Pergamon

Bioorganic & Medicinal Chemistry Letters 11 (2001) 2437–2440

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Solid-Phase, Single Nucleotide Primer Extension of DNA/RNA Hybrids by Reverse Transcriptases

Michael C. Pirrung,* Janice D. Worden, Joanne P. Labriola, Michael P. Montague-Smith and Laura J. Weislo

Department of Chemistry, Levine Science Research Center, Duke University, Durham, NC 27708-0317, USA

Received 18 April 2001; accepted 26 June 2001

Abstract—A method for RNA analysis based on primer extension by reverse transcriptase is described.

The use of arrays of nucleic acids ('DNA chips') is revolutionizing many aspects of genetic analysis,[1–3] diagnostics,[4] drug discovery,[5] and basic structural biology.[6] We are developing a novel, high-fidelity method for nucleic acid analysis that involves the polymerase-based, solid-phase extension of primer–template complexes with labeled dideoxynucleotide terminators (Fig. 1). The method is Arrayed Primer Extension (APEX)[7,8] when applied to an array of primers immobilized on a DNA microchip. Primer extension methods have been earlier applied to the analysis of single-nucleotide polymorphisms (SNPs) in genomic DNA. Genetic Bit Analysis (GBA)[9] attaches primers to microtiter plates for SNP genotyping. The mini-sequencing method[10–12] involves attachment of templates to the support. Enzymatic addition of radiolabeled terminator nucleotide triphosphates to hybridized primers readily identifies heterozygous, homozygous wild-type, and homozygous mutant alleles, and can distinguish genotypes for nine disease mutations with an order of magnitude greater power than hybridization with allele-specific probes.[13] Thus, compared to hybridization-based analysis, primer extension methods offer a high signal-to-noise ratio and consequent high fidelity. This is because APEX uses a pre-equilibrium step followed by an essentially irreversible addition of a labeled dideoxyoligonucleotide terminator (Fig. 2). Both steps are dependent on the match between primer and template. DNA polymerases require a perfect hybrid near the incorporation site (3′ end of the primer); only mismatches distal to the 3′ end can lead to extension and thereby contribute to background. The covalent attachment of the fluorescent tag to primers in APEX permits very stringent washing steps, and no labeling of analyte nucleic acid is required.

As powerful as these methods are in detecting mutations at the DNA level, the diagnostic utility of DNA sequence information is limited in that it is focused on sequences that are eventually translated into proteins: genes. Diagnostic methods focused on genomic DNA do not take into account effects that transcriptional controls, RNA processing (splicing, editing, etc.), or translational controls may have on the transfer of information from DNA to protein. While details of expression controls are contained within genomic sequences, efforts to elucidate their complex interacting

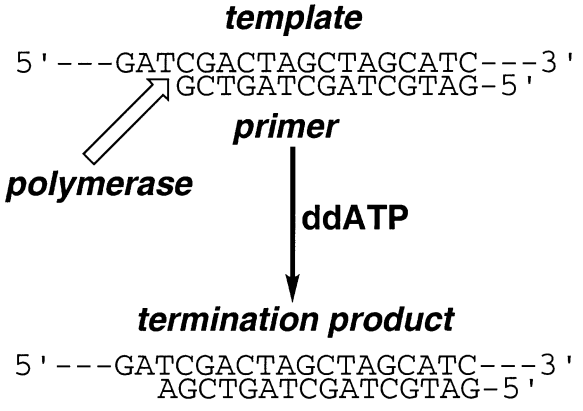

Figure 1. The APEX method involves single-base extension of primer–template complexes to give a termination product. One primer is required for each site in the template to be interrogated.

