PNA Directed Genome Rare Cutting: New Developments

Maxim D. Frank-Kamenetskii
and Vadim V. Demidov

Abstract

This chapter deals with the PNA-assisted rare cleavage (PARC) of duplex DNA. The technique is a variety of the general ‘Achilles’ heel’ cleavage strategy and uses PNA oligomers to protect very few sites on genomic DNA against enzymatic methylation. As a result, the PARC technique makes it possible to convert common restriction enzymes into a pool of infrequent genome cutters. These artificial genome cutting systems cover the range of recognition specificities, where very few, if any, cutters are now available. Here, we present the PARC-based method for robust purification of yeast artificial chromosomes (YACs) from host chromosomes with similar lengths and demonstrate the PARC potential for rare fragmentation of human DNA. Further progress in the PARC approach also includes the use of pseudocomplementary PNAAs (pcPNAAs) as sequence-unrestricted duplex DNA-binding ligands.
Background

DNA rare fragmentation is used in genomics for genome analysis and may find applications in DNA technology for handling with large DNA molecules. However, a limited number of rare-cutting restriction enzymes (1-4) and homing endonucleases have been identified (4-7). To extend the range of natural DNA rare cutters, their directed modification or selection (8, 9), coupling with DNA methylases (10) and employment of specially inserted sites (11, 12) could be used. Also, artificial ‘restriction enzymes’ combining a variety of DNA binding and cleavage functions have been created (13-15). Still, the progress in this direction is limited, especially for creation of 8-12 bp cutters (16). Considering that this range of recognition specificities corresponds to the most useful sizes of large DNAs—from several dozen kilobases to few megabases, further expansion of approaches for rare cutting the genomes is highly desirable.

In 1996, we proposed a variety of the ‘Achilles’ heel’ rare cleavage strategy (17, 18), the PARC approach (19-21; see refs. 16 and 20 for schematics), which is based on targeting the double-stranded DNA (dsDNA) with cationic pyrimidine bis-PNAs (22-27). A major idea underlying this method is that sequence-specific binding of bis-PNA to relatively short dsDNA sequences should block DNA recognition

Figure 1. Schematics of the PARC method involving pcPNAs. In contrast to triplex-forming pyrimidine bis-PNAs, sequence-unrestricted recognition of dsDNA by pcPNAs via the double-duplex invasion makes it now possible to completely cover the desired methylation/restriction site. For the original similar PARC schematics with bis-PNAs, see refs. 16 and 20.
by the DNA methylase in case the PNA binding site overlaps with the enzyme binding site. At that time, only pyrimidine PNAs could be used for dsDNA targeting, and hence for PARC. Nevertheless, this variety of PNAs has potentially provided, despite its sequence-limitations, with a large pool of different nonnatural DNA rare cutters (16, 19-21).

Recently, a new kind of PNAs, pseudocomplementary PNAs (pcPNAs), has been designed that can bind to essentially any chosen dsDNA site (28-30). In contrast to pyrimidine PNAs that form with dsDNA the invasion triplexes (23, 27, 31), pcPNAs sequence-specifically target the DNA duplex via the double-duplex invasion mode of DNA recognition (28-30). These novel PNAs, when employed in the framework of ‘Achilles’ heel’ strategy (Figure 1), make it also possible to infrequently cut genomic DNA (29), thus further extending the range of PARC-based DNA rare cutters.

Altogether, the use of pyrimidine bis-PNAs and pcPNAs can readily yields a number of PARC cutters with 8-12 bp recognition specificities (see examples below). As a result, a difficult task of rare cutting the genomes with unknown sequences using the diverse cutters becomes possible. This may significantly improve and simplify numerous techniques in the entire field of DNA analysis, including genome mapping, cloning and sequencing.

