INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9022111

Structure-function analysis of the bacteriophage PRD1 DNA terminal protein: Nucleotide sequence, overexpression, and site-directed mutagenesis of the terminal protein gene

Hsieh, Jui-Cheng, Ph.D.

The University of Arizona, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

STRUCTURE-FUNCTION ANALYSIS OF THE BACTERIOPHAGE PRD1 DNA TERMINAL PROTEIN: NUCLEOTIDE SEQUENCE, OVEREXPRESSION, AND SITE-DIRECTED MUTAGENESIS OF THE TERMINAL PROTEIN GENE

by

Jui-Cheng Hsieh

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY

In Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MICROBIOLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1990

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by JUI-CHENG HSIEH

entitled STRUCTURE-FUNCTION ANALYSIS OF THE BACTERIOPHAGE PRD1 DNA TERMINAL PROTEIN: NUCLEOTIDE SEQUENCE, OVEREXPRESSION, AND SITE-DIRECTED MUTAGENESIS OF TERMINAL PROTEIN GENE

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Junetsu Ito	1/25/90
Harris Bernstein	Date 1/25/90
Neil H. Mendelson	Date 1-25-90
Mary Ann Hanlon	Date 1-25-90
Charles P. Kerk	Date 1-25-90
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

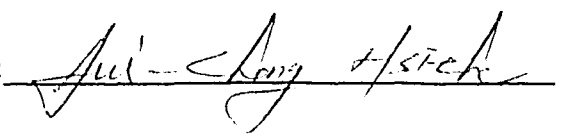
Junetsu Ito — 1-25-90

Dissertation Director Junetsu Ito — Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interest of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: 

TO MY PARENTS, WIFE AND CHILDREN

ACKNOWLEDGEMENTS

I am especially indebted to my dissertation advisor, Dr. Junetsu Ito, without whose guidance this work would not have been possible.

Thanks are also due to the members of my dissertation committee, Dr. Harris Bernstein, Dr. Neil H. Mendelson, Dr. Charles P. Gerba and Dr. Martinez J. Hewlett for their help.

Sincere appreciation to my close friends and colleagues, Mark Leavitt, Seung-ku Yoo, Dan Gerendasy, Michael J. Moran and Jason D. Bannan for our discussions and collaborations.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure

LIST OF TABLES

ABSTRACT

The nucleotide sequence of the PRD1 terminal protein gene has been determined. The coding region for PRD1 terminal protein is 777 base pairs long and encodes 259 amino acid residues (29,326 daltons). The initiation codon, ATG, is preceded by an identifiable ribosome binding site, putative promoter sequences, phosphate box, and plasmid pBR322 origin-like sequence. In addition, there are unique palindromic sequences between the initiation codon ATG and the "-10" promoter sequence.

The deduced amino acid sequence of PRD1 terminal protein reveals no overall homology with other known terminal proteins or related proteins. A closer examination revealed a highly conserved amino acid sequence, YSRLRT, exist among all identified DNA terminal proteins including PRD1, PZA, Nf, ϕ29 and adenovirus. This is the first conserved amino acid sequence that has been found in all identified DNA terminal proteins. Not only is the YSRLRT sequence conserved, but its spatial location is similar as well. Therefore, the significance of the YSRLRT conserved sequence is suggested by both its conservative spatial location and high degree of homology across species. To

study the structure-function relationship of the YSRLRT sequence of PRD1 terminal protein, in vitro site-directed mutagenesis was performed to determine the role of each amino acid in this conserved region. The PRD1 terminal protein and DNA polymerase genes were cloned into phagemid pEMBLex3, and the recombinant plasmid used for constructing mutants. Eleven PRD1 terminal protein mutant clones were examined for their priming complex formation activities. Our results have strongly demonstrated that the positive charge residue of arginine-174 plays an important role for PRD1 terminal protein function.

There are 13 tyrosine residues in the predicted PRD1 terminal protein. It was of interest to know which tyrosine is actually linked to terminal nucleotide of the PRD1 DNA. We used a new approach involving replacing the tyrosine residues with phenylalanine residues in the carboxyl terminal portion of the protein. From analyses, the tyrosine-190 has been determined to be the most likely linkage site between terminal protein and PRD1 DNA.

