A simple method for preparing pools of synthetic oligonucleotides with random point deletions

Daniel K. Treiber and James R. Williamson*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received June 27, 1995; Accepted August 4, 1995

Mutagenized pools of nucleic acid sequences are powerful tools for identification of positions that have an important role in the function of a particular RNA or DNA molecule, and are a central feature of directed molecular evolution technology. Synthetic oligonucleotides containing either defined mutations or a predetermined level of degeneracy at one or more positions are often used in these studies. We describe a simple method for introducing random point deletions into synthetic oligonucleotides. The method is compatible with, and complements, current methods for constructing pools of partially degenerate DNA sequences and should prove useful in a wide range of studies.

In our method, random point deletions are introduced at a predetermined frequency by altering the concentration of the detritylation reagent used for DNA synthesis. The 5' terminal base on the solid support used in automated DNA synthesis harbors a dimethoxytrityl (DMT) protecting group on the 5' hydroxyl to prevent the addition of multiple bases during each synthesis cycle. Before a subsequent base can be coupled, the DMT group must be removed by acid treatment. If detritylation is incomplete, then a fraction of the molecules will not undergo coupling during that cycle; however, these molecules are not irreversibly terminated and retain the capacity to detritylate and couple in subsequent cycles. The product molecules will thus contain point deletions at a frequency that is inversely related to the detritylation efficiency. The efficiency of detritylation can be simply altered by adjusting the acid concentration in the detritylation solution. In contrast to a previously described method for synthesizing oligonucleotides with random point deletions (1), our method does not require adjustment or optimization of the phosphoramidite concentrations. As a result, our method is simple, inexpensive to optimize, and fully compatible with current protocols for synthesizing partially randomized sequences (2).

All DNA synthesis was performed using an Applied Biosystems 391 DNA synthesizer (PCR-Mate) using the 0.2 μ mol scale synthesis cycle (63 steps). Manually assembled 1 μ mol scale column bodies loaded with 0.2 μ mol of DNA synthesis resin were used throughout. Alternative column configurations are also acceptable but may require an independent optimization of detritylation conditions. Detritylation efficiency was altered by substituting varying concentrations of dichloroacetic acid (DCA) in methylene chloride for the standard detritylation reagent [2.5% trichloroacetic acid (TCA) in methylene chloride]. The efficiency of removal of the DMT group from each of the four bases was determined at four DCA concentrations using a two-step

protocol. The base was first coupled to the resin, then detritylated with the desired concentration of DCA, collecting the released DMT group for quantitation by the standard colorimetric assay. Briefly, the DMT-containing waste fraction was mixed with 7 ml of 0.1 M *p*-toluene sulfonic acid (in acetonitrile) and the absorbance at 498 nm was measured. The DCA bottle was then replaced with 2.5% TCA (w/v)—the standard reagent used to remove 100% of the DMT groups. The resin was subjected to a second detritylation to remove any DMT groups not liberated by the DCA treatment. The efficiency of detritylation was calculated by dividing the absorbance of the DCA sample by the sum of the TCA and DCA sample absorbances.

The results of the detritylation tests are shown in Table 1. The detritylation efficiencies for each base were highly reproducible, and concentrations of DCA between 0.17 and 0.85% (v/v) encompass the efficiency range that would be of practical use for most applications. Also evident in Table 1 is the observation that the four bases have different rates of detritylation, with the rates following the trend G>>A>C>T. The variation in detritylation rates will have the potentially undesirable effect of biasing the frequency of deletions at certain positions, with most deletions 5' to pyrimidines and a smaller fraction 5' to purines. The difference between the predicted deletion probability 5' to A, C and T is clearly significant, however, at all DCA concentrations the maximum difference is less than a factor of 10, and often less than a factor of five. As a result, pool complexity should not be greatly compromised at these positions. Deletions 5' to G, however, will be greatly under-represented at DCA concentrations useful for the other three bases (0.85–0.5%). There are at least two ways to circumvent this potential problem. First, since the coupling efficiency at each cycle is 98–99%, some basal deletion frequency might be expected by omitting the capping step from the synthesis cycle. With this approach, deletions 5' to G are guaranteed to be included in the population at a low, but significant, frequency. Secondly, the DCA concentration can be reduced to <0.5% after coupling each G to correct for the rapid detritylation of this base. This approach could have the effect of skewing the distribution of tandem deletions in the vicinity of Gs—which may or may not be desirable.

