Sequence identity of the $n$–1 product of a synthetic oligonucleotide

Jamal Temsamani, Michael Kubert and Sudhir Agrawal*

Hybridon, Inc., One Innovation Drive, Worcester, MA 01605, USA

Received March 8, 1995; Revised and Accepted April 21, 1995

ABSTRACT

After synthesis and purification of an oligonucleotide, the final product usually contains a low level of $n$–1 congeneric species. We have sequenced the $n$–1 population of a 25mer phosphodiester oligonucleotide. The $n$–1 band was cut from the gel and eluted. Oligonucleotides were tailed with dA and annealed to a dT-tailed plasmid. The recombinant plasmid was ligated and used to transform competent bacteria. Our results show that the $n$–1 population was heterogeneous. The frequency of truncated nucleotides at the 3′-end was much higher than at the 5′-end of the oligomer. No truncated nucleotides were found in the last four nucleotides at the 5′-end. Our results also show that the chain of oligonucleotides can grow on unreacted sites of a controlled-pore glass support.

INTRODUCTION

Oligonucleotides, because of their strong affinity and theoretical selectivity for nucleic acid targets, have many potentially exciting applications. Oligonucleotides are being used as diagnostic probes, as tools in molecular biology, e.g. hybridization techniques, PCR (1) and as therapeutic agents (2,3). Several antisense oligonucleotides are presently being studies for their efficacy in human clinical trials (4–6).

The application of synthetic oligonucleotides and their analogs has grown considerably because of recent advances in oligonucleotide synthetic chemistry (7). The solid-phase approach using phosphite triester (8), phosphoramidite (9) or H phosphonate (10), is the most common method used for the synthesis of oligonucleotides and various analogs. The current synthetic approaches to oligonucleotide synthesis involve building an oligomer chain that is anchored to a solid support through its 3′-OH group, and is elongated by coupling to its 5′-OH. The yield of each coupling step in a given chain-elongation cycle will generally be <100%. For an oligonucleotide of length $n$, there are $n$–1 linkages and the maximum yield of a desired sequence will be [coupling efficiency]$^{n-1}$. For a 25mer, assuming a coupling efficiency of 98%, the calculated yield of the full-length product will be 61%. The other 39% consist of all possible shorter length oligonucleotides resulting from inefficient monomer coupling. The desired oligonucleotide can be purified from this mixture by purification steps using ion exchange or reverse phase chromatography. The purification procedures used routinely are not 100% effective and fail to eliminate these populations. The final product is therefore contaminated with a low level of $n$–1 (and to some extent $n$–2 and $n$–3) congeneric species. Since oligonucleotides are used as diagnostic probes and as therapeutic agents, the contaminating species should be identified. Development of suitable methods for establishing the identity, uniformity and purity of synthetic oligonucleotides and analogs is a critical challenge. Recently, we reported a protocol for sequencing oligonucleotides and analogs (11), modified from Okayama/Berg’s method (12). In this report, we have applied this method to the sequencing of the ($n$–1) population of a 25mer oligonucleotide to confirm the identity of the sequence.

MATERIALS AND METHODS

Oligonucleotide synthesis

The 25mer phosphodiester oligonucleotide (5′-CTC TCG CAC CCA TCT CTC TCC TTC T-3′) was synthesized using phosphoramidite chemistry on an automated synthesizer (Millipore, model 8700) on a 1 μmol scale. After assembly of the required sequence, the controlled-pore glass (CPG) bound oligonucleotide preparation contained ~50% full-length material as determined by polyacrylamide gel electrophoresis (PAGE).

Sequencing

The crude oligonucleotide was separated by 20% PAGE. The band corresponding to $n$–1 (24mer) was cut from the gel under UV shadowing and eluted overnight in 0.5% ammonium acetate. The $n$–1 oligonucleotide was then desalted on Waters Sep-Pak C18 cartridges (Waters). Sequencing was carried out as described in (11). Briefly, 1 μg of the oligonucleotide ($n$–1) was phosphorylated at the 5′-end with ATP and T4 polynucleotide kinase (13). The phosphorylated oligonucleotide was then tailed with 0.2 mM dATP and 2.5 U terminal transferase in 100 mM cacodylate (pH 6.8), 1 mM CoCl2, 0.1 mM dithiothreitol, 0.1 mg/ml BSA for 30 min at 37°C. Ten μg of the plasmid pSL1190 (Pharmacia) was cut with PstI. Tails of dT were added to the 3′ terminus as described above, and the plasmid was cut with Smal to leave just one tailed 3′-end. 400 ng of oligonucleotide and 100 ng of plasmid were annealed by virtue of their complementary tails in 50 mM NaCl, 10 mM Tris–HCl (pH 7.5), 5 mM MgCl2, 7.5 mM dithiothreitol.