*Corresponding author. Fax: +1-919-660-1591; e-mail: pirrung@chem.duke.edu

0960-894X/01/$ - see front matter
PII: S0960-894X(01)00465-6

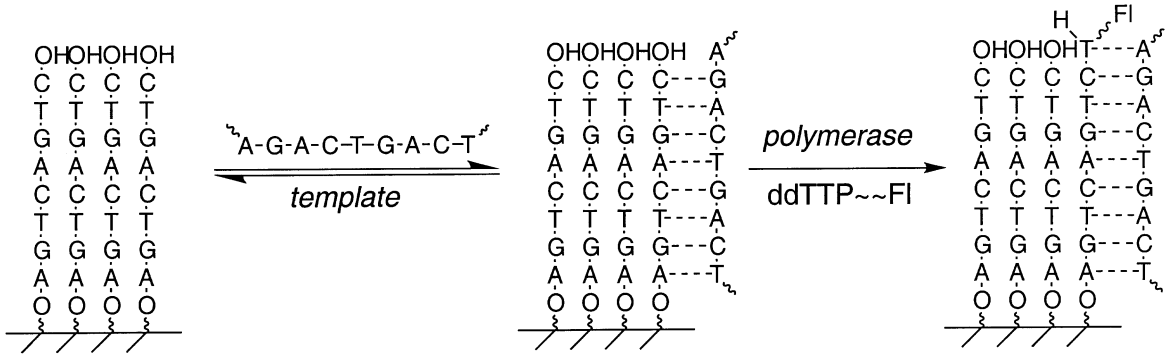

Figure 2. APEX involves a pre-equilibrium step driven by the thermodynamics of duplex formation followed by an enzyme-catalyzed, irreversible addition of a terminator. The selectivity in both steps is sensitive to mismatches.

mechanisms are ongoing and long term. Available DNA-based methods for detecting genetic disorders do not account for the possibility that whether a particular allele relates to disease can be more a function of its temporally and spatially proper expression than its existence. Consequently, many genetic disorders are not amenable to diagnostic methods based upon DNA sequence information alone.

For the foregoing reasons, RNA analysis may be equally relevant to diagnosis of genetic disease. Several examples of human disease-related expression defects are known. Improper expression of adenosine deaminase[14] or lymphokine (interleukins)[15] mRNA is associated with forms of combined immunodeficiency. Incorrect expression of epidermal growth factor receptor has been implicated in several human cancers.[16] Technology for detecting expression errors related to a genetic disorder at one of the downstream control points (transcription or translation) would allow diagnosis of such problems even when they may not be amenable to DNA sequence-based methods. The detection of improperly expressed coding regions of the genome must therefore be directed toward the gene products, mRNA or actual proteins, rather than the DNA sequence itself. Methods are needed to detect a particular mRNA expressed within cells from the total mRNA population in a time-efficient and practical manner. In practice, the mRNA sequence would be one associated with a particular cell type and/or expression-related disorder. Whether the mRNA is present, at what level, and in what form (e.g., splicing mutants) is the information sought. DNA chips have been used extensively for the study of RNA, particularly in gene expression analysis,[17,18] but as is the case for mutation detection, RNA-based primer extension may be a much more powerful and reliable method for analysis of RNA. RNA-based arrayed primer extension would effectively permit thousands of simultaneous northern blots.

As a prelude to APEX RNA analysis, we wished to examine in simple solid-phase assays the specificity of primer extensions on DNA/RNA hybrids catalyzed by reverse transcriptase, paralleling the early development of APEX with DNA/DNA hybrids.[19] We unexpectedly discovered RTs are indiscriminate of some mismatches.

The eukaryotic RNA template for these initial experiments in primer extension was obtained from a recombinant yeast strain inducible for overproduction of the mRNA for aminocyclopropanecarboxylic acid (ACC) synthase. This clone was created by engineering the cDNA of the winter squash ACC synthase (pCMW33)[20] using PCR and subcloning into the yeast expression vector pYES2. This vector allows expression of genes cloned downstream from the GAL1 promoter when cells are grown with galactose. Several features were introduced into the amplicon to promote expression.[21] The C at −3 from the AUG start codon of pCMW33 was changed in the 5′ PCR primer to an A. Translation of yeast messages is maximal with A at this position.[22] The 3′ primer was designed to amplify only the first 438 codons of the gene, producing a post-translationally processed form of the enzyme with higher activity.[23] The 1350 bp amplified sequence containing the ACC synthase gene is flanked by *Eco*RI sites for subcloning into pYES2. The resulting plasmid, named pWDE12, was introduced into the yeast strain INVSc1, and cells containing the plasmid were selected by growth on synthetic media minus uracil. Several colonies containing pWDE12 were chosen and screened for expression of ACC synthase enzymatic activity by measuring ACC secretion into the growth medium (by bleach oxidation to ethylene). This method was also used to monitor induction.

The best pWDE12/INVSc1 strain was grown in a non-inducing medium, diluted 1:41 with inducing medium (2% galactose), grown at 37 °C until the culture reached its exponential growth phase (>24 h), and total RNA was isolated, giving ~3 mg RNA/L culture. Polyadenylated mRNA (380 μg/L culture) was isolated from total RNA by affinity chromatography on a polyT column. RNA was isolated from non-induced cultures as a control.

