T7 DNA polymerase in automated dideoxy sequencing

Tom Kristensen¹, Hartmut Voss, Christian Schwager, Josef Stegemann, Brian Sproat and Wilhelm Ansorge*

European Molecular Biology Laboratory, Heidelberg, FRG and ¹Research Institute for Internal Medicine, University of Oslo, Norway

Received March 3, 1988; Accepted March 17, 1988

ABSTRACT

T7 DNA polymerase with chemically inactivated 3'-5' exo-nuclease activity, as well as unmodified T7 DNA polymerase, were used for sequencing by the dideoxy method in an automated system with fluorescence labelled primer and on-line detection of laser-excited reaction products. An analysis of signal intensity variations in the C track revealed that low C signals were usually preceded by a T in the sequence. This effect was modified by surrounding nucleotides. Signal intensities were more uniform with T7 polymerase than with the Klenow fragment of DNA polymerase I. Some sequences ambiguous with the Klenow enzyme could easily be evaluated with the T7 enzyme. One sequence could only be read by the unmodified T7 polymerase, while both the Klenow fragment and the chemically modified T7 enzyme gave uninterpretable data.

INTRODUCTION

DNA sequencing using DNA polymerases with DNA cloned in M13 vectors as templates, complementary oligodeoxyribonucleotides as primers, and dideoxyribonucleoside triphosphates (ddNTPs) as chain terminators has become the method of choice for many sequencing projects. Offering a simple system for DNA labelling by the incorporation of labelled nucleotides or labelled primers, it forms the basis for automated sequence analysis with on-line identification of fluorescently labelled bands and computer storage of the sequence (1-3).

The large variation in signal intensity often observed in dideoxy sequencing puts high demands both on detection systems and the sequence interpretation in automatic sequencing. The main factor responsible for the signal intensity variations observed in sequencing reactions appears to be the DNA polymerase employed. The Klenow fragment of the DNA polymerase I from E. coli, which is the standard enzyme used in most applications, has
several shortcomings. The frequency of incorporation of a ddNTP with concurrent chain termination at a given position appears to be strongly dependent on the surrounding sequence; the resulting large ratio between strong and weak signals in one lane may limit resolution of multiple peaks. In addition, the Klenow fragment is sensitive to secondary structure in the template, leading occasionally to chain termination without incorporation of ddNMPs, resulting in sequence ambiguity. Consequently, enzymes such as reverse transcriptase and other viral DNA polymerases have been used for analysis of sequences difficult to read with the Klenow enzyme (4). But usually they have also peculiarities leading to difficulties in the interpretation of the sequence.

Recently a chemically modified T7 DNA polymerase was made commercially available ("Sequenase"), with claims of less signal variation and lower sensitivity to template secondary structure than the Klenow enzyme (5). The modification involves inactivation of the very active 3'-5' exonuclease activity of the polymerase by locally produced free radicals (6). In the present work we show that this enzyme, as well as an unmodified T7 DNA polymerase can be used with advantage in an automated DNA sequencing system with a fluorescent primer and on-line detection of laser-excited bands. We have evaluated their performance with special emphasis on sequences where the performance of the Klenow fragment is insufficient.

MATERIALS AND METHODS

DNA polymerase I large fragment (Klenow fragment) was from BRL, deoxy- and dideoxynucleotides from Boehringer, and Sequenase sequencing kit, including a T7 DNA polymerase with a chemically inactivated 3'-5' exonuclease, from USB Corporation, Cleveland, Ohio, USA. Cloned, unmodified T7 DNA polymerase was obtained from Pharmacia PL Fine Chemicals.

Preparation of template DNA was performed as described previously (7), by precipitation of M13 phage particles from culture supernatants with acetic acid and filtration on glass fiber filters, followed by protein removal by treatment with 4 M sodium perchlorate and elution of DNA with a low ionic strength buffer. In the experiments described we have used the cloning
vector M13 mpl8 as well as various inserts cloned in M13 mpl8 and mpl9 as template DNA.

The conditions used in sequencing reactions with the Klenow fragment and a fluorescent primer were essentially as described previously (3), with equal concentrations of all four dNTPs, and a reduced ddNTP/dNTP ratio as compared to radioactive labelling. This modification is necessary in end labelling procedures since the amount of incorporated label will be proportional to the molarity rather than to the mass of the different products, so that a higher amount of the larger size products is needed to get a signal sufficiently strong for their detection. As for the two T7 DNA polymerases, the conditions recommended in the Sequenase manual were modified for the same reason. We found that equal amounts of each dNTP, at concentrations 5 times higher than recommended in the manual during the labelling reaction gave the desired result when combined with stop reaction conditions as in the manual.

Fluorescent reaction products were analysed by polyacrylamide gel electrophoresis and on-line detection of laser-induced fluorescence in an automated system as described previously (2,3,7).

