Overexpression of the Bacteriophage PRD1 DNA Polymerase

Guhung Jung

Department of Biology Education, Seoul National University, Seoul 151-724, KOREA

In order to overexpress bacteriophage PRD1 DNA polymerase in *E. coli* cells, the 2 kb HaeII fragment was isolated from phage genomic DNA. This fragment was then cloned into pEMBLex 3 expression vector. A specific 57 bp deletion was performed by using uracil containing ss DNA and oligonucleotide spanning each region to remove an unwanted non-coding region. After this deletion, the PRD1 DNA polymerase gene is totally under the control of the vector promoter and SD sequence. Upon heat induction, a protein with an apparent size of 68 kdal was overexpressed as an active PRD1 DNA polymerase. The expression of PRD1 DNA polymerase was about 1% of total *E. coli* protein.

KEY WORDS □ PRD1, DNA polymerase, overexpression.

A small lipid-containing bacteriophage PRD1 specifies its own DNA polymerase which utilizes terminal protein as a primer for DNA synthesis (11). The PRD1 DNA polymerase gene has been sequenced and its amino acid sequence deduced (6). This protein-primed DNA polymerase consists of 553 amino acid residues with a calculated molecular weight of 63,300 (6). Thus, it is the smallest DNA polymerase ever isolated from prokaryotic cells.

Sufficient quantities of homogeneous PRD1 DNA polymerase are required to perform detailed biochemical studies of the enzyme. The major difficulty in purifying this enzyme is its naturally low abundance. To overcome this problem, two approaches may be taken. One is to find a relatively abundant source. The other is to clone the DNA polymerase gene into an expression vector and overproduce the enzyme within a suitable host. Although, the natural abundance of a virus-encoded polymerase is high compared to DNA polymerase III of *E. coli* and mammalian polymerases, cloning of the PRD1 DNA polymerase gene into an expression vector will make isolation of PRD1 DNA polymerase much easier. This approach has been used successfully with the Klenow fragment of pol I (5). The large quantities of the protein generated in this way have facilitated subsequent biochemical and crystallographic studies (12). In order to overexpress PRD1 DNA polymerase in *E. coli* cells, the polymerase gene was cloned into pEMBLex3 expression vector.

MATERIALS AND METHODS

Bacterial, Phage and Plasmid strains

E. coli NM522 (*lac-proAB*), *thi,hsd* 5, *sup*E, [*F'*, *pro*AB *lac*IqZ M15] and *E. coli* RZ1032 [HfrKL16 PO/45 {*lys*A(61-61)}, *dut*1, *ung*1, *thi*1, *rel*A1] (7) were the strains used as the hosts for all recombinant plasmids. These were grown in 2X YT or YT broth. Bacterial strain *E. coli* NM522 is an *E. coli* 71/18 mutant which lacks a restriction system. *E. coli* 71/18 is *E. coli* K-12 (*lac-pro*) [*F'* *lac*Iq*lac*Z M15pro] *sup*E. The phagemid expression vector pEMBLex3 (16) was a generous gift of Dr. G. Cesareni of the European Molecular Biology Laboratory. The helper phage M13KO7 (19) was purchased from IBI.

DNA Preparation

Phage were grown on *S. typhimurium* LT2 in LB broth (containing kanamycin 50 μg/ml) at 37°C. The phage DNA was isolated by phenol extraction after digestion with proteinase K in the presence of 0.2% SDS (1). Plasmid DNA was prepared by the alkali lysis method (9). (Table 1).

DNA Sequencing

To confirm mutants after site-specific mutagenesis, sequenase (United States Biochemical, Cleveland, OH) was used for sequencing according to the manufacturer's recommendations.