CHAPTER 1

INTRODUCTION

An Overview of Proteins Which Covalently Bind to DNA

There are several types of proteins that covalently bond to nucleic acidsthat are involved in genomic replication or recombination. The first type of proteins are involved in breaking a polynucleotide chain and simultaneously bind covalently to phosphate groups at nicked sites. These proteins are the E. coli and Micrococcus luteus omega proteins, DNA gyraseA subunit of M. luteus (Tse et al., 1980), T4 DNA topoisomerase(Sugino et al., 1977; Rowe et al., 1984), the Rat liver nicking-closing enzyme (Champonx, 1981), and ϕx 174 A and A* proteins (Roth et al., 1984; Sanhueza et al., 1984). All of these proteins are capable of cleaving single-stranded DNA and bind covalently to a 5'- or 3'-phosphorylated terminus via a phosphodiester bond between a specific tyrosine residue and a terminal phosphate group.

Formation of covalent linkage between DNA and these types of proteins may be transient, serving as intermediates in DNA metabolism. For example, T4 DNA topoisomerase, DNA gyrase, or ϕx174 A protein act under specific conditions as endonucleases, the nicking activity being accomplished by covalent

linkage formation of the enzyme to the newly generated terminal phosphate. These processes could allow either the DNA to undergo topological isomerization or a newly generated 3'-OH group to serve as a primer for DNA polymerase (Eckard, 1982).

The second type of proteins are site-specific recombinases such as λ integrase, resolvase of transponsons Tn3 and rδ, the Cre protein of phage P1, the Gin protein of phage Mu, the Hin protein of the phaseinversion system of S. typhmurium, and the FLP protein of the 2 um plasmid of S. cerevisiae (Sadowsk, 1986). The hallmarks of all site-specific recombinases are that they promote reciprocal recombination between two limited DNA sequences, and their target sequences are inherently asymmetric. All of these types of protein cause phosphodiester bond breaks in the recombination site and remain covalently bound to the phosphoryl terminus at the break. The Int, Cre and FLP recombinases are bound to the 3'-phosphoryl terminus, whereas resolvase and invertase are bound to the 5'-phosphoryl end (Andrews et al., 1985; Craig et al., 1983; Hoess et al., 1985; Reed et al., 1981). The FLP protein is attached to the DNA via a phosphotyrosine linkage (Gronostajski et al., 1985), whereas resolvase is bound through a phosphoserine residue (Reed et al., 1984). In this respect, site-specific recombinases are similar to the topoisomerases in that they

convert the energy of the phosphodiester bond in a protein-phosphate linkage and use that energy for subsequent religation of the phosphodiester backbone (Gellert, 1981).

The third type of proteins are the adenylated enzymes including E. coli and T4 DNA ligases (Gumport and Lehman, 1971), E. coli glutamine synthetase (Kingdon et al., 1967) and E. coli RNA polymerase (Chelala et al., 1971). To our knowledge, the E. coli and T4 DNA ligases are the first instances in which the AMP moiety is linked to the C-amino group of a single lysine residues of the enzymes through a phosphoamide bond, and these C-amino-linked lysine-AMP intermediates have been isolated by Gumport and Lehman (1971).

The fourth type of proteins are natively linked to the 5'-terminal phosphate of viral nucleic acids and nonviral linear plasmids.Because of their location on the end of the DNA molecule, these proteins have been referred to as terminal proteins. The DNA/RNA genomes that have a proven or candidate terminal protein are listed in Table 1. The linking amino acid residues of terminal proteins to nucleic acids can be serine (adenovirus, ϕ29, and cowpea mosaicvirus), threonine (Cp-1), or tyrosine (PRD1, poliovirus, foot-mouth diseases virus, encephalo-myocarditis virus, and human rhinoviruses). In general, the known 5'-linked terminal proteins of double-stranded DNA are larger than

TABLE 1

Viruses and Plasmids with Genome-Linked Terminal Proteins(TP)