Relative signal strengths were estimated by comparing the peak intensity of a given signal with the mean peak intensity of the 5 neighbour signals on each side in the same lane.

RESULTS

The Klenow fragment of DNA polymerase I from E. coli is the best characterized and most widely used enzyme for chain termination sequencing. The three-dimensional structure of the protein has been determined and various activities assigned to distinct domains (8). More is known about how template sequence affects its catalytic properties than for any other polymerase used for sequencing purposes. Part of this knowledge has been condensed in a few "sequencing rules" (9, 10) which emphasizes the behavior of the Klenow enzyme on runs of two or more repeats of a given nucleotide:

1. The second C in a run of Cs will always be the most intense.
Fig. 1. Analysis of runs of As, Cs, Gs and Ts with the Klenow fragment and T7 polymerase. Top panels: Klenow fragment, lower panels: modified T7 polymerase. Comparable data for the two enzymes are always from the same sequence.

2. The second G in a run of Gs will be the most intense only if preceded by a T.

3. The first A in a run of As will usually be the most intense if preceded by a G or a C. If preceded by a T, it will still be stronger, but the difference will be less pronounced.

As a first step in the comparison of the enzymes we have evaluated their behavior when reading sequences such as those described in the three "rules" above. Fig. 1 shows a comparison of the Klenow enzyme and the T7 polymerase in resolution of multiple neighbouring peaks. The peak intensities were significantly more uniform with the T7 polymerase than with the Klenow fragment. The largest deviations from ideal behaviour of the T7 polymerase in
AATCCAC 0.45  CATCCCT 0.63  CGTCTCT 0.37  GGGCTCC 0.53
AATCCAT 0.54  CATCTAC 0.76  CTTCAAG 0.33  GGTCATG 0.57
AATCGTC 0.42  CATCTGC 0.48  CTTCGCT 0.54  GGTCATA 0.50
ACACGTC 0.71  CCTCAAG 0.66  CTTCTGC 0.63  TAGTCGG 0.73
AGCCGGA 0.58  CCTCATC 0.71  CTTCTTC 0.38  TCTCAAC 0.63
AGTCCCTC 0.69  CCTCCGC 0.67  GCGCTTT 0.55  TCTCGGA 0.79
ATTCCAC 0.58  CCTCGTT 0.50  GCTCCAG 0.76  TCTCTTG 0.54
ATTGTCTG 0.48  CGTCCGT 0.52  GCTCTGC 0.71  GTGACCC 0.58
CATCAAG 0.61  CGTCCCTG 0.38  GCTCTGA 0.51  TGTCCGG 0.65
CATCCCA 0.71  CGTCGTC 0.35  GGGCATC 0.60  TTTCCGT 0.61
CATCCCT 0.37

<table>
<thead>
<tr>
<th>Relative position to the C</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>T</td>
</tr>
</tbody>
</table>

Table 1. Sequences which give reduced intensities in the C track when read by T7 polymerase. Three bases on each side of the C are shown. The values given are the intensity of the C peak in question, relative to the mean intensity of the five closest Cs on each side in the sequence. The frequencies of the four bases at each position are summarized in the bottom table.

these tests was observed in some runs of Cs when preceded by T, but in no case was the ratio between high and low peak higher than 3.5. With the Klenow fragment the strong peaks may be more than ten times higher than the weak ones.

The variations in peak intensities obtained with the T7 polymerase seemed to be most frequent in the C track. An analysis of the sequence context of 41 Cs, each with a signal intensity of less than 80% of the mean intensity of the surrounding ten Cs, is shown in Table 1. The results demonstrate that low Cs were almost always (in 36 out of 41 cases) preceded by a T. A search in about 1500 sequenced base pairs revealed only 8 examples of TC sequences where the C was not significantly lower than its neighbours (Table 2). Interestingly, in all 8 cases the C is preceded by one of the sequences GAT, GCT or GTT. Of these, only GCT is present in the same position in Table 1.

Evaluation of examples of unusually low and high peaks from reactions performed with T7 polymerase showed that low Ts were
Table 2. T7 polymerase does not always give reduced C intensities in the sequence -TC-. A collection of sequences containing -TC without significant reduction of the C intensity is displayed, consisting of the C and three bases on each side. The values given are the intensity of the C relative to the mean intensity of the five closest Cs on each side in the sequence.

often preceded by a G, while high Ts were often preceded by a double C. All in all, however, these differences in peak intensities were smaller, less consistent, and less frequent than those observed in the C track, and much smaller than the variability obtained with the Klenow fragment.

Furthermore, we analysed sequences where one or more of the enzymes investigated did not read the sequence properly. One such example is shown in Fig. 2 and involves an A-C-rich region in M13 mp18. Various batches of the Klenow fragment had reproducible problems with this region, as shown in the figure. When T7 polymerase was tested, the sequence could be read without ambiguities.

