Nucleic Acids Research, 1998, Vol. 26, No. 22 5225–5227

The use of tailed octamer primers for cycle sequencing

S. Ball[+], M. A. Reeve, P. S. Robinson[§], F. Hill[1], D. M. Brown[1] and D. Loakes[1,*]

Nycomed Amersham plc, Amersham Laboratories, White Lion Road, Amersham, Buckinghamshire HP7 9LL, UK and [1]Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received September 10, 1998; Revised and Accepted October 5, 1998

ABSTRACT

Studies have been carried out on the use of octamer oligonucleotides tailed with different base analogues as primers in cycle sequencing reactions. 5-Nitroindole tails improved the performance as primers of a number of octamers. A tail length of three or four 5-nitroindole residues significantly increased the sequencing signal intensity for almost all primers. The use of incomplete libraries of tailed octamer primers for primer walking is discussed.

Random sequencing of shotgunned DNA fragments, a feature of the earliest sequencing projects by Sanger's laboratory, remains the preferred approach to large-scale sequencing. However, obtaining the final few percent of any particular sequence (closure) requires a disproportionate effort unless primer walking is employed. A number of attempts have been made, with some success, to use presynthesised sets (libraries) of short primers to avoid the delays inherent in the repeated synthesis of new primers (1,2). A disappointing feature of this approach has been that many short oligonucleotides fail to prime in cycle sequencing. Thus in one report, only approximately one half of 121 nonamer primers worked (3). For any small library of oligomers to be useful, the great majority would have to work, and work well. We have examined whether adding tails of modified bases to the 5′ ends improves the priming properties of short oligonucleotides.

A number of analogues that behave indiscriminately in their base-pairing properties towards the natural DNA bases, while helping to stabilise the primer/template duplex, have been described. Ten different base analogues were tested: 3-nitropyrrole (4), 4-, 5- and 6-nitroindole (5), 5-fluoroindole, benzimidazole (6), 5-nitroindazole, hypoxanthine (which is often used as a universal base analogue; 7) and a 50:50 mixture of the degenerate bases P and K (8). Tails with up to 12 residues were tested in an attempt to stabilise the 8mers; the effect of each of these modifications has been examined in cycle sequencing reactions.

Initially we used an octamer primer comprising the eight terminal bases of the (–40) M13 universal primer (5′-GTCAC-GAC) (9). This works reasonably well without any modification (Fig. 1a), making it possible to demonstrate deleterious as well as beneficial effects of each modification. 5′-tails (3, 6, 9 or 12 residues in length) of each of the modified bases were added to the control octamer primer. The reactions were performed using either 5 or 25 pmol of primer and the annealing temperature was varied from 20 to 50°C, in increments of 10°C, to determine the optimum annealing temperature for cycle sequencing.

Of the 10 analogues tested, the most effective were 5-nitroindole and 5-nitroindazole. However, due to its ease of synthesis, 2′-deoxyribosyl-5-nitroindole was preferred. Full details of the other analogues used will be reported elsewhere. Of the other analogues, the only two that gave any sequence data were 3-nitropyrrole and benzimidazole, though they were much less effective than 5-nitroindole. The 5-nitroindole tails were expected to stabilise the primers while behaving indiscriminately in base-pairing (5,10). Longer tails (>6 residues) were detrimental; they may loop back on themselves, destabilising the primer. Previous work has shown that runs of 5-nitroindole residues can form secondary structures, such as hairpins (11,12), due to self-stacking. Short tails, on the other hand, might increase duplex stability by intercalation. A tail of three 5-nitroindole residues improved the intensity of the sequence data compared to the control primer lacking a tail, at an annealing temperature of 30°C (Fig. 1 and Table 1). At 40°C, good quality sequence data was still obtained with the tailed universal octamer while the control failed to any give analysable data.