Preparation of Uracil-containing ss-DNA

A culture of *E. coli* RZ1032 cells which contains recombinant pEMBLex3 in 2X YT supplemented

Table 1. *List of Plasmid*

Plasmid	Characteristics	Selectie Marker	Size (Kb)
pLM3	pBR322 derivative carrying PRD1 DNA polymerase and terminal protein genes.	Tet	7.3
pLM141	pBR322 derivative carrying PRD1 DNA polymerase gene and part o terminal protein gene.	Tet	6.8
pEMBLex3	Phagemid expression vector carrying pR promoter.	Lac, Amp	4.9
pEJ3	pEMBLex3 derivative carrying PRD1 DNA polymerase and terminal protein genes.	Amp	7.9
pEJ2.5	pEMBLex3 derivative carrying DNA polymerase gene and part of terminal protein gene	Amp	7.5
pEJ2	pEMBLex3 derivative carrying DNA polymerase gene and part of terminal protein gene.	Amp	6.9
pEJG	pEMBLex3 derivative carrying only DNA polymerase gene	Amp	6.9

with 0.001% thiamine, 150 μg/ml ampicillin and 0.25 μg/ml uridine was grown to OD_{600} of 0.08 at 30°C and incubated for 3 min at 37°C.

This culture was then infected with M13KO7 at a multiplicity of infection of 2-10 for 75 min at low rpm. Afterwards the cells were diluted to an OD_{600}<0.2 and kanamycin was added to a final concentration of 70 μg/ml. The culture was then grown for 20-24 hr at 30°C and 300 rpm. The cells were pelleted by centrifugation and supernatant was removed to a fresh tube. One quarter of the supernatant volume of 20% PEG and 3.5 M ammonium acetate were then added, placed on ice for 30 min. This viral pellet was isolated by centrifugation (12,000 g, 15 min). The superntant was then discarded and the tube was tamped dry on tissues. The pellet was then resuspended in 400 μl TE buffer by vortexing, then centrifuged 2 min to remove insolubles.

A 200μl of buffered phenol was added, the solution was vortexed for 2 min, and then centrifuged for 2 min. The aqueous layer was removed to a fresh tube and the extraction was repeated with phenol/chloroform until only a slight interface was visible. The single-stranded DNA was then precipitated by adding 0.5 volume of 7.5 M ammonium acetate and 2 volumes of 95% ethanol.

After the pellet dried, it was resuspended in TE buffer (10 mM Tris cl, 1 mM EDTA).

Site-specific Deletion of pEJ2

A region of 57 bp was deleted from the recombinant plasmid pEJ2 (Fig. 1). Th e specific 57 bp deletion was created by making uracil-containing ss-DNA and utilizing an oligo-nucleotide spanning the desired deletion. The uracil-containing ss-DNA pEJ2 DNA (1 g) and the 5′ phosphorylated oligonucleotide were mixed in equimolar ratio into a buffer containing 20 mM Tris-HCl pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl.

The mixture was then cooled gradually from 80° to 30°C over a period of 5 hr in a large waterbath. Synthesis of the second strand was done according to the method of Kunkel (7). A 29 mer oligonucleotide, 5′TGTACTAAGG-AGGTTTGATATGCCGCGCC3′, was synthe-sized chemically on an Applied Biosystems DNA synthesizer. The final 57 bp deletion removes a PstI site from the vector (Fig. 3). Plasmids containing the deletion were identified by digesting DNA isolates with Pst I and analyzing the products by agarose gel electrophoresis. DNA sequencing was used for confirmation of this deletion.

Analysis of Cloned Bacteriophage PRD1 Protein

A 1 ml of LB with ampicillin (100 μg/ml) was inoculated with an overnight culture of *E. coli* NM 522 containing a recombinant expression vector to a final dilution of 1:200. This 1 ml culture was induced at mid-log phase by increasing the temperature to 42°C for 4 hr. 100 or 200 μl of the cell culture was centrifuged (12,000 g, 3 min), and the cell pellet was resuspended and boiled in 20 μl of a gel loading buffer (50 mM Tris-HCl, pH 6.8, 10% glycerol, 5% 2-mercaptoethanol, 2% SDS and 0.05% bromophenol blue). This sample was loaded on a 10% SDS-polyacrylamide gel. The total cellular protein in the gel was visualized by staining with Coomassie Blue.

RESULT

Cloning of Bacteriophage PRD1 DNA Polymerase Gene into the Expression Vector pEMBLex3

Nucleotide sequence analysis revealed that the 3 kb PRD1 insert within recombinant plasmid pLM3 contains complete open reading frames for PRD1 DNA polymerase and terminal protein (4, 6). Therefore, the 3 kb Pst I fragment was isolated and inserted into the Pst I site of pEMBLex3 (Fig. 1). pEMBLex3 has a single Pst I site located downstream from a lambda pR promoter (16).