Virus/Plasmid	Example	Genome	MW of TP
Plant viruses:			
Comovirus	Cowpea mosaic virus	2 ssRNAs	4-5kDa
Luteovirus	Potato leafroll virus	ssRNA	7kDa
Nepovirus	Tobacco ring-spot virus	2 ssRNAs	4-6kDa
Potyvirus	Tobacco etch virus	ssRNA	6-24kDa
Sobemovirus	Southern bean virus	ssRNA	10-14kDa
Animal viruses:			
Adenovirus	Adenovirus	dsDNA	55kDa
Calicivirus	Vesicular exanthema virus	ssRNA	10-15kDa
Picornavirus	Poliovirus Hepatitis A virus	ssRNA	3-12kDa
Parvovirus	Minute virus of mice	ssDNA	83kDa
Birnavirus	Infectious pancreatic necrosis virus	dsRNA	110kDa
Hepadnavirus	Hepatitis B virus	dsDNA(partially)	
Bacteriophages:			
ø29, Nf, PZA, M2Y(Bacillus phages)		dsDNA	30kDa
PRD1(E. coli phage)		dsDNA	29kDa
Cp-1(Pneumococcal phage)		dsDNA	28kDa
Linear plasmids:			
pSLA1, pSLA2(Streptomyces)		dsDNA	
pGKL1, pGKL2(Kluyveromyces)		dsDNA	
S1, S2(Maize)		dsDNA	

those of single-stranded RNA.

The DNA terminal protein is an intriguing protein. It provides a primer for the initiation of linear DNA replication. In addition, the terminal protein may enhance the fidelity of DNA replication at the 5' terminus of the linear DNA. Since chemical mismatches between the bases can occur with a frequency of 10^{-5} (Topal et al., 1976), which is substantially higher than the frequency of spontaneous mutation, it can be inferred that the accuracy of DNA synthesis is maintained. It has also been well documented that the terminal protein is essential for DNA encapsidation in the ϕ29 system (Bjornsti et al., 1984). Other possible functions of terminal protein are indicated by its ability (i) to protect the DNA from attack by exonucleases (Dunsworth-Browne et al., 1980), (ii) to show the DNA nicking-closing, DNA ligase, and/or DNA topoisomerase activities (Ruben et al., 1983), and (iii) to prevent internal starts and somehow stabilizes the initiation complex (Tamanoi and Stillman, 1982; Bergen et al., 1983).

Protein-Priming Mechanism Specific for Linear DNA Replication

All known DNA polymerases are only able to elongate existing polynucleotide chains and can not de novo synthesize DNA without a primer.

Hence, a fundamental problem arises for these linear DNA molecules. As shown in Figure 1, the gap created at the 5' end of a new chain, after the initiating RNA primer has been removed, can not be filled in by DNA polymerase because there is no adjacent 3'-OH group to serve as a primer and DNA polymerases act only in the 5' to 3' direction. No such problem of replication arises in the case of circular chromosomes which have no ends. This may explain why most bacterial and viral genomes are circular. There are four models which have been proposed to circumvent this problem of replication. They are: the cohesive end model (i.e., phage λ), the terminal redundant end model (i.e., phage T7), the palindromic end model (parvovirus), and the cross-linked end model (vaccinia virus). It is suggested that genome circularization occurs in phage λ (Weige et al., 1973) and concatermerization occurs in phage T7 (Watson, 1972). Both mechanisms require identical sequences at the two ends of the genome. According to the palindromic end model (Cavalier-Smith, 1974), parvovirus can form a self-complementary hairpin loop and start DNA replication. This is followed by activity of a DNA ligase, a specific endonuclease, and DNA polymerase. Finally, the cross-linked end model suggests that one end of the cross-linked viral DNA unwinds to form a single-stranded DNA bubble where replication can be started (Bern and Silverman,