The method was tested by synthesizing a 62mer oligonucleotide in which random deletions were introduced into the central 32 nucleotide (nt) region. The 15 nt flanking sequences served as primer binding sites for subsequent PCR reactions and were synthesized using 2.5% TCA to guarantee complete detritylation.

* To whom correspondence should be addressed

The central 32 nt region contained equal numbers of the four bases and was synthesized using 0.7% DCA. In addition, the capping step was omitted from the cycle to increase the probability of incorporating deletions 5' to G. Based on the detritylation efficiencies measured in Table 1, and assuming a Poisson distribution of deletions, an average of approximately three deletions per molecule was predicted (the mathematics of pool diversity are discussed in reference 3). Deprotected products of the synthesis were 5' end-labeled with [32 P]ATP and separated by electrophoresis through a 20% denaturing polyacrylamide gel. A control 62mer, synthesized using 2.5% TCA throughout, was also prepared for comparison. Traces of the gel image recorded on a Phosphorimager are shown in Figure 1. As expected, the majority of products synthesized with 0.7% DCA contained 3–4 deletions (Fig. 1A), and only a small fraction was full length, whereas the control oligonucleotide products were predominately full length (Fig. 1B).

Table 1. Base-specific detritylation efficiencies as a function of DCA concentration

% DCA (v/v)	A	C	G	T
0.85	98.3	98.0	>99.0	92.0
0.70	96.8	90.0	>99.0	77.0
0.50	93.2	85.5	>99.0	67.0
0.17	43.0	27.0	81.0	17.0

Percent detritylation values are the average of three trials, and the standard errors were generally <1–2%.

The deletion pool of oligonucleotides was converted to double stranded DNA by performing PCR with primers complementary to the 15 nt constant regions. The primers were tagged with restriction enzyme recognition sites such that the PCR products could be ligated into a pGem cloning vector for DNA sequence analysis. Plasmid DNA isolated from 10 independent clones was sequenced, and a compilation of the results is shown in Figure 1C. The frequencies and positions of the observed deletions were essentially as predicted by the results of Table 1 and Figure 1A. Most deletions were 5' to pyrimidines, with a smaller fraction observed 5' to As. None of the deletions could be unambiguously assigned to positions 5' of G; however, this result was anticipated because only a small number of clones was analyzed, and deletions at these positions were expected at a low frequency.

In conclusion, we describe a simple, flexible method for constructing pools of oligonucleotides containing random point

Figure 1. Characterization of a point deletion library. (A) Size distribution of a 62mer oligonucleotide in which the internal 32 bases were synthesized with 0.7% DCA. (B) The 62mer was prepared using standard detritylation conditions throughout (2.5% TCA). (C) Sequence analysis of the deletion pool from (A). A compilation of the data from 10 independent clones is shown, and the bold G at the 3' terminus was the first base to be detritylated with 0.7% DCA.

deletions. The method can be used in conjunction with partial randomization procedures and should have applications in fields ranging from structure–function analysis to directed molecular evolution.

ACKNOWLEDGEMENTS

We thank members of the Williamson laboratory for their helpful contributions. This work was supported by a grant from the NIH (GM-39589) to J.R.W. and an American Cancer Society postdoctoral fellowship to D.K.T.

REFERENCES

1. Ner, S.S., Atkinson, T.C. and Smith, M. (1989) *Nucleic Acids Res.*, **17**, 4015–4023.
2. Beaudry, A.A. and Joyce, G.F. (1992) *Science*, **257**, 635–641.
3. Breaker, R.R. and Joyce, G.F. (1994) *Tibtech*, **12**, 268–275.