To determine the optimum length of the 5′-tail, primers containing one to five 5-nitroindole residues were tested in cycle sequencing reactions, and compared to the control octamer itself. An annealing temperature of 30°C was used and the signal intensities for each reaction are shown in Table 1. These results suggest that the optimum length for the 5-nitroindole tail is four residues.

To determine whether this effect was generally applicable, a further seven octamers were tested both with and without 5′-tails of 5-nitroindole. As in our initial study, we used primers with three 5-nitroindole residues, although improved data may be obtained with four (see example in Table 1). These results are also summarised in Table 1. All the tailed primers produced good sequence data, as shown by the signal intensity data. In particular, in two cases (primers 2 and 5), the addition of a tail proved essential for obtaining sequence data. In all but one case (primer 4), addition of a tail increased the signal intensity. The increased signal intensity (an arbitrary photomultiplier output measurement) improved base calling as measured by the number of 'Ns' and, therefore, the amount of useful data.

*To whom correspondence should be addressed. Tel: +44 1223 248011; Fax: +44 1223 412178; Email: dml@mrc-lmb.cam.ac.uk

Present addresses: [+]Department of Biochemistry, Astra Charnwood, Bakewell Road, Loughborough, UK and [§]Genomics, UK Genetics, Glaxo Wellcome, Medicines Research Centre, Stevenage, UK

Table 1. Summary of sequencing data using octamers with 5′-tails of 5-nitroindole (denoted by 5) of increasing length (top) and of seven pairs of primers

Primer	Sequences	G	A	T	C
Universal	GTCACGAC	168	154	136	143
	5GTCACGAC	197	195	186	183
	55GTCACGAC	237	226	244	219
	555GTCACGAC	240	256	244	236
	5555GTCACGAC	291	223	262	264
	55555GTCACGAC	277	238	253	244
1	GGAAGCCC	301	344	352	248
	555GGAAGCCC	395	421	673	374
2	AAATCCGC	43	64	82	53
	555AAATCCGC	235	356	225	221
3	TTCAGCGG	115	218	150	103
	555TTCAGCGG	268	524	367	240
4	AAGTGCCG	236	397	219	201
	555AAGTGCCG	201	317	181	192
5	AACGAGCG	106	223	152	91
	555AACGAGCG	128	337	196	116
6	CGCTGGTT	–	–	–	–
	555CGCTGGTT	70	131	63	62
7	GCGGTTTG	222	174	78	63
	555GCGGTTTG	155	223	146	146

The primers differ by the presence or absence of a three residue 5′-tail of 5-nitroindole. Data is presented as the signal intensities, a fluorescence photomultiplier tube output measurement recorded by the automated sequencer, for each set of reactions. The reactions were performed as described in the Figure 1 legend.

A universal library of primers would accelerate DNA sequencing by primer walking. However, huge numbers of oligomers (4^8 or 65 536 octamers, for example) would be needed unless a sample of the complete library could suffice. Two complementary approaches to reducing the number of primers are discussed here. The first approach is to calculate the probability of finding an octamer, present in a sample of the complete library, within a given distance from the end of a sequence run. The model assumes that each of the four bases have an equal probability of being found at any position. A full account of these calculations is provided as Supplementary Material. This shows that in a library of only 1500 octamer primers, there is a 90% probability of finding an appropriate primer within 99 bases of the end of the previous sequence read, a 99% probability within 198 bases and a 99.9% probability within 297 bases. Although this smaller library can result in long overlaps, this increased redundancy should improve the accuracy of the final sequence. Modification by 5-nitroindole does not increase the size of the primer library.