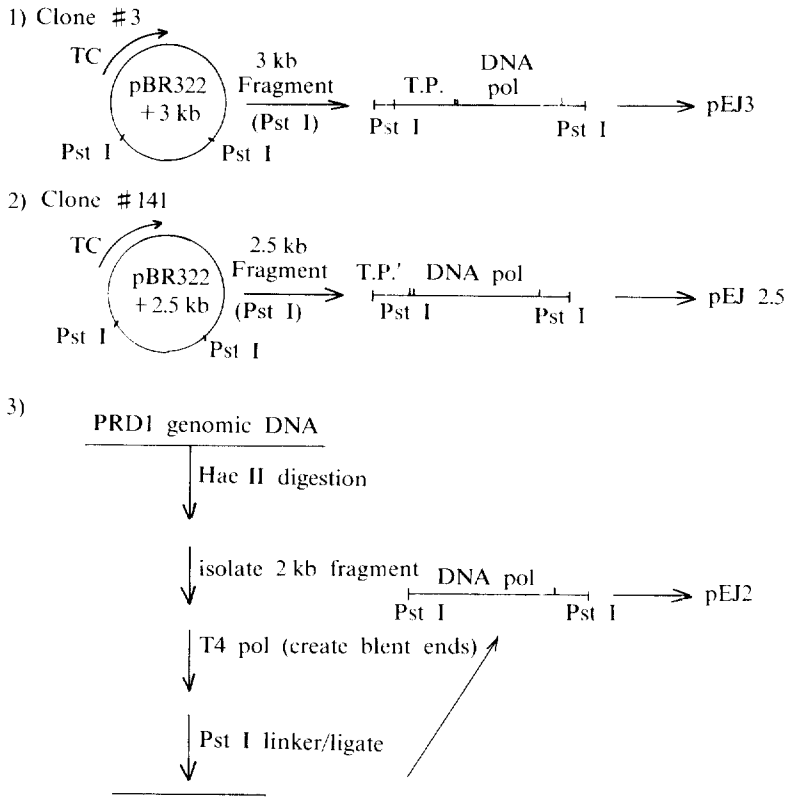

Fig. 1. *Construction of recombinant plasmids; pEJ3, pEJ2.5 and pEJ2.*

Transcription of a gene inserted at this site will be under the control of the pR promoter of the vector. The PRD1 terminal protein and the DNA polymerase gene constitute an operon (6). Transcription initiates upstream of the terminal protein gene and proceeds through both genes and terminates at some point downstream of the DNA polymerase gene (15). When cloned in the correct orientation, the 3 kb Pst I fragment from pLM3 will synthesize PRD1 terminal protein and DNA polymerase using both the vector and viral promoter (16). Plasmid isolates containing 3 kb insert in the correct orientation were identified by restriction enzyme analysis. This plasmid was designated pEJ3. Another plasmid which carried the insert in the reverse orientation was designated pEJ3R. The orientation of the 3 kb insert was further verified by DNA sequence analysis using single-stranded DNA obtained from pEJ3 and the helper phage M13KO7.

As shown in Figure 2, cells transformed with the recombinant plasmid pEJ3 produce two new protein bands with approximate molecular weights of 68 kD and 28 kD. The other cells, containing either plasmid pEMBLex3 or pEJ3R, do not synthesize these two proteins. The molecular weights of the two new proteins which were produced at 42 are consistent with the reported values for PRD1 DNA polymerase and terminal protein (11), and roughly agree with the 63 kD and 29 kD values deduced from the nucleotide sequence (4, 6).

Two other DNA fragments containing the PRD 1 DNA polymerase gene were obtained (described in Materials and Methods). These fragments, 2.5 kb and 2.0 kb in size, each contain the PRD1 polymerase gene and part of the terminal protein gene. Both fragments were cloned into Pst I site of pEMBLex3 in order to create a vector which synthesizes only PRD1 DNA polymerase (Fig. 1). The plasmids which contain the 2.5 kb and 2.0 kb inserts are designated pEJ2.5 and pEJ2, respectively. Expression of the cloned gene from pEJ2.5 and pEJ2 was very low, based on analysis of coomassie blue-stained SDS polyacrylamide protein gels (Fig. 2).