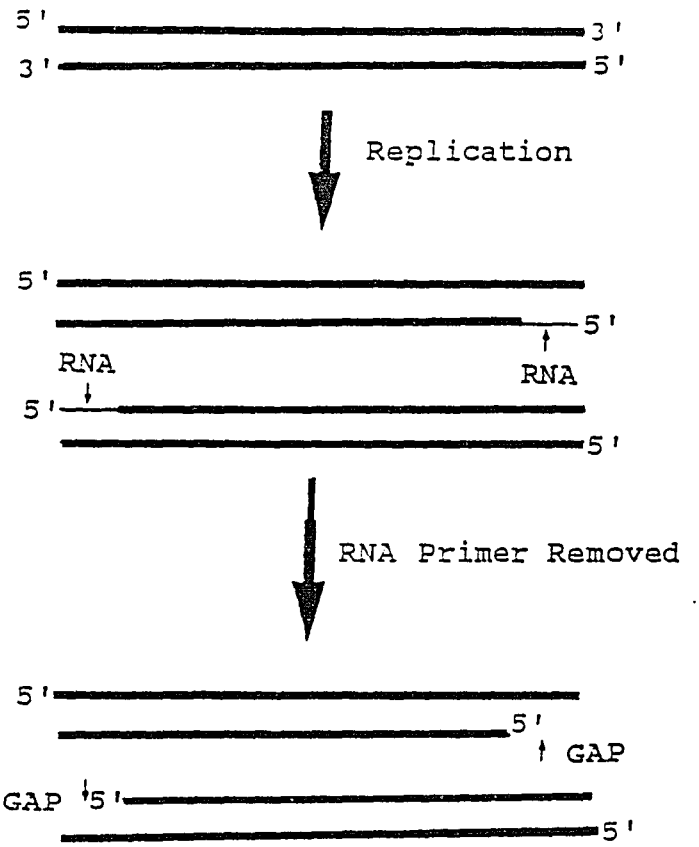


Figure 1. Gap created at 5' ends of progeny strands after RNA-primed DNA synthesis of a linear genome.

1970; Holowczak, 1976; Esteban et al., 1977). This appears to be the case in vaccinia virus.

None of the above models can account for the initiation of adenovirus, ϕ29, or PRD1 DNA replication, because their genomes do not contain such structures. The distinguished feature of these linear DNA genome is the existence of a protein covalently linked to the 5' ends. The idea that 5'-terminal protein may serve as a primer for initiation of linear DNA replication was first proposed by Rekosh et al. (Rekosh et al., 1977) in the adenovirus system. Up to now, this "protein-priming" model has been demonstrated in various systems including ϕ29 (Ito et al., 1979; Harding and Ito, 1980; Inciarte et al., 1980; Watebe and Ito, 1983; Watabe et al., 1983; Watabe et al., 1984), PRD1 (Yoo and Ito, 1989), adenovirus (Challberg et al., 1979; Lichy et al., 1981; Challberg et al., 1980; Horowitz et al., 1978), and poliovirus (Takeda et al., 1986), and encephalomyocarditis (Vartapetian et al., 1984). Therefore, it is clear that protein-priming provides a mechanism by which the 5' end of linear DNA can be preserved.

Molecular Aspects of Phage PRD1 and Its Replication

Bacteriophage PRD1 is the prototype of a group of small icosahedral,

lipid-containing bacterial viruses (Olsen et al., 1974; Bamford et al., 1981). This phage infects a wide variety of gram-negative bacteria harboring plasmids of the P, N or W incompatibility types (Olsen et al., 1974; Mindich et al., 1982). The receptor for PRD1 is believed to be plasmid encoded pili (Bradley, 1976). The PRD1 genome is a linear, double-stranded DNA molecule of about 14.7 kilobases with a 29 kd terminal protein covalently linked at both 5' ends (Bamford et al., 1984; McGraw et al., 1983). The linkage between the terminal protein and PRD1 DNA involves a phosphodiester bond between a tyrosine residue of the terminal protein and 5'-terminal nucleotide dGMP (Bamford and Mindich, 1984). The PRD1 DNA contains a perfect inverted terminal repeat of 110-111 bp which does not bear substantial homology to that of any known viral DNAs with terminal protein (Savilahti and Bamford, 1986; Gerendasy and Ito, 1987).

A number of nonsense and temperature sensitive mutants PRD1 have been isolated. The genetic map was generated by using cloned fragments of the PRD1 genome which were capable of complementing the nonsense mutants in marker rescue experiments (Mindich et al., 1982). Up to 19 genes have been arranged in a linear genetic map and the DNA terminal protein gene is mapped at the left end of the PRD1 genome (McGraw et al., 1983). The

genetic and physical maps of the PRD1 genome are shown in Figure 2. The gene organization of PRD1 indicates that the early functions (e.g., replication and transcription control) are located at the ends of the genome and the structural or morphogenic genes are at the center (Mindich et al., 1982). PRD1 induced the synthesis of at least 25 proteins in the infected cells and 17 of these proteins were components of the virion. On the basis of the timing of the synthesis, the viral-encoded proteins were classified into 3 groups: very early proteins (P1, P8 and P12), medium early proteins (P15 and P19), and late proteins. The twenty three PRD1-encoded proteins are summaried in Table 2.