The second approach exploits the non-randomness of real DNA sequences to choose which primers to include in the library. This approach is based on previous analyses of model organisms by others (13–15). In effect, guesses are made, based on known oligonucleotide frequencies, as to which sequences are likely to be present in the material to be sequenced. Natural DNA sequences are non-random; it has long been appreciated that GC compositions vary widely. This skew from randomness is increased as longer sequences are considered. Thus dinucleotide and trinucleotide frequencies deviate from the values expected based on base composition (15). Markov chain analyses, performed some years ago (13–15), examined the skewed distribution of oligomer frequencies in *Escherichia coli*, *Saccharomyces cerevisiae* and *Drosophila melanogaster*. A pertinent feature of the *Drosophila* analysis was the table of 50 nonamers (Table 6 in 15). Each of these nonamers was predicted to be present in 50% of cosmid clones from that organism. Thus, a library of the most frequent octamers in human coding sequences, for example, could be used for completing EST sequences.

In conclusion, the addition of a 5′-tail of the universal base 5-nitroindole stabilises short (octamer) primers; this improves both the quality of sequence data produced, and the fraction of octamers that can be used, in cycle sequencing. 5-nitroindole was found to be the best of the 10 analogues investigated by us, and the optimum tail length was shown to be four. Further development of these findings should expedite primer walking.

See supplementary material available in NAR Online.