Site-specific Deletion of pEJ2.

Because synthesis of PRD1 DNA polymerase from pEJ2 and pEJ2.5 was very low, a new plasmid construction was prepared. 57 bp of pEJ2 DNA was deleted (Fig. 5). This deletion removes a small segment of 3' end coding sequence for

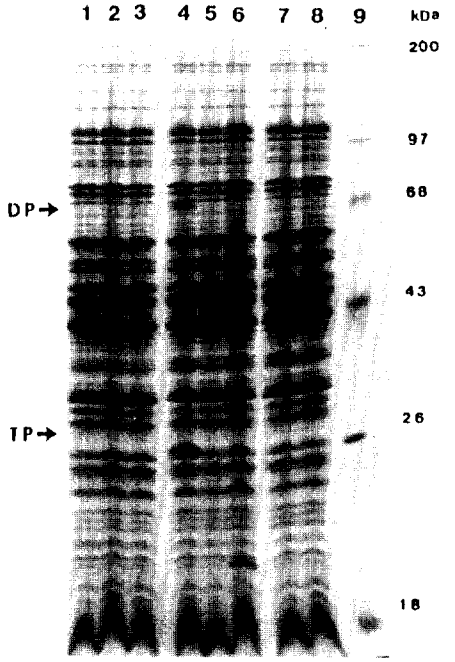

Fig. 2. *Expression of PRD1 DNA polymerase (DP) and terminal protein (TP). Lane 1, 4 and 7 represent the samples of cells harboring pEJ2.5, pEJ3 and pEJ2, respectively and 3 and 6 represent the samples of cell harboring pEJ2.5R and pEJ 3R, respectively. 2, 5 and 8 are control (pEMBLex3). Lane 9 contains standard protein molecular markers. Numbers indicate size in kDa.*

terminal protein, the ribosome binding site of the DNA polymerase gene, and the small N-terminal sequence of the lac Z gene of pEMBLex3 (Fig. 3). After this specific deletion, the PRD1 DNA polymerase gene remained under the control of the lambda pR promoter and utilized the ribosome-binding site of the lambda cro gene.

The specific 57 bp deletion was created by making uracil-containing single stranded DNA and utilizing a 29-mer oligonucleotide, which spanned the desired deletion. The first 16 bp of the oligonucleotide corresponded to vector sequence including the ribosome binding site(SD) of pEMBLex3. The remaining nucleotides represent the PRD1 DNA polymerase gene. The distance between the vector SD sequence and the initiation codon of the PRD1 DNA polymerase gene was 6 nucleotides long (Fig. 3). Details of the procedure used to create the deletion are described in Materials and Methods. The 57 bp deletion removes a Pst I site from pEJ2. Plasmids containing the deletion were identified by digestion with Pst I.

The efficiency of deletion mutagenesis was 63%. Several plasmids were checked for expression of DNA polymerase on SDS polyacrylamide protein gels. Every clone except one expressed the 68 kD PRD1 DNA polymerase well and no expression of terminal protein was observed (Fig. 4).

The plasmid created by specific deletion which has only the PRD1 DNA polymerase gene was named pEJG. The expression level of DNA polymerase from pEJG was slightly better than that observed in pEJ3. Further confirmation of

