listed in columns) represents the number of occurrences of a particular dinucleotide (out of a total of 1094 dinucleotides obtained from the sequences shown in Fig. 4). Each listing in the Sum column represents the number of occurrences for each of the 4 nt from which the nt frequencies (nt freq. column) are calculated.

DISCUSSION

The methods described in this paper permit the synthesis and the cloning of random DNA sequences. Various aspects of the method are worthy of comment. (*i*) The cloning efficiency is sufficiently high that it is possible to create large collections of recombinant molecules, each containing a single segment of random DNA. In other experiments, we have generated pools composed of 10^6 independent phages. (*ii*) It is possible to vary the length of the random DNA sequence. We have synthesized oligos whose random DNA regions are as short as 14 bp and as long as 55 bp. For longer oligos, the yields can be increased significantly by omitting the capping reaction after each step of random addition although this also has the consequence that the collection is more heterogeneous in length. (*iii*) Using appropriate 5′ and 3′ end sequences, the random oligos can be cloned into a wide variety of ds DNA molecules. This is particularly important because most of the useful applications of random DNA require the insertion into specially devised molecules with severe constraints on available restriction endonuclease cleavage sites. The only constraint on the synthesis of the oligo is that the 3′ end must be palindromic. The 5′ end does not have to be recognized by a restriction endonuclease as the mutually primed synthesis generates blunt ends. (*iv*) The nt composition of the central region can be varied simply by using appropriate concentrations of the nt precursors.

The homoduplex molecules produced by mutually primed synthesis represent an accurate reflection of the original oligo mixture because they are generated in a manner that is independent of the DNA sequences of the central region. Thus, by using appropriate concentrations of the nt precursors at appropriate steps of the DNA synthesis, this method can be used to clone oligos with any degree of degeneracy including pure species. This modification makes it possible to use the products of a single DNA synthesis to saturate a defined nt sequence with bp substitutions. For example, we have generated a large number of mutations of the yeast *HIS3* regulatory site by synthesizing and cloning a degenerate oligo whose central region was mutated by including low concentrations of the non-wild-type nt precursor during each step of the synthesis (Hill et al., 1986). In a related mutagenesis procedure, degenerate oligo mixtures for both strands are hybridized and then cloned as heteroduplex molecules (Matteucci and Heyneker, 1983). The ability to clone degenerate oligos as homoduplex molecules eliminates biases due to differential stability or preferential repair of heteroduplexes, and it is essential for the cloning of highly degenerate oligos.

The ability to generate libraries of recombinant DNA molecules containing individual random-sequence oligos makes it possible to select for functional nt sequences by genetic or biochemical means (Oliphant and Struhl, 1986). This general method should prove useful for determining the consensus sequence for a wide variety of genetic elements. Although this method is most easily employed for genetic elements that are defined by short stretches of DNA such as promoter/regulatory elements, it can be adapted to study the requirements for specific regions of proteins. Finally, random nt sequences of defined length will be useful as 'neutral spacers' in studies designed to investigate the spacing relationships of genetic elements, as well as 'control sequences' in a wide range of biochemical, genetic, and physical experiments.

ACKNOWLEDGEMENTS

We thank Sarah Bray and Richard Kolodner for comments on the manuscript. This work was supported by grants from the National Institute of Health (GM 30186) and the Chicago Community Trust (Searle Scholars Program) to K.S. and a grant from the Milton fund to A.L.N.

REFERENCES

Beaucage, S.L. and Caruthers, M.H.: Deoxynucleoside phosphoramidites — a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. 22 (1981) 1859–1862.

Hill, D.E., Oliphant, A.R. and Struhl, K.: Mutagenesis with degenerate oligodeoxynucleotides: an efficient method for saturating a defined DNA region with base pair substitutions. Methods Enzymol. (1986) in press.

Matteucci, M.D. and Heyneker, H.L.: Targeted random mutagenesis: the use of ambiguously synthesized oligonucleotides to mutagenize sequences immediately 5′ of an ATG initiation codon. Nucl. Acids Res. 11 (1983) 3113–3121.

Matteucci, M.D. and Caruthers, M.H.: Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. 103 (1981) 3185–3193.

Messing, J.: New M13 vectors for cloning. Methods Enzymol. 101 (1983) 20–79.

Oliphant, A.R. and Struhl, K.: The use of random-sequence oligodeoxynucleotides for determining consensus sequences. Methods Enzymol. (1986) in press.

Sanger, F., Coulson, A.R., Barell, B.G., Smith, A.J. and Roe, B.A.: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143 (1980) 161–178.

Zoller, M.J. and Smith, M.: Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. Nucl. Acids Res. 10 (1982) 6487–6500.

Zon, G., Gallo, K.A., Samson, C.J., Shao, K-L., Summers, M.F. and Byrd, R.A.: Analytical studies of 'mixed sequence' oligodeoxyribonucleotides synthesized by competitive coupling of either methyl- or β-cyanoethyl-*N*,*N*-diisopropylamino phosphoramidite reagents including 2′-deoxyinosine. Nucl. Acids Res. 13 (1985) 8181–8196.

Communicated by A.J. Podhajska.