A site downstream from the cDNA insertion site in pYES2 was targeted for mRNA primer extension based on a database search showing its uniqueness in the yeast genome. This inducible mRNA template is translated to a protein with readily assessable enzymatic activity. DNA primers were designed for primer extension experiments (Tables 1 and 2) with a 5′-biotin group for attachment through an avidin link to agarose beads.

The bead represents a surface environment, but enables the extended primer to be removed from the support by avidin denaturation so the extension product can be analyzed by gel electrophoresis. Both sites are in the 3′-untranslated region of the mRNA. The overall scheme for the assay is given in Figure 3. The longer primer set included two perfect matches to target sites, single mismatches at −3 and −12 (relative to the incorporation site; the nucleotide at the 3′-end is defined as 0), and a double mismatch at −3, −4 (**1**–**6**). The shorter primer set included a perfect match to the target site, single mismatches at the 0, −1, −2, −5 and −6 sites (**7**–**13**), and a double mismatch at −3, −4 (not shown). Random, multiply-mismatched primers were used as negative controls. Primers were prepared on an automated synthesizer and purified by reverse-phase HPLC (C18).

The first enzyme studied for extension of primer–template complexes (Fig. 4) is the common AMV reverse transcriptase. Following heating of the mRNA (4 μg/μL) fraction to remove secondary structure, hybridization was performed at low to medium stringency (42 °C for heptadecanucleotides, 37 °C for decanucleotides, no salt/formamide). Primer–template complexes were incubated with AMV reverse transcriptase (RT) and α-^{35}S-ddATP (42 °C, 2 h). The products were heat-denatured from the support (90 °C), separated by 6 M urea denaturing PAGE, and detected by autoradiography. An $n+1$ band on the gel signified hybridization and extension of the primer by rt.

Fully complementary primers, and those with single mismatches at sites beyond −4 from the 3′-end, showed extension products. Unexpectedly, 0 position mismatches also result in extension products. Primer–template complexes with single mismatches at −1, −2, −3, double mismatches, or multiple mismatches do not promote extension. No extension was observed with control mRNA. A second, thermostable RT was also studied. The *Tth* DNA polymerase has effective and stable RT activity in the presence of Mn^{2+},[24] which has enabled its use in RT–PCR protocols. It was used with a selected set of the decanucleotide primers (Table 2). In general, results were comparable to those with AMV RT. Surprisingly, this enzyme is not sensitive to mismatches at the 3′-end of the primer, while it is sensitive to mismatches further toward the 5′-end of the primer.

Table 1. 5′-Biotinylated heptadecanucleotides used in solid-phase primer extension reactions with ACC synthase mRNA target sites 5′-∼∼UG GAT CUA CGG AGG CUA∼∼-3′ and 5′-∼∼UGC GAC CGG AGC GAG CGA∼∼-3′. Mismatches underlined

Oligo	Sequence	Calcd T_m (°C)	AMV extension
1	5′-TAG CCT CCG TAG ATC C	50	+
2	5′-TCG CTC GCT CCG GTC GC	60	+
3	5′-TCG CTC GCT CCG GAC GC	53	−
4	5′-TCG CAC GCT CCG G$\overline{\text{T}}$C GC	53	+
5	5′-TCG C$\overline{\text{T}}$C GCT CCG C$\underline{\text{A}}$C GC	45	−
6	5′-AAA AAA AAG GGG $\overline{\text{G}}\underline{\text{G}}$G GG	—	−

Table 2. 5′-Biotinylated decanucleotides used in solid-phase primer extension reactions with ACC synthase mRNA target site 5′-∼∼UG GAT CUA CGG∼∼-3′. Mismatches underlined

Oligo	Sequence	Calcd T_m (°C)	AMV extension	*Tth* extension
7	5′-CCG TAG ATC C	42	+	+
8	5′-CCG TAG ATC $\underline{\text{G}}$	41	+	+
9	5′-CCG TAG ATG $\overline{\text{C}}$	41	−	nd
10	5′-CCG TAG A$\underline{\text{A}}\overline{\text{C}}$ C	41	−	−
11	5′-CCG TTG A$\overline{\text{T}}$C C	41	+	nd
12	5′-CCG T$\overline{\text{A}}$G ATC $\underline{\text{T}}$	41	+	+
13	5′-CCG $\underline{\text{A}}$AG ATC $\overline{\text{C}}$	41	+	nd