*To whom correspondence should be addressed
Steps in the construction and sequencing of oligonucleotide-plasmid recombinant.

Figure 1. Analysis of the synthesis of a 25mer oligonucleotide by PAGE. After synthesis and deprotection, 0.5 optical density (OD) of the crude oligonucleotide was analyzed on a 20% polyacrylamide gel.

RESULTS AND DISCUSSION

The process of synthesizing oligonucleotides generates shorter-length oligonucleotides, especially n−1 product, as well as the full-length product. Even after purification, the final product is contaminated with a low level of this n−1 species. Very little, if anything, is currently known about this n−1 population, but given the highly specific nature of oligonucleotides, it is important to identify it.

Recently, we described a protocol to sequence oligonucleotides and analogs (11) (Fig. 1). Briefly, the oligonucleotide is tailed with dATP at the 3′-end and then annealed to a dTTP-tailed plasmid. After ligation, the plasmid is used to transform competent bacteria, at which individual clones are picked up randomly and sequenced, once plated. Since, theoretically, each clone contains one type of molecule, this method is ideal for sequencing a mixture of oligonucleotides.

We used this protocol to sequence the n−1 product of the synthesis of a 25mer phosphodiester oligonucleotide. After synthesis of the 25mer oligonucleotide and deprotection in ammonia, the oligonucleotide was analyzed by gel electrophoresis. Figure 2 shows that the synthesis contained ~50% of the full-length product and the remaining product was a population of shorter-length oligonucleotides (n−1, n−2, n−3, etc.). The n−1 band represented ~5% of the total product. This band was cut out from the gel and eluted overnight. In order to be sure that only the n−1 band (24mer) was cut, a small aliquot of the n−1 product was labeled with [γ-32P]ATP and T4 polynucleotide kinase and compared to the 25mer oligonucleotide. Figure 3 shows that only one band was observed, confirming that only the n−1 band was cut out from gel. Tailing of the oligonucleotide n−1 with dATP and terminal transferase showed that almost all of the n−1 population was tailed. The size of the tail was between 12 and 20 dA residues (Fig. 4). After annealing with plasmid and cloning, we picked up more than 70 clones and sequenced them as described in Materials and Methods. Figure 5 shows a sample of these sequencing reactions. Table 1 shows the sequence of all the n−1 population obtained. Since the sequence of the 25mer oligonucleotide contains two pairs and one set of three identical nucleotides, it is difficult to know which nucleotide in the set is missing. Therefore, in Table 1, the frequency is represented for two or three nucleotides of the set. The sequencing method also determines the length of the oligonucleotide—100% of the clones were 24mers, confirming that only the n−1 band was cut out from the gel.
**Table 1. Sequence of the n-1 products**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>No of clones</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-CTC TCG CAC CCA TCT CTC TCC TTC T-3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC TCC TTC X T</td>
<td>11</td>
<td>15.7</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC TCC XX T</td>
<td>12</td>
<td>17.1</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC TCC X X C T</td>
<td>9</td>
<td>12.8</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC TXX TTC T</td>
<td>6</td>
<td>8.75</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC XCC TTC T</td>
<td>4</td>
<td>5.7</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC XX T</td>
<td>3</td>
<td>4.3</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CXC TTC T</td>
<td>3</td>
<td>4.3</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT TTC T</td>
<td>11</td>
<td>15.7</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TCT CTC TTC T</td>
<td>2</td>
<td>2.8</td>
</tr>
<tr>
<td>CTC TCG CAC CCA TX T CT T T T T</td>
<td>3</td>
<td>4.3</td>
</tr>
<tr>
<td>CTC TCG CAC CCA X CT CTC TTC T</td>
<td>3</td>
<td>4.3</td>
</tr>
<tr>
<td>CTC TCG CAC CCA C X T CT TTC T</td>
<td>2</td>
<td>2.8</td>
</tr>
<tr>
<td>CTC TCG CAC CCA CX X AT CT CTC TTC T</td>
<td>2</td>
<td>2.8</td>
</tr>
<tr>
<td>CTC TCG CAC CCA X C X TTC T</td>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>CTC TCG CAC CCA X C X X CT CTC TTC T</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CTC TCG CAC CCA X C X X X X CT CTC TTC T</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CTC TCG CAC CCA X X X X X X CT CTC TTC T</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Seventy clones containing the n-1 molecules were sequenced as described in the Materials and Methods. The number of clones obtained for each sequence and the percentage are represented.