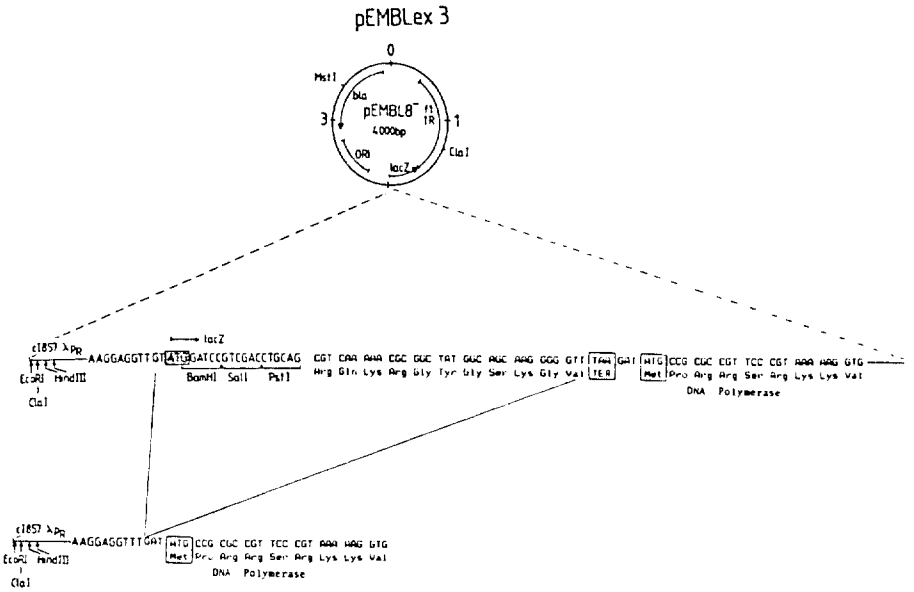

Fig. 3. *The nucleotide sequence of the deletion region in pEJG before and after 57 bp specific deletion.*

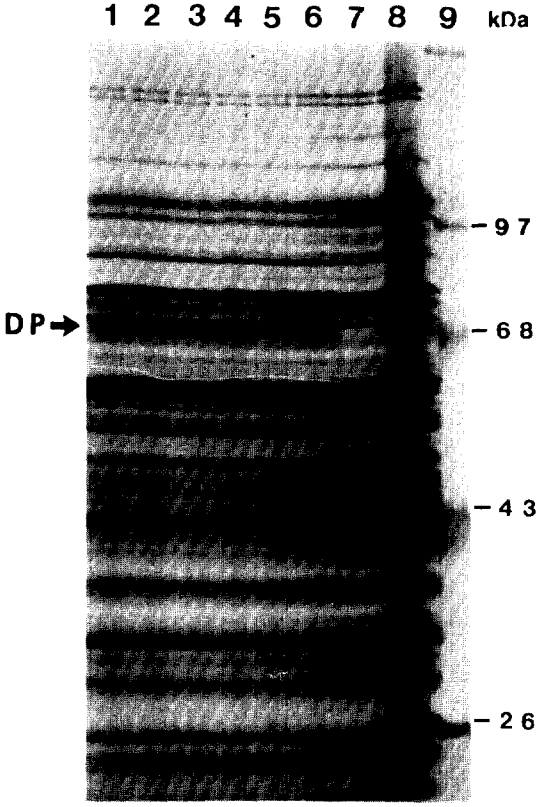

Fig. 4. *Expression of PRD1 DNA polymerase (DP) after specific 57 bp deletion (pEJG). Lanes from 1 to 6 represent the different clones of cells harboring pEJG plasmid. Lane 7 and 8 represent the samples of cells harboring pEMBLex3 and pEJ3, respectively. Lane 9 contains standard protein molecular markers. Numbers indicate size in kDa.*