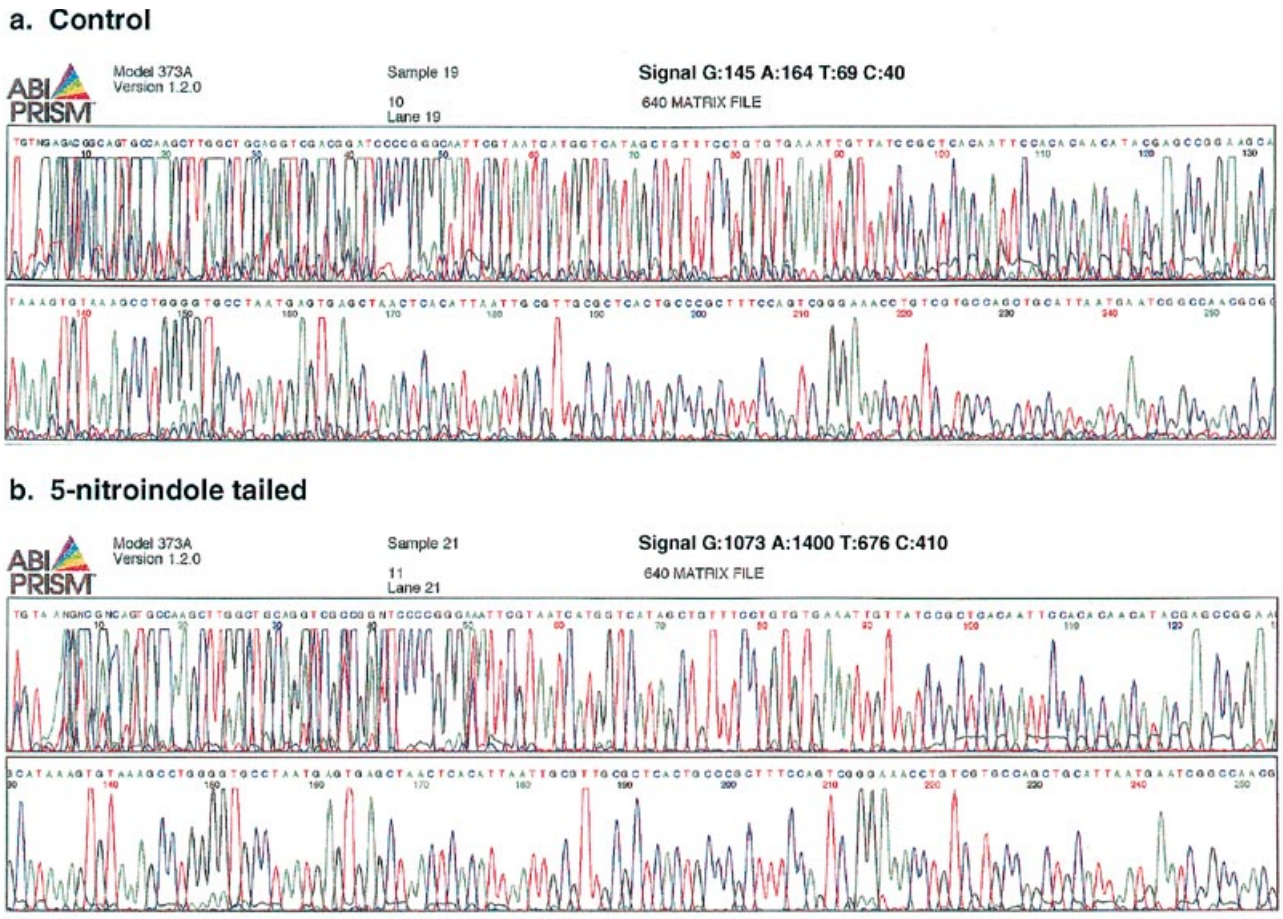

Figure 1. Sequence traces obtained using either (**a**) the control primer (5′-GTCACGAC) or (**b**) the 5-nitroindole (denoted as 5) tailed primer (5′-555GTCACGAC). Note the increased signal intensities in (a) compared to (b). Sequencing was performed using ThermoSequenase™ dye-terminators and samples were run on an ABI 373 fluorescent sequencer. The concentration of each primer was 25 pmol; the template used was 1 µg of caesium chloride purified single-stranded M13mp8 DNA. Thermal cycling conditions used were: 95°C for 5 min followed by 25 cycles of: 95°C for 30 s, annealing for 15 s at 30°C, extension at 60°C for 4 min.

ACKNOWLEDGEMENTS

We thank Richard Grenfell and Jan Fogg for oligonucleotide synthesis, and Nycomed Amersham plc for financial support (to D.L.).

REFERENCES

1 Kieleczawa,J., Dunn,J.J. and Studier,F.W. (1992) *Science*, **248**, 1787–1791.
2 Azhikina,T., Veselovskaya,S., Myasnikov,V., Potapov,V., Ermolayeva,O. and Sverdlov,E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11460–11462.
3 Bock,J.H. and Slightom,J.L. (1995) *Biotechniques*, **19**, 60–64.
4 Bergstrom,D.E., Zhang,P., Toma,P.H., Andrews,P.C. and Nichols,R. (1995) *J. Am. Chem. Soc*., **117**, 1201–1209.
5 Loakes,D. and Brown,D.M. (1994) *Nucleic Acids Res*., **22**, 4039–4043.
6 Papageorgiou,C. and Tamm,C. (1987) *Helv. Chim. Acta*, **70**, 138–141.
7 Ohtsuka,E., Matsuki,S., Ikehara,M., Takahashi,Y. and Matsubara,K. (1985) *J. Biol. Chem*., **260**, 2605–2608.
8 Kong Thoo Lin,P. and Brown,D.M. (1992) *Nucleic Acids Res*., **20**, 5149–5152.
9 Duckworth,M.L., Gait,M.J., Goelet,P., Hong,G.F., Singh,M. and Titmas,R.C. (1981) *Nucleic Acids Res*., **9**, 1691–1706.
10 Loakes,D., Brown,D.M., Linde,S. and Hill,F. (1995) *Nucleic Acids Res*., **23**, 2361–2366.
11 Vallone,P.M. and Benight,A.S. (1997) *Biophys. J*., **72**, TH426 (A421) and personal communication.
12 Loakes,D., Hill,F., Brown,D.M. and Salisbury,S.A. (1997) *J. Mol. Biol*., **270**, 426–435.
13 Phillips,G.J., Arnold,J. and Ivarie,R. (1987) *Nucleic Acids Res*., **15**, 2611–2626.
14 Arnold,J., Cuticchia,A.J., Newsome,D.A., Jennings,W.W. and Ivarie,R. (1988) *Nucleic Acids Res*., **16**, 7145–7158.
15 Cuticchia,A.J., Ivarie,R. and Arnold,J. (1992) *Nucleic Acids Res*., **20**, 3651–3657.