Current understanding of the replication of linear, nonredundant, duplex DNA with terminal protein is as follows: replication initiation starts at either end of the DNA molecule via protein-primed strand displacement. The free PRD1 terminal protein molecule interacts with 5'-terminal nucleotide dGTP and forms a complex. This PRD1 terminal protein-dGMP complex contains a free 3'-OH group which can be used for subsequent chain elongation by the catalysis of PRD1-encoded DNA polymerase. The proposed protein-priming mechanism of the PRD1 DNA replication is shown in Figure 3. The subsequent elongation of the DNA chain displaces one parental strand. The displaced single-stranded DNA is then replicated by complementary DNA synthesis. To date, the actual

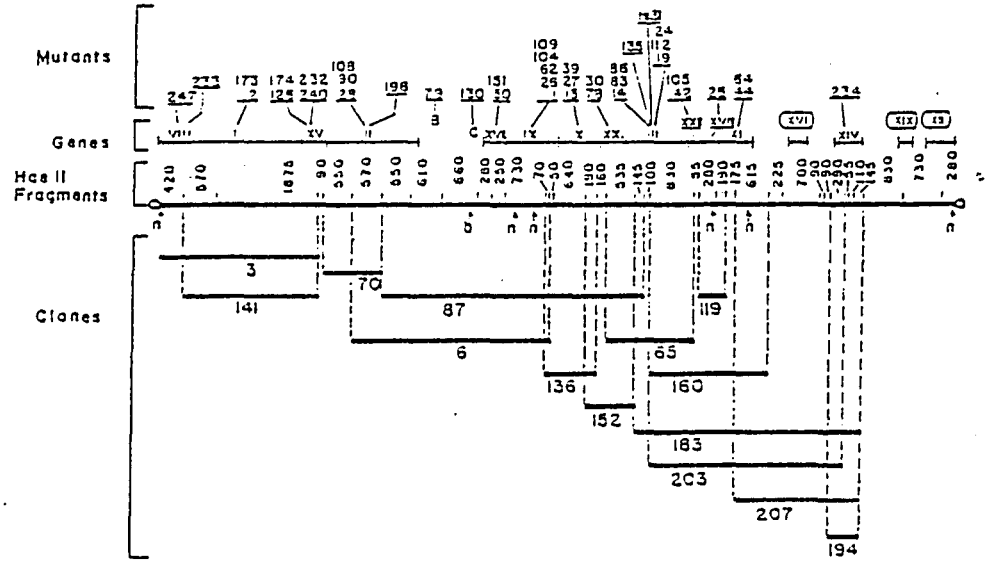


Figure 2. Composite of genetic and physical maps of the PRD1 genome (From McGraw et al., 1983).

TABLE 2

PRD1 Genes, Time of Expression, Molecular Weight and Role

Gene	Time of expression	Mol. wt. (KD)	Role
1	Early	63	DNA polymerase
2	Late	68	Attachment to host
3	Late	42	Major capsid protein
5	Late	30	Minor capsid protein
8	Early	29	Terminal protein
9	Late	28	DNA packaging
10	Late	26	Assembly
11	Late	23	Infectivity
12	Early	20	DNA synthesis
14	Late		Infectivity
15	Middle early		Lysin
16	Late	17	Infectivity
17	Late	15	Assembly
18	Late	11	Infectivity
19	Middle early	9	
20	Late		
22	Late		
23	Late		Infectivity