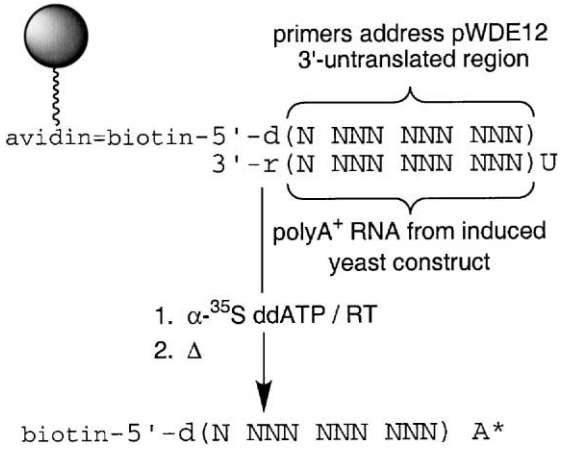

Figure 4. Solid-phase APEX with RNA templates and reverse transcriptase.

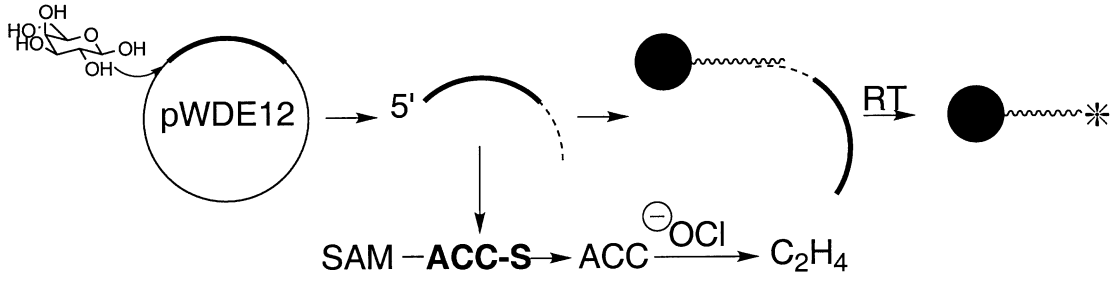

Figure 3. Induction of transcription of the ACC-S cDNA harbored on pWDE12 is induced by galactose. The mRNA of interest is produced bearing a 3′-untranslated region that is targeted by APEX primers bound to agarose beads. Reverse transcriptase adds labeled terminator nucleotide, which is analyzed after denaturation and PAGE. Expression of the target mRNA is monitored through its translation product, the ACC synthase protein, which converts SAM to ACC. The bleach oxidation of ACC gives ethylene, which is monitored by gas chromatography.

The results reported here show that APEX is a viable method for RNA analysis. It is interesting that reverse transcriptase seems much less sensitive to mismatches in the primer–template complex than we have earlier observed for DNA polymerases.[7] The type of mismatch and its position influence extension of the mismatched primer–template complex. Known thermodynamics of mismatches in DNA–RNA duplexes are helpful in analyzing these data.[25] The most thermodynamically stable mispair (ΔT_m 4 °C) is dT–rG (wobble), and we observe that it is extended (primer **12**). A dA–rA mismatch is one of the less stable mispairs (ΔT_m 12–17 °C), and is not extended (primer **10**). We studied two dG–rG mismatches (ΔT_m 3–8 °C) at adjacent positions in the primer. Interestingly, one of these is extended and one is not (primers **8** and **9**). Overall, the results with the mesophilic and thermophilic enzymes show reasonable agreement.

These results address mismatches on the DNA strand of the primer–template complex because those bases are the most easily varied using synthetic nucleic acid. The influence of mismatches on the RNA strand is most easily addressed using synthetic RNA templates, rather than native eukaryotic RNA, as studied here. For full RNA analysis, improvements will center on reduced mishybridization/extension from unintended RNA templates. Increased stringency in the hybridization/extension step will be facilitated by use of novel RTs. Longer, more specific primers should also be helpful.

Acknowledgements

Financial support was provided by NIH GM46720 and an NIH NRSA Fellowship (MPM-S).