Protocol

Materials

High-molecular-weight DNA samples with freely dissolved or agarose-embedded (for very large, sub-megabase or longer DNAs; see refs. 32, 33) molecules; a pair of methylation/restriction enzymes with coincident DNA recognition specificities; bis-PNA or pcPNAs having 7-or 8-mer DNA target site overlapping with or embodying the enzymatic targets (bis-PNAs can be purchased from some biotechnological companies (e.g., from Applied Biosystems) or obtained by manual (34) and automated (35; see also this book)
syntheses in a lab from the commercially available protected PNA monomers; pcPNA oligomers and corresponding protected monomers are presently not commercially available); buffer solutions: PNA binding buffer (buffer A: any of suitable buffers (10-20 mM) with pH 6.5-7.0; addition of 1-2 mM EDTA and 10 mM NaCl is recommended, in case they are not the buffer components); DNA methylase buffer (buffer B: use an appropriate commercially available or self-prepared buffer for enzymatic methylation with pH close to 7.0 containing EDTA or some other chelating reagent); ‘stop´ buffer solution (buffer C: 1% SDS; 500 mM NaCl; 10 mM EDTA; 50 mM Tris-HCl; pH 9.0; preincubate buffer at 37°C before use to dissolve all components); restriction enzyme buffer (buffer D: use any appropriate commercially available buffer for restriction digest; to prevent precipitation, avoid K+ cations if possible; addition of SAM or BSA is recommended for some restriction enzymes).

**Step 1: PNA Targeting**

For DNA solution, incubate the DNA sample (typically ≤1 µg) with the requisite PNA(s) in 10-100 µl of buffer A for about 1 h at 37°C. The necessary PNA concentration is usually within 0.2-0.5 µM range depending on the PNA binding affinity and incubation time. After that, remove unbound PNA by gel filtration of the sample through Sephadex G-50.

In case of agarose-embedded DNA, PNA-DNA complexes are formed at 37°C by shaking incubation of the 10-50 µl gel sample containing ~1 µg of DNA for several hours with 50 µl of 0.2-0.5 µM PNA(s) dissolved in buffer A. Note that more concentrated PNA may be required for the in-gel PNA targeting, as compared to the similar treatment in solution. The gel-filtration procedure is replaced here by washing the gel sample twice for 30 min in 100 µl of buffer B.

To enhance the specificity of PNA binding (if necessary), prior washing from PNA in high-salt buffer solution (up to 1 M NaCl) and incubation of the sample at elevated temperature (~50°C) for a short time (≤10 min) is recommended: incorrect PNA-DNA complexes are less stable
than correct ones, therefore their differential dissociation may improve the selectivity of DNA targeting (24, 27, 36).

**Step 2: DNA Methylation**

Next, the gel sample containing DNA targeted by PNA is saturated at 4°C (a low temperature is used to prevent the contaminating endonuclease activity) for about 10 h with the appropriate quantity of DNA methylase (usually, 10-20 U of freshly diluted methylase are necessary) in 50 µl of buffer B. To make the enzyme more stable, an addition of BSA (up to 100 µg/ml) to buffer B is advised at this stage. Periodical tipping of the sample is required. Then, SAM is added up to 300 µM and DNA methylation is performed for 2-4 h at 37°C with shaking. Incubating the sample in 100 µl of buffer C for 20-30 min at 37°C stops the methylation. Thereafter the methylase is finally inactivated and the PNA-DNA complex is dissociated by incubation of the sample for about 1 h at near 60°C (dependent on the melting point of the agarose matrix) in the same buffer. Similar treatment is performed with DNA solutions except less amount of methylase and methylation time are necessary, and gel filtration is used here to change the solution.

**Step 3: DNA Cleavage**

Afterwards, the ‘stop’ solution is removed, and DNA sample is equilibrated at least thrice in 100-200 µl of buffer D for 30 min at 37°C with shaking. In case that buffer D contains K+ cations, washing the DNA sample twice in 100 µl of 50 mM Tris-HCl (pH 7.5) for 30 min at 37°C before its equilibration in buffer D is strongly recommended to avoid precipitation of dodecyl sulfate potassium salt from buffer C. DNA digestion is performed by 5-15 U of a desired restriction enzyme for about 5 h at 37°C (in some cases, other temperatures may be required for restriction enzyme activity) with periodical shaking by tipping. The digestion is stopped by incubating the sample for 20 min at 37°C in 50 mM EDTA (pH 8.0). Similar treatment is performed with DNA solutions except less restriction
enzyme and digestion time are necessary, and gel filtration is used here to change the solution.

**Step 4: PFGE Separation**

If necessary, equilibrate the sample in TE buffer (pH 7.5) and check the PARC digest by pulsed-field gel electrophoresis (PFGE) or use this procedure for preparative purposes. Additional materials and equipment: the pulsed-field certified or chromosomal grade agarose and a PFGE apparatus; use the electrophoretic-pure reagents on all previous stages and the low-melting agarose plugs compatible with PFGE.