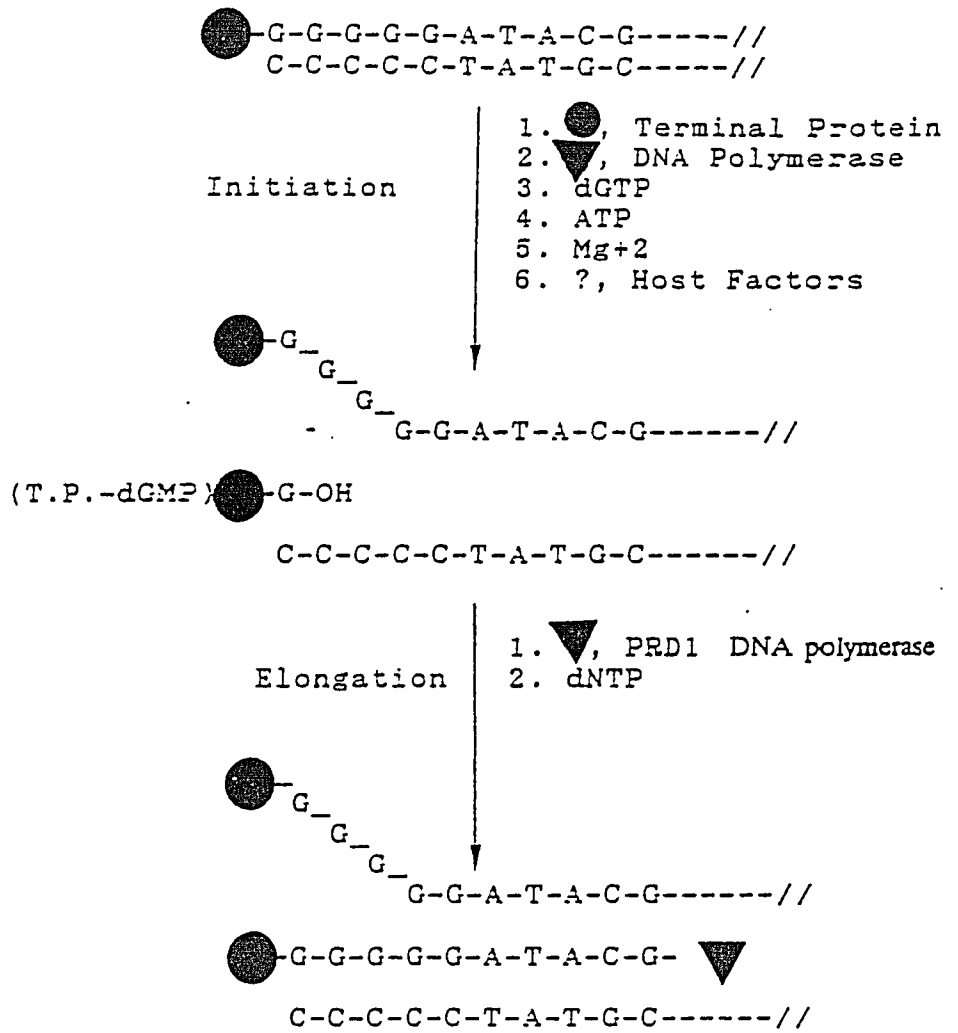


Figure 3. Protein-priming mechanism of the PRD1 DNA replication.

mechanism of complementary strand synthesis is not known, however, in the case of adenovirus DNA replication, it has been suggested that the displaced single-stranded DNA could make a panhandle structure via the inverted terminal repeats on both ends (Daniell, 1976; Lechner and Kelly, 1977). The DNA replication of adenovirus is shown in Figure 4. Although genetic studies suggest that panhandled structure molecules may be the replicative intermediates in vivo (Stow, 1982), such molecules have not yet been identified (Challberg and Kelly, 1982). However, this is a plausible model for the replication of such phages as PRD1 and Cp-1 because they contain long inverted terminal repeat sequences. In the case of phage ϕ29, there exists only a short inverted terminal repeat (only 6 bp), and therefore, ϕ29 might not be to able form a stable panhandled intermediate.

The origin and direction of PRD1 DNA replication in vitro has been determined by restriction enzyme analysis of 32-P labeled PRD1 DNA products. The results demonstrated that the replication starts at both ends of the linear PRD1 DNA template (Yoo and Ito, 1989). However, it is not known whether PRD1 DNA replication starts at both ends simultaneously. In the cases of adenovirus and ϕ29 DNA replication systems, simultaneous initiation at both termini did not take place (Lehner and Kelley, 1977; Harding and Ito, 1980;