In the sequences analysed above, we saw no difference between the T7 DNA polymerase with the chemically inactivated exonuclease activity and the unmodified enzyme. Differences did, however, appear when we examined examples of "pile-ups", where the polymerase is stopped by secondary structures in the template, resulting in signals in all four lanes on the gel. One example of such reproducible pile-ups obtained with the Klenow fragment and with the chemically modified T7 DNA polymerase, together with the unambiguous sequence signals obtained with the unmodified polymerase, are shown in Fig. 3. Whether the increased ability of the unmodified enzyme to read through secondary structures is due to the nature of the alteration or to other differences in the enzyme preparations remains to be determined.
Fig. 2. Example of a sequence which can be resolved with T7 DNA polymerase, but not with the Klenow fragment. a, Klenow fragment; b, T7 DNA polymerase. The correct sequence is AAT TCC ACA CAA CAT ACGAGC C and was found in M13 mp18, 110 bases away from the primer.

DISCUSSION

Variations in signal intensities observed in dideoxy sequencing may be caused by factors having long-range or short-range effects. Long-range variations are usually due to reaction conditions, such as the nucleotide concentration or the ratio of deoxy- and dideoxyribonucleotides, and reflect the normal distribution in product sizes obtained. By a careful choice of reaction conditions the experimenter may manipulate the mean product size; the distribution of sizes around this mean will depend on several factors, such as the processivity of the polymerase used, and the primer/template/polymerase ratios.

On the other hand, short-range variations in signal intensity are mainly the result of sequence-dependent variations in the interaction between the polymerase and its substrates, both the primer-template complex and the nucleotides. Some sequences will give rise to secondary structures in the template, which will have more or less profound effects on the products formed and their corresponding signal intensity, depending also on the
Fig. 3. Difference in performance between two types of T7 DNA polymerase. Data obtained with a, the Klenow fragment; b, chemically modified T7 DNA polymerase; c, unmodified T7 DNA polymerase. The correct sequence is CAA ACC CCT ACC CCC TG.

type and quality of the polymerase used, as well as other factors (4,11). From Fig. 3 it is evident that whereas the Klenow fragment and one of the T7 polymerases were hindered by the secondary structure in the template to the extent that they were unable to read through it, the unmodified T7 polymerase from Pharmacia PL read through the structure without problems. In
other cases, short-range variations in the amounts of product formed must be due to other causes, such as direct neighbouring nucleotide effects on the different enzymatic activities of the polymerase used, i.e. the polymerase activity and the 3'-5' exonuclease activity. The "sequencing rules" derived for the Klenow fragment reflect some such interactions, and the present work describes some observations which reflect similar interactions between the T7 polymerase and specific template sequences. It appears that although some quite simple "rules" may be formulated for the T7 polymerase ("C will be weak after T -"), some complications appear ("- except in the sequence GATC, where the C will be normal or strong") indicating that the signal variations observed may be the effects of quite complex interactions between the polymerase and its substrates.

Tabor and Richardson (5) noted the presence of some unusually strong signals in sequencing reactions with the modified T7 polymerase (TCT, AAG, GCA, CCT, underlined bases give strong signals). We observed similar effects except for GCA (results not shown). These effects were, however, less consistent and pronounced than the effects described for the C track. The absence of a GCA effect in our system could probably be due to the use of end labelled primers, rather than radioactive dATP in a low concentration in the reaction mixture used in (5).

Despite the irregularities described here, the signal variations observed in sequencing reactions catalyzed by T7 DNA polymerase were definitely smaller than those obtained when the Klenow fragment was used. This indicates that the T7 polymerase may be a better choice for a sequencing enzyme than the Klenow fragment, especially when used in automated on-line signal detection and sequence extraction, where large variations in signal intensities may complicate the sequence interpretation.

It should also be noted that the increased use of automated sequencing systems with acquisition and storage of signal intensity data, as well as of sequence data, will give us a large amount of information on polymerase-template interactions as reflected in variations in signal intensity. This information will be an added bonus to the sequence information obtained, and should be useful for the continued development of automated procedures.
sequencing systems. It would allow the formulation of detailed "sequencing rules" for the various polymerases used in dideoxy sequencing; these rules could then be incorporated in the software for the sequence reading and simplify the interpretation of problem sequences. It should also be an important element in the mapping of polymerase-substrate interactions and reaction mechanisms which eventually may allow the engineering of polymerases with less substrate-derived product artefacts than those derived from natural sources.

ACKNOWLEDGEMENTS

Professor L. Philipson is thanked for support. T.K., who was a Fellow of the Norwegian Society for Fighting Cancer during the major part of this work, thanks EMBO for providing a short term fellowship.

*To whom correspondence should be addressed

REFERENCES