**Examples**

**Example 1: Purification of YACs**

YACs are commonly used as convenient vectors to clone very long pieces of DNA and to transfer large genes or gene clusters into mammalian cells (37-39). One of the problems with the YAC technology stems from the fact that YACs are maintained in yeast cells as single-copy objects and are very similar to yeast endogenous chromosomes in their general construction and size. As a result, the YAC isolation, which is usually based on the PFGE separation (37, 40), is often complicated by the presence of comigrating or closely migrating host chromosomes. Therefore, significant contamination of PFGE-isolated YACs by endogenous yeast DNA often occurs, which can compromise their use for subsequent DNA transfer, subcloning, sequencing and amplification procedures (38, 40, 41). Here we show an application of the PARC method to PFGE purify YACs from comigrating endogenous yeast chromosomes with similar lengths (42).

The sequencing data for *S. cerevisiae* indicate that each yeast chromosome contains several unique 12-bp-long sites with generic sequence Pu7GCGCPy suitable for the PARC strategy. Specifically, the Pu7G octapurine part of these sites can be targeted by bis-PNAs
with sequences Py\textsubscript{7}C-linker-JPy\textsubscript{7} (here and below, J denotes pseudoisocytosine and linker consists of three 8-amino-3,6-dioxoctanoic acid units (22)). The PNA binding will mask the adjacent recognition site of DNA methylase \textit{HhaI} (GCGC) from methylation. Consequently, the restriction enzyme \textit{HaeII} should cut, after the PNA removal, only those designated yeast chromosome that overlaps with a specific YAC during PFGE and carries the unique unmethylated site PuGCGCPy.

Note that any chosen 12 bp DNA sequence is met once per ~10 Mb, on average. It is therefore highly unlikely that a YAC, whose sequence is usually unknown but normally does not exceed 2 Mb, will carry the same target site for PARC. Hence, YAC should remain intact whereas the comigrating chromosome will be quantitatively digested. In case YAC will also be cut by the given bis-PNA/M\textit{HhaI}/R\textit{HaeII} combination (M and R stand here for methylation and restriction), the PARC method with another bis-PNA targeted to different site can be applied to eliminate only the required endogenous yeast chromosome. It is virtually improbable that the same YAC will be cut again.

Figure 2 demonstrates the feasibility of this approach (42). Yeast strain 731\textsubscript{G}_2 (CEPH human YAC library) was chosen, which is derived from the commonly used \textit{S. cerevisiae} AB1380 and contains a YAC vector with an insert from human chromosome 2. The size of this YAC practically coincides with the size of yeast Chr II (see Figure 1, lane 3; here and below, Chr followed by a Roman numeral abbreviates a specific yeast host chromosome). Chr II contains near its center the AG\textsubscript{4}A\textsubscript{2}GCGC\textsubscript{2} site suitable for the PARC procedure. We accordingly split Chr II into two fragments by using cationic bis-PNA [TC\textsubscript{4}T\textsubscript{2}C-linker-JT\textsubscript{2}J\textsubscript{4}T]\textsuperscript{(+4)} (PNA 1) and a pair of M/R enzymes (Figure 2A). As a result, we segregated YAC from Chr II (Figure 2BC): quantitative densitometry confirmed that at least 90% of Chr II was cut while ~90% of YAC remained unaltered.

To check the efficiency of the PARC-based YAC purification for DNA subcloning, we sequenced DNA fragments recovered from the gel slices corresponding to YAC after PFGE separation and \textit{HindIII} digestion and subcloned into pUC19. Before the PARC treatment, 5
out of 10 randomly chosen subclones carried Chr II DNA. By contrast, none of 10 randomly chosen subclones obtained after the PARC treatment contained yeast DNA: they all carried only human DNA derived from YAC.

One more example with another YAC not shown here demonstrates the universality of the proposed approach (42). This YAC carries a fragment of human chromosome 6 and significantly overlaps with both Chr II and Chr XIV (second yeast chromosome from the top in Figure 2). Thus, now two yeast chromosomes have to be selectively cut by the PARC method leaving YAC intact. To this end, PNA 2,
PNA Directed Genome Rare Cutting

[T2C2T3C-linker-JT3J2T2](+5), was additionally employed to target Chr XIV. This chromosome carries the PARC target site A2G2A3GCCG2 for PNA 2/M\textit{Hha}I/R\textit{Hae}II combination near the chromosomal end. Hence, the PARC digest is expected to result in a large fragment of Chr XIV coinciding with Chr X (third yeast chromosome from the top in Figure 2), which migrates well ahead of YAC.