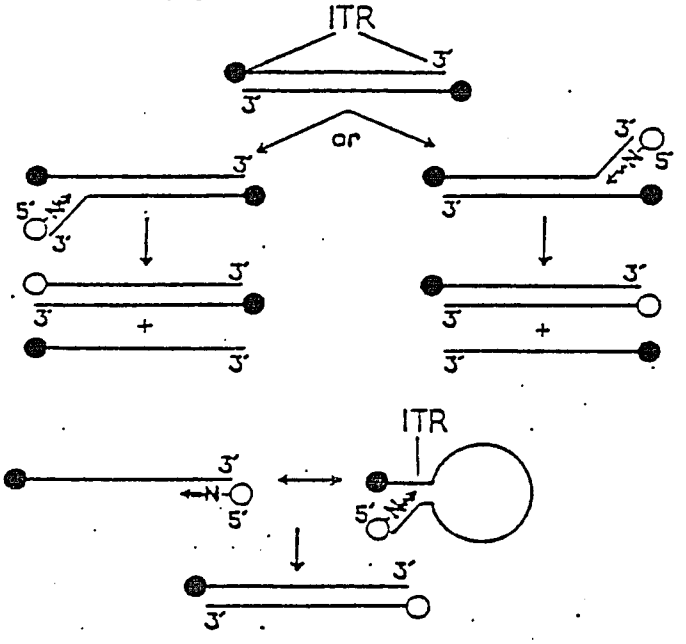


Figure 4. The model of protein-primed replication of adenovirus linear DNA. The solid circle denotes the parental terminal protein; the open circle denotes the free priming terminal protein. ITR is inverted terminal repeat sequences. (From Vartapetian and Bogdanov, 1987).

Inciarte et al., 1980). Another interesting feature is the comparison of PRD1 DNA replication with the ϕ29 system. ϕ29 requires at least six viral-encoded proteins (P1, P2, P3, P5, P6, and P17) for DNA replication, whereas PRD1 only requires three viral-encoded proteins (P1, P8, and p12) (Mindich, et al., 1982). Despite the differences in the proteins involved, both the terminal protein and DNA polymerase are absolutely required for DNA replication in both organisms.

On the basis of the following facts: (a) PRD1 genome is the smallest known linear duplex DNA containing long inverted terminal repeat sequences with terminal protein; (b) PRD1 can infect gram negative bacteria such as E. coli and S. typhimurium; and (c) numerous conditionally lethal mutant phages have been isolated; the PRD1 system provides the useful model to study the "protein-priming" mechanism of linear DNA replication, through both genetic and biochemical manipulation.

SPECIFIC AIMS

Unlike RNA polymerase, all known DNA polymerases cannot begin synthesis de novo; they must have a free 3'-OH available to serve as an acceptor for the transfer of the first nucleotide and to form the first phosphodiester bond. The molecules that offer the required 3'-OH have been

termed "primers". The primer can be a small RNA molecule, tRNA, nicked DNA or a hairpin structure with a 3'-OH end. A new type of DNA polymerase has been discovered in both human adenovirus and phage ϕ29. These two DNA polymerases utilize terminal proteins as primers for the initiation of DNA chain synthesis. This process is referred to as protein-primed DNA replication. Although many terminal proteins have been discovered in various systems, they have not been well characterized. Many aspects of the molecular biology of terminal protein are not understood. These aspects include: (a) Do DNA terminal proteins share a conserved amino acid sequence for performing their common function ? (b) What is the nature of association of terminal protein with DNA polymerase ? (c) How do the terminal proteins recognize each other between free priming terminal protein and parental template DNA-linked terminal protein ? In order to further understand the mechanism of protein-primed DNA replication in detail, one must first know the structure of the terminal protein.

The DNA replication of bacteriophage PRD1 is also initiated by a "protein-primed" mechanism similar to those of ϕ29 and adenovirus. The objective of this dissertation is to study the structure-function analysis of PRD1 terminal protein molecule.