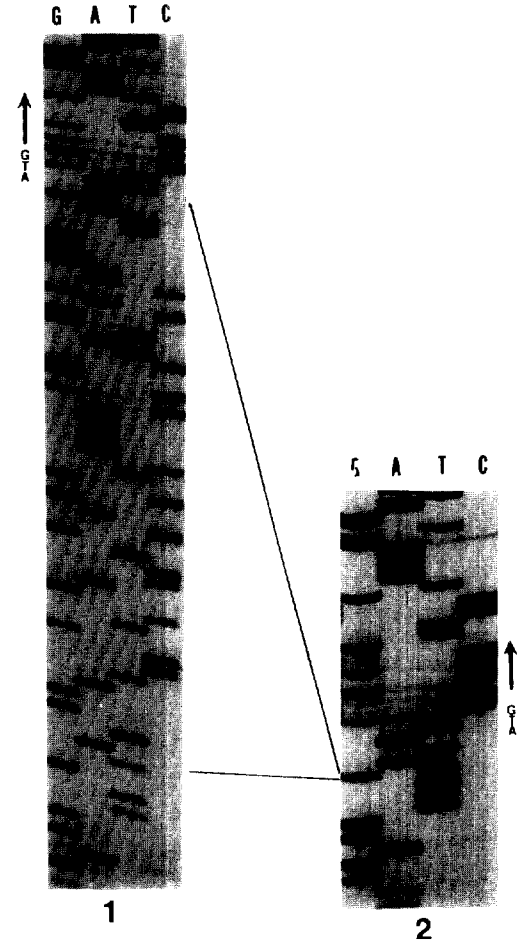

Fig. 5. *Autoradiogram showing nucleotide sequence of the region that has been deleted. 1 represent before deletion (pEJ2) and 2 represent after deletion (pEJG). ATG indicates start codon of PRD1 DNA polymerase gene and arrow represent the direction of gene.*

the deletion was obtained by producing pEJ2 and pEJG single-stranded DNA using the helper phage M13K07. The single-stranded DNAs were used to sequence the deleted region by the dideoxy chain-termination method (Fig. 5). Unexpectedly, nucleotide sequence analysis revealed that the vector used in these experiments, pEMBLex3, deviates from the published sequence (16) at three positions located between the pR promoter and the start codon of the lac Z′ gene.

The published nucleotide sequence adjacent to the ribosome binding site is given as AGGAGGTTTTTTATG and the vector used here contained the sequence AGGAGGTTGTATG, where the underlined bases represent the Shine-Dalgarno sequence and the start codon for lac Z′, respectively. The 29 mer oligonucleotide used to create a deletion in these experiments was designed based on the published sequence, but fortunately contained enough homology in the region of the altered bases to anneal properly. It is likely that the published sequence of pEMBLex 3 is incorrect because further investigation revealed that the sequence obtained in this work is identical to that published for the lambda pR region including the disputed base pairs (14). As described earlier, one plasmid derived from the deletion mutagenesis appeared to contain the 57 bp deletion, but did not make significant amounts of PRD1 DNA polymerase. Single stranded DNA obtained from this clone was sequenced to reveal that there is a one base-pair alteration (G-A) located between the ribosome-binding site and the ATG start codon of the DNA polymerase gene (Fig. 6). The clone which

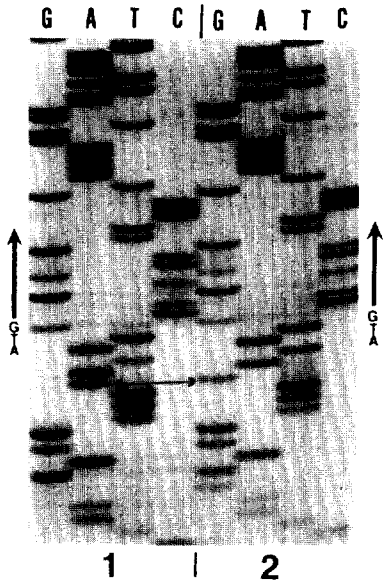

Fig. 6. *Autoradiogram showing nucleotide sequence of pEJG plasmids which show different expression. 1 represents the clone which shows poor expression and 2 represent clones which shows higher expression. Double-headed arrow indicates the base difference between two these clones.*

produces large amouts of PRD1 DNA polymerase, pEJG, does not have this base change. It is interesting that a nucleotide difference at one position between the ribosome binding site and the start codon appears to decrease expression of PRD1 DNA polymerase significantly.

DISCUSSION

Many processes contribute to the level of protein synthesis inside cells. The initial rate of protein synthesis is determined by the rate of transcription, mRNA stability and translation efficiency. The final steady-state level of a protein inside the cell also depends on its turn over rate. The strength of a promoter is probably the most important factor in transcription.

The lac, trp, lambda pL and tac promoters have been used most often for the construction of expression vectors. The lambda pL promoter is the strongest in this group. The lambda pR promoter is as strong as the pL promoter. pR and pL are not only strong promoters but also can be regulated by a repressor and thus transcription of a gene, cloned downstream of either promoter, can be turned on after a bacterial cell culture has reached the desired density. This regulation is particularly important when the expressed protein is lethal to the cell when overproduced. Regulation is also important when the cloned gene must be expressed to a high level in the cells because cells devote a large portion of their energy to the synthesis of the cloned gene when it is overexpressed from a multicopy plasmid.