Several significant points emerge from the analysis of the results in Table 1. First, the n-1 population is heterogeneous. The truncated nucleotide was missing from almost all positions, but the frequency was higher at the 3'-end of the oligonucleotide than at the 5'-end. More than 45% of the clones analyzed were found to be truncated by one nucleotide out of the last 4 nt at the 3'-end. We could not detect any truncated nucleotides in the last 4 nt at the 5'-end. In the central region, this frequency was almost constant, between 2 and 4%.

The n-1 population probably results from inefficient coupling followed by incomplete capping and/or detritylation, and/or oxidation. During synthesis, the dimethoxytrityl (DMTr) protecting group at the 5'-end of the growing oligonucleotide chain is removed by treatment with acid. The newly liberated 5'-end of the oligomer is then coupled to the 3'-end phosphoramidite derivative of the next deoxynucleoside to be added to the chain. Any unreacted 5'-end (the coupling efficiency is <100%) is capped by acetylation, to block its extension in subsequent coupling reactions and prevent the extension of erroneous oligonucleotides. However, the capping reaction is not quantitative, it does not proceed at 100% efficiency. Theoretically, one might expect the frequency to be the same all along the chain of the oligonucleotide. Our results suggest, however, that the coupling reaction is not consistent during synthesis of the oligonucleotide.

The fact that nucleotides are missing more frequently from the 3'-end suggests that coupling/capping/detritylation/oxidation reactions proceed with lower efficiency at the 3'-end than the 5'-end, probably due to a steric hindrance. It is possible that at the 3'-end, the short growing chains are close to each other and the encumbrance of DMTr groups impedes the reaction and makes the chain elongation proceed with lower efficiency. This steric hindrance can also decrease the efficiency of the detritylation reaction that removes the DMTr groups, leaving some of the chains protected and unavailable for the next coupling. When the growing chain is >6 nt, there is less steric hindrance, therefore, the frequency goes down to 2-5%. At the 5'-end, our data suggest that the reaction proceeds at 100% efficiency.

One surprising result, shown in Table 1, was the high frequency of the truncated nucleotide at the 3'-end, which is attached to the CPG support, indicating that the synthesis can grow on uncapped sites of the CPG support. The nucleoside coupling reactions used to attach the first nucleoside to the linker arm of the CPG support,
never proceed quantitatively, so it is always necessary to cap the unreacted linker sites, otherwise, these surface groups may react to produce shorter chains (14). Inefficient capping may lead to the presence of unreacted amino-linker sites and silanol moieties, as clearly happened in our experiments.

We recently reported that we could reduce the n−1 population by adding a pre-capping step in the synthesis and two-step deprotection with ammonium hydroxide (15). High loading capacity of the first nucleoside on CPG support also reduces the number of unreacted sites (15). In our case, the loading efficiency of the first monomer (T) on the CPG was about 30–40 μmol/g of CPG.

From a therapeutic point of view, a fair amount of this n−1 population can be active and bind to the target nucleic acid. More than 45% of the truncated nucleotides are located in the first 4 nt which suggests that these truncated species can bind to the target with ≥21 Watson–Crick base-pairings.

In summary, our results shed light on the identity of the n−1 population. They indicate that the n−1 is a heterogeneous population, which results from inefficient coupling followed by incomplete capping and/or detritylation, and/or oxidation. The efficiency of coupling is not consistent during synthesis of oligonucleotides. The oligonucleotide chain growth proceeds with low efficiency at the 3′-end and then improves, reaching 100% at the 5′-end. The n−1 population can also result from inadvertent chain growth on uncapped sites of the CPG support. Addition of the pre-capping step (15), increasing the time of the detritylation reaction during the first steps of synthesis (16), or use of alternate capping reagents may reduce the n−1 contents (17). These results will help to improve the synthesis of oligonucleotides and analogs synthesized by the phosphoramidite method. Although this report describes only the n−1 sequence of a phosphodiester oligonucleotide, the identity of n−1 should be the same for phosphorothioate, phosphorodithioate, oligonucleotide and other oligonucleotide analogs.

REFERENCES