CHAPTER 2

Bacteria, Phage, and Plasmids

Phage PRD1 was kindly provided by both R. Olsen and L. Mindich. Phage PRD1 was grown in Salmonella typhimurium LT2 carrying the drug resistance plasmid pLM2. The plasmid is necessary for PRD1 absorption. E. coli HB94 (pLM2,pLM3) which contains PRD1 genes 8 (terminal protein gene) and 1 (DNA polymerase gene) was generously provided by L. Mindich at the Public Health Research Institute of the City of New York. E. coli NM522 (lac-proAB), thi, hsd 5, supE, [F', proAB lacIqZ M15] (Grough and Murray, 1983) and E. coli RZ1032 [HfrKL16 PO/45{lysA(61-61)}, dut1, ung1, thi1 recA1] (Kunkel, 1985) were used as host strains for site-directed mutagenesis experiments. The plasmids pT7-6 in E. coli strain HMS174 (recA1 hdsR) and pGP1-2 in E. coli strain K38 (HfrC λ) (Tabor et al., 1985) were used for the exclusive expression of inserted genes. The phagemid pEMBLex3 (Sollazzo et al., 1985) was a kind gift from Dr. G. Cesareni of the European Molecular Biology Laboratory. The helper phage M13KO7 (Vieira and Messing, 1987) was purphased from IBI company.

Enzymes and Chemicals

Restriction endonucleases were purchased from Pharmacia and New England Biolabs. Calf intestinal alkaline phosphatase, T4 polynucleotide kinase, T4 DNA ligase, RQ1 DNase, and proteinase K were obtained from Sigma and Promega Corporations. [α-32P]dGTP, [γ-32P]ATP, and [35-S]-methionine were purchased from New England Nuclear. Sequenase kits were acquired from United State Biochemical Corporation. An in Vitro Muta-Gene Enzyme kit was obtained from Bio-RAD Corporation and a Geneclean kit purchased from BIO 101 Inc. Oligonucleotide synthesis chemicals were from Biosearch Corporation.

Preparation of Phage PRD1 and Plasmid DNA

Phage PRD1 were grown in S. typhimurium LT2 in LB medium containing kanamycin at a concentration of 50 ug/ml at 37°C. Host cells were infected with PRD1 at an M.O.I. of 5 at a cell density of 100 klett units. After infection, the phage PRD1 were allowed to adsorb without shaking for 10 minutes and incubation was continued until host cell lysis occurred. The liberated PRD1 phages were concentrated using 12% PEG and 0.4 M NaCl and purified through a 5% to 20% (W/V) linear sucrose gradient. The phage PRD1 terminal protein-DNA templates were further purified by boiling for 7.5 minutes and

concentration via a CsCl gradient centrifugation at 150,000 g for 36 hours. Deproteinized PRD1 DNA was isolated by suspending PRD1 particles in 2% SDS, 400 ug/ml proteinase K and heating for 2' at 65°C. The DNA was then extracted with phenol/chloroform and precipitated with ethanol.

DNA Sequencing

The pLM3 plasmid was first cleaved with PstI and the resulting DNA fragment containing genes 8 (terminal protein gene) and 1 (DNA polymerase gene) subcloned into phage M13mp9. A nested set of deletions was generated by the method of Dale et al., (1985). Each set of deletions was sequenced by the dideoxy chain termination method (Sanger et al., 1977). Additionally, the 3kb restriction maps of HaeIII, BanI, RsaI, TaqI and HinfI sites were contructed using partial and double digested methods. For confirmation, most regions were also sequenced by the chemical method (Maxam and Gilbert, 1980).

Computer Analysis of Sequence Data

Determination of open reading frames, deduced amino acid sequences, amino acid composition, codon usage and restriction sites were performed using the microcomputer programs of Mount and Conrad (1984). NBRF protein

database searches were performed by a FASTP micro computer program (Lipman and Pearson, 1985). The hydropathic profile was determined using a program based on the method of Kyte and Doolittle (1982). Secondary structure was determined using a program based on the method of Chou and Fasman (1978).

Cloning of PRD1 terminal protein and DNA polymerase genes into T7 RNA polymerase/promoter system

A 3kb DNA fragment that contains PRD1 terminal protein gene, DNA polymerase gene and the adjacent 90% of the lysin gene, was cloned into the T7 RNA polymerase/promoter system. This system contains two compatible plasmids and exploits the unique specificity of T7 RNA polymerase for its own promoter to control the expression of inserted genes. The cloning strategy is shown in Figure 5. The first plasmid, pT7-6, was used for cloning the 3kb insert in the PstI site and its subsequent transformation into E. coli NM522. After selecting the recombinant plasmid, pT7-3K, that contained the 3kb insert in the correct orientation, pT7-3k was transformed into the second host E. coli K38, a strain harboring the second plasmid pGP1-2. The pGP1-2 plasmid contains the temperature-sensitive λ repressor cI857 and T7 RNA polymerase gene.