Using this pair of PNAs and the \textit{M\textit{Hha}I}/\textit{R\textit{Hae}II} pair of enzymes, successful segregation of the second YAC from two comigrating chromosomes has been achieved (42). Altogether, these data prove that the PARC procedure based on the \textit{M\textit{Hha}I}/\textit{R\textit{Hae}II} pair of enzymes and a pool of appropriate bis-PNAs may be an efficient tool for purification of intact YACs by creating convenient PFGE ‘windows’ in the corresponding range of DNA lengths. Due to very high selectivity of the PARC strategy, YACs must remain intact whereas the host chromosomes will be completely degraded directly in the agarose gel plug carrying extracted YACs.

Example 2: Use of pcPNAs

We have found that pcPNAs as short as octamers form stable and sequence-specific complexes with duplex DNA and effectively compete with DNA methylases for common binding sites (29). Interference of pcPNAs with methylases protects selected DNA sites containing all four nucleobases from enzymatic methylation. The pcPNA-assisted protection against enzymatic methylation is more efficient when the PNA-binding site embodies the methylase-recognition site rather than overlaps it (see schematics in Figure 1). The experiment, whose results are presented in Figure 3, demonstrates that this pcPNA ability can be used for sequence-unrestricted DNA rare cleavage, similar to pyrimidine bis-PNAs.

In this study, the self-pseudocomplementary PNA 3, \[^{3}U^\text{UGD}^\text{UCDD}\](+4), was employed that contains 2,6-diaminopurine (D) and 2-thiouracil (\(^3U\)) as pseudocomplementary substitutes for A and T, respectively (29). 35.9-kb-long adenovirus-2 DNA that carries the pcPNA-binding sites at positions 4.0 and 29.4 kb was subjected to
the PARC treatment using the PNA 3/M\textit{dam}/R\textit{DpnII} combination. Note here that binding site for PNA 3 embodies the M\textit{dam}/R\textit{DpnII} recognition site, GATC. Also note that besides a pair of correct PNA-binding sites, two adenoviral DNA sites partially complementary to PNA 3 are located at 15.3 and 32.9 kb. They have the identical sequence, [G]TGATCAA, which is end-mismatched relative to the pcPNA binding site (the mismatched nucleobase is in brackets). These are the only singly-mismatched PNA-binding sites on adenoviral DNA that include the correct methylation/restriction site.

Figure 3, lane 2 shows that unprotected/unmethylated adenoviral DNA is digested by R\textit{DpnII} into many short fragments. When this restriction enzyme is used after binding of 1 \(\mu\text{M}\) PNA 3 and subsequent methylation with M\textit{dam}, only a few large fragments are observed (Figure 3, lane 5). The lengths of the resulting DNA fragments (4, 6.5, 11, 15 and 25 kb) correspond to the PNA-assisted protection of three
sites in adenoviral DNA. Thus, as expected, full protection of the two correct PNA-binding sites plus partial protection of one of the end-mismatched sites (located at 15.3 kb) was observed. Although both end-mismatched PNA-binding sites are sequence-identical to each other, only one of them was evidently protected from methylation by PNA 3. The reason for this is presently unknown and requires further studies.

Lower concentrations of PNA 3 resulted in incomplete digestion of viral DNA due to incomplete binding of PNA and, as a result, lack of full protection of PNA-binding sites from enzymatic methylation (not shown). Interestingly, three-fold increase of the PNA concentration also yielded non-digested DNA along with a pair of longer, about 29 and 32 kb, fragments again due to incomplete PNA binding (Figure 3, lane 4). Note that at a higher PNA concentration 11 and 15 kb fragments corresponding to mismatched PNA binding were observed with much lower yield. We ascribe this effect to self-inhibition of the pcPNA invasion into dsDNA due to the polycationic character of PNA 3 (30). At elevated PNA concentrations, positively charged PNA stabilizes the DNA double helix against PNA invasion, as other cations do (29).