Furthermore, if the presence of the cloned gene on a multicopy plasmid is deleterious to cell growth, then faster-growing segregants may eventually comprise a large fraction of a cell culture. The mutant repressor lambda cI857 is a temperature sensitive repressor which becomes non-functional at 42°C. Thus expression of a cloned gene transcribed from a pR or a pL can be regulated by a change in temperature when cI857 is present.

DNA polymerase is one of the most important and presumably one of the oldest enzymes, because its function is genetic information transmission from one generation to the next. Despite the importance of this enzyme, only a few DNA polymerase genes have been sequenced and very few DNA polymerase have been cloned into expression systems for subsequent study. The low natural abundance of this enzyme in the cell is one of the most important problems encountered in studying DNA polymerase.

Up to this time only the Klenow fragment of *E. coli* DNA polymerase I (3) which is the most intensively studied DNA polymerase, phage T4 DNA polymerase(8), phage subunit of T7 DNA polymerase (13, 18) and phage 29 DNA polymerase (2) have been cloned and the proteins purified for study from overexpressing cells.

Three DNA polymerases (Klenow fragment, T4 and T7) were synthesized to more than 10% of total cellular protein under the best conditions. Expression of PRD1 DNA polymerase in pEJ3 was low, about 1% or less of total protein. This is better than the results obtained with the 29 DNA polymerase expression system.

High level expression of the klenow fragment, the T4 DNA polymerase and the T7 subunit of DNA polymerase was achieved by placing a strong promoter upstream of the ribosome binding site and start codon of each gene. Therefore, the wild type translational regulatory sequence was present in these constructions.

Expression of a cloned gene is not limited solely by the rate of trancription initiation. After transcription, the mRNA structure determines its stability and influences translational efficiency. Also, the initiation codon, which is generally AUG of GUG, the Shine-Dalgarno (SD) sequence and the distance between these two sites are very important factors in determining the rate of translation. The nucleotide sequence composition from about -20 to $+15$ (with respect to the start codon) are also important factors (17). The nucleotide sequence encompassing the ribosome-

binding site and adjacent base-pairs of pEMBLex 3 is derived from the lambda Cro gene, whose product is known to be translated very efficiently. Thus the reason that low expression of pEJG is observed may be due to a characteristics of the nucleotide sequences after the ATG start codon.

It is well known that mRNA secondary structure around the Shine-Delgarno sequence and/or the start codon can inhibit translation initiation (3), but computer analysis predicts no secondary structure around this region of pEJG. mRNA stability is also a very important factor in determining gene expression. Most *E. coli* messenger RNA's have a half-life of about one to two minutes. Therefore gene expression might be drastically increased by enhancing the half-life of an mRNA, but the factors which determine mRNA stability are not known.

There exists a controversy over the correlation between codon usage and gene expression. Williams *et al.*, (20) have shown that a synthetic gene whose codon bias resembled that found in highly expressed *E. coli* genes (high codon preference) was expressed at levels 16 times higher than that of the native gene. They also confirmed that this higher expression was a direct result of the high codon preference, and was not due to altered promoter efficiency, mRNA secondary structure or mRNA stability. The PRD1 DNA polymerase gene was examined using a Codonpreference program (The University of Wisconsin Genetics Computer group) and was shown to have low codon preference. The low level of PRD1 DNA polymerase expression, in comparison to other DNA polymerases, is quite possibly a result of poor codon usage.