References and Notes

1. Pease, A. C.; Solas, D.; Sullivan, E. J.; Cronin, M. T.; Holmes, C. P.; Fodor, S. P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5022.
2. Southern, E. M. *Trends Genet*. **1996**, *12*, 110.
3. Gerhold, D.; Rushmore, T.; Caskey, C. T. *Trends Biochem. Sci*. **1999**, *24*, 168.
4. Whitcombe, D.; Newton, C. R.; Little, S. *Curr. Opin. Biotechnol*. **1998**, *9*, 602.
5. Debouck, C.; Goodfellow, P. N. *Nat. Genet*. **1999**, *21*, 48.
6. Sohail, M.; Akhtar, S.; Southern, E. M. *RNA* **1999**, *5*, 646.
7. Pirrung, M. C.; Connors, R. V.; Montague-Smith, M. P.; Odenbaugh, A. L.; Walcott, N. G.; Tollett, J. J. *J. Am. Chem. Soc*. **2000**, *122*, 1873. Pirrung, M. C.; Davis, J. D.; Odenbaugh, A. L. *Langmuir* **2000**, *16*, 2185.
8. Tollett, J. J.; Kurg, A.; Shah, A.; Roa, B. B.; Richards, C. S.; Nye, S. H.; Pirrung, M.; Metspalu, A.; Shumaker, J. M. *Am. J. Hum. Gen*. **1997**, *61* (Suppl. S), 1322. Metspalu, A.; Saulep, H.; Tonisson, N.; Soo, V.; Kurg, A.; Shumaker, J. M. *Am. J. Hum. Gen*. **1997**, *61* (Suppl. S), 1301. Kurg, A.; Tollett, J. J.; Shah, A.; Roa, B. B.; Richards, C. S.; Nye, S. H.; Pirrung, M.; Metspalu, A.; Shumaker, J. M. *Am. J. Hum. Gen*. **1997**, *61* (Suppl. S), 68.
9. Nikiforov, T. T.; Rendle, R. B.; Goelet, P.; Rogers, Y.-H.; Kotewicz, M. L.; Anderson, S.; Trainor, G. L.; Knapp, M. R. *Nucl. Acids Res*. **1994**, *22*, 4167.
10. Syvänen, A. C.; Ikonen, E.; Manninen, T.; Bengtstrom, M.; Söderlund, H.; Aula, P.; Peltonen, L. *Genomics* **1992**, *12*, 590.
11. Syvänen, A. C.; Aalto-Setälä, K.; Harju, L.; Kontula, K.; Söderlund, H. *Genomics* **1990**, *8*, 684.
12. Syvänen, A. C.; Landegren, U. *Hum. Mutat*. **1994**, *3*, 172.
13. Pastinen, T.; Kurg, A.; Metspalu, A.; Peltonen, L.; Syvänen, A. C. *Genome Res*. **1997**, *7*, 606.
14. Daddona, P. E.; Schwach, D. S.; Kelley, W. N.; Argos, P.; Markham, A. F.; Orkin, S. H. *J. Biol. Chem*. **1984**, *259*, 12101.
15. Geha, R. S.; Castigili, E.; Chatila, T. *Clin. Immun. Immunopath*. **1991**, *61*, S16.
16. Gullick, W. J. *Brit. Med. Bull*. **1991**, *47*, 87.
17. Lockhart, D. J.; Winzeler, E. A. *Nature* **2000**, *405*, 827. Harrington, C. A.; Rosenow, C.; Retief, J. *Curr. Opin. Microbiol*. **2000**, *3*, 285.
18. Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, *270*, 467.
19. Shumaker, J. M.; Metspalu, A.; Caskey, C. T. *Hum. Mutat*. **1996**, *7*, 346. Shumaker, J. M.; Metspalu, A.; Caskey, C. T. In *Lab. Protoc. Mutat. Detect*; Landegren, U., Ed.; Oxford University Press: Oxford, 1996; p 93.
20. Nakajima, N.; Mori, H.; Yamazaki, K.; Imaseki, H. *Plant Cell Physiol*. **1990**, *31*, 1021.
21. Weislo, L. J.; Pirrung, M. C., *Biochem. J*. Submitted for publication.
22. Moir, D. T.; Davidow, L. S. *Methods Enzymol*. **1991**, *194*, 491.
23. Mori, H.; Nakagawa, N.; Ono, T.; Imaseki, H. In *Cellular and Molecular Aspects of the Plant Hormone Ethylene*; Pech, J. C., Latché, A., Balagué, C., Eds.; Kluwer: Dordrecht, 1993; p 1.
24. Rüttiman, C.; Cotoras, M.; Zaldivar, J.; Vicuna, R. *Eur. J. Biochem*. **1985**, *149*, 41. Myers, T. W.; Gelfand, D. H. *Biochemistry* **1991**, *30*, 7761.
25. Sugimoto, N.; Nakano, M.; Nakano, S.-I. *Biochemistry* **2000**, *39*, 11270.