Based on these data, we therefore conclude that use of pcPNAs within the framework of the PARC strategy makes it possible to convert 4-bp frequent DNA cutter into 7-8 bp rare cutters. They both recognize and cut the dsDNA sites with mixed sequence of all four nucleobases. Mention that thus obtained PARC-based cutters have recognition specificities, TGATCAA and TTGATCAA, previously being unavailable.

**Example 3: Rare Cut of Human DNA**

Figure 4 demonstrates the results of the ‘blind’ PARC treatment of the YAC-cloned human DNA without prior knowledge of its sequence. In this example, 1.8-Mb-long YAC from yeast strain 680_C_10 (CEPH human YAC library) was chosen that carries a large insert from human chromosome 16. Both bis-PNA (PNA 1) and pcPNA (PNA 3) were independently tested here.
Figure 4. PFGE analysis of the PARC cleavage of the YAC-cloned human DNA, as visualized by staining with EtBr (left lanes) or by blot hybridization with human Alu repeats (right lanes). (A) PARC cleavage with the use of bis-PNA 1. C: the intact control showing YAC (gray arrow) and the most of yeast chromosomes; R: the R.HpaII treatment only; P: the PARC treatment; black arrows: the PARC-generated fragments. (B) PARC cleavage with the use of 3 μM pcPNA 3. Controls C1 and C2 correspond respectively to the initial sample and the sample with all incubations required in the PARC protocol but without any treatment with PNA and enzymes. These previously unpublished experiments were performed jointly with K. I. Izvolsky and A. N. Lebedev.
Figure 4A shows that using the PNA 1/M.\textit{HpaII}/R.\textit{HpaII} combination, the YAC under study is rarely cut into four major fragments with lengths from $\sim$100 to $\sim$800 kb. Thus, the PARC combination used in this experiment works as a 9-bp rare cutter. Figure 4B shows that use of the PNA 3/M.\textit{dam}/R.\textit{DpnII} combination yields dozens of unresolved YAC fragments within the 50-90 kb size range, as expected for a 8-bp rare cutter. This digest is compatible with that of the natural 8-bp cutter, R.\textit{PmeI}, whose recognition site, GTTTAAAC, has the same base composition as the pcPNA-based PARC cutter. Note that R.\textit{DpnII} alone cleaves YAC into a number of very short ($\leq$1 kb), fast moving fragments that leave the gel during the PFGE analysis of larger fragments and hence are not detected in Figure 4B.

We may therefore conclude that by choosing an appropriate combination of bis-PNA and/or pcPNA with some pair of methylation/restriction enzymes, the recognition specificity of PARC cutters can be adjusted in a desired way to be used in a rare fragmentation of genomic DNA with unknown sequence.

**Discussion**

The data presented in this chapter clearly demonstrate the promising PNA potential to provide the genomic researchers with essentially unlimited number of DNA rare cutters featuring 8-12 bp recognition specificities. Consequently, the PNA-assisted DNA rare fragmentation could be done both in a directed way using the known DNA sequence and absolutely ‘blindly’ without any prior sequence information, as our examples show. Note in this connection that another quite popular method of DNA rare cleavage with similar to PARC methodology, RARE (most recently reviewed in ref. 43), can cut DNA only at sites with known sequences.

One more drawback of RARE, the non-specific background cleavage, is due to both the relatively low specificity of RecA-assisted hybridization (44, 45) and the presence of magnesium-activated methylase-contaminating nucleases (46). Since the PARC method does not require magnesium cations at the PNA-targeting/methylation steps
and because the formation of PNA-DNA complexes is a highly sequence-specific process (23-30), PARC can become the method of choice in some applications.

Finally note that the PARC method, as any ‘Achilles’ heel’-based approach, leaves the resulting fragmented genome thoroughly methylated, which may impose some limitations on the further use of the PARC products. As an option, another PNA-based, methylation-free method can be used for DNA rare cut in case non-methylated products are required. This alternative method involves the use of PNA openers (47; also see the next chapter of this book) and endonucleases specific to single-stranded DNA, e.g. S1 or mung bean nucleases, which site-specifically attack the DNA site locally exposed by PNA openers (48, 49).

**Acknowledgements**

We thank Peter E. Nielsen, Michael Egholm and James M. Coull for the continuous providing our research group with PNAs. We appreciate all colleagues contributed to these studies. We are also grateful to Boston University and the National Institutes of Health for financial support.

**References**