REFERENCES

1. **Bamford, D.H., Rouhiainen, L., Takkinen, K. and H. S derlund,** 1981. Comparision of the lipid-containing bacteriophages PRD1, PR3, PR4, PR5 and L17. *J. Gen. Virol.* **57**, 365-373.
2. **Blanco, L. and M. Salas,** 1984. Characterization and purification of a phage 29-encoded DNA polymerase required for the initiation of replication. *Proc. Natl. Acad. Sci. USA* **81**, 5325-5329.
3. **Buell, G.N., M.F. Schulz, G. Selzer, A. Chollet, N.R. Movva, D. Semon, S. Escanez and E. Kawasima,** 1985. Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-c (IGF-I). *Nucleic Acids Res.* **13**, 1923-1938.
4. **Hsieh, J., G. Jung, M.C. Leavitt, and J. Ito,** 1987. Primary structure of the DNA terminal protein of bacteriophage PRD1. *Nucleic Acids Res.* **15**, 8999-9009.
5. **Joyce, C.M. and N.D.F. Grindly,** 1983. Construction of a plasmid that overproducts the large proteolytic fragment (Klenow fragment) of DNA polymerase I of *E. coli*. *Proc.Natl.Acad.Sci. USA* **80**, 1830-1834.
6. **Jung, G., M.C. Leavitt, J.-C. Hsieh and J. Ito,** 1987. Bacteriophage PRD1 DNA polymerase: Evolution of DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**, 8287-8291.
7. **Kunkel, T.A.,** 1985. Rapid and Efficient sitespecific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488-492.
8. **Lin, T.-C., J. Rush, E.K. Spicer, and W.H. Konigsberg,** 1987. Cloning and expression of T4 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**, 7000-7004.
9. **Maniatis, T., E.F. Fritsch and J. Sambrook,** 1982. "Molecular Cloning: A Laboratory Manual" Cold Spring Harbor, N.Y.
10. **Mindich, L. and T. McGraw,** 1983. Molecular cloning of Bacteriophage PRD1 Genomic Fragment. *Mol. Gen. Genet.* **190**, 233-236.
11. **Mindich, L., D.H. Bamford, D. Goldthwaite, M. Laverty, and G. Mackenzie,** 1982. Isolation of Nonsense Mutants of Lipid-Containing Bacteriophage PRD1. *J. Virol.* **44**, 1013-1020.
12. **Ollis, D.L., P. Brick, R. Hamlin, N.G. Xuong and T.A. Steitz,** 1985. Structure of large fragment of *E. coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762-766.
13. **Reutiman, H., B.-M. Sjoberg and A. Holmgren,** 1985. Bacteriophage T7 DNA polymerase: Cloning and high-level expression. *Proc. Natl. Acad. Sci. USA* **82**, 6783-6787.
14. **Sanger F., Coulson A.R., Hong G.F., Hill D.F. and G.B. Peterson,** 1982. Nucleotide sequence of bacteriophage DNA. *J. Mol. Biol.* **162**, 729-773.
15. **Savilahti, H. and D.H. Bamford,** 1987. The complete nucleotide sequence of the left Gene very early region of *E. coli* bacteriophage PRD1 coding for the terminal protein and the DNA polymerase. *Gene* **57**, 121-130.
16. **Sollazzo, M., R. Frank and G. Cesareni,** 1985. High-level expression of RNAs and Proteins: the use of oligonucleotides for the precise fusion of coding-to-regulatory sequences. *Gene* **37**, 199-206.
17. **Stormo, G.D., T.D. Schneider and L.M. Gold,** 1982. Characterization of translational initiation sites in *E. coli*. *Nucleic Acids Res.* **10**, 2971-2996.
18. **Tabor, S. and C.C. Richardson,** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**, 1074-1078.
19. **Vieira, J. and J. Messing,** 1987. Methods in Enzymology **153**, 3-11. Academic Press, Inc., N.Y.
20. **Williams, D.P., D. Regier, D. Akiyoshi, F. Genbauffe and J. R. Murphy,** 1988. Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed *E. coli* genes. *Nucleic Acids Res.* **16**, 10453-10467.

(Received February 19, 1992)
(Accepted March 3, 1992)

초 록: Bacteriophage PRD1 DNA Polymerase의 과잉 발현

정구홍 (서울대학교 사범대학 생물교육과)

Bacteriophage PRD1 DNA polymerase를 *E. coli*에서 발현시키기 위하여 phage genome DNA로 부터 2 kb의 HaeII 단편을 분리하여 pEMBLex3 expression vector에 cloning 하였다. Uracil을 갖고있는 ssDNA를 이용하여 원하지 않는 non-coding region 57 bp를 제거하였다. 이 57 bp를 결실시킨 후에, PRD1 DNA polymerase 유전자는 전적으로 vector promoter와 SD sequence의 조절하에 있게 되었다. Heat induction시, 68 kdal 의 크기를 가진 PRD1 DNA polymerase로 발현되었으며, 발현 정도는 *E. coli* 전체 단백질의 약 1% 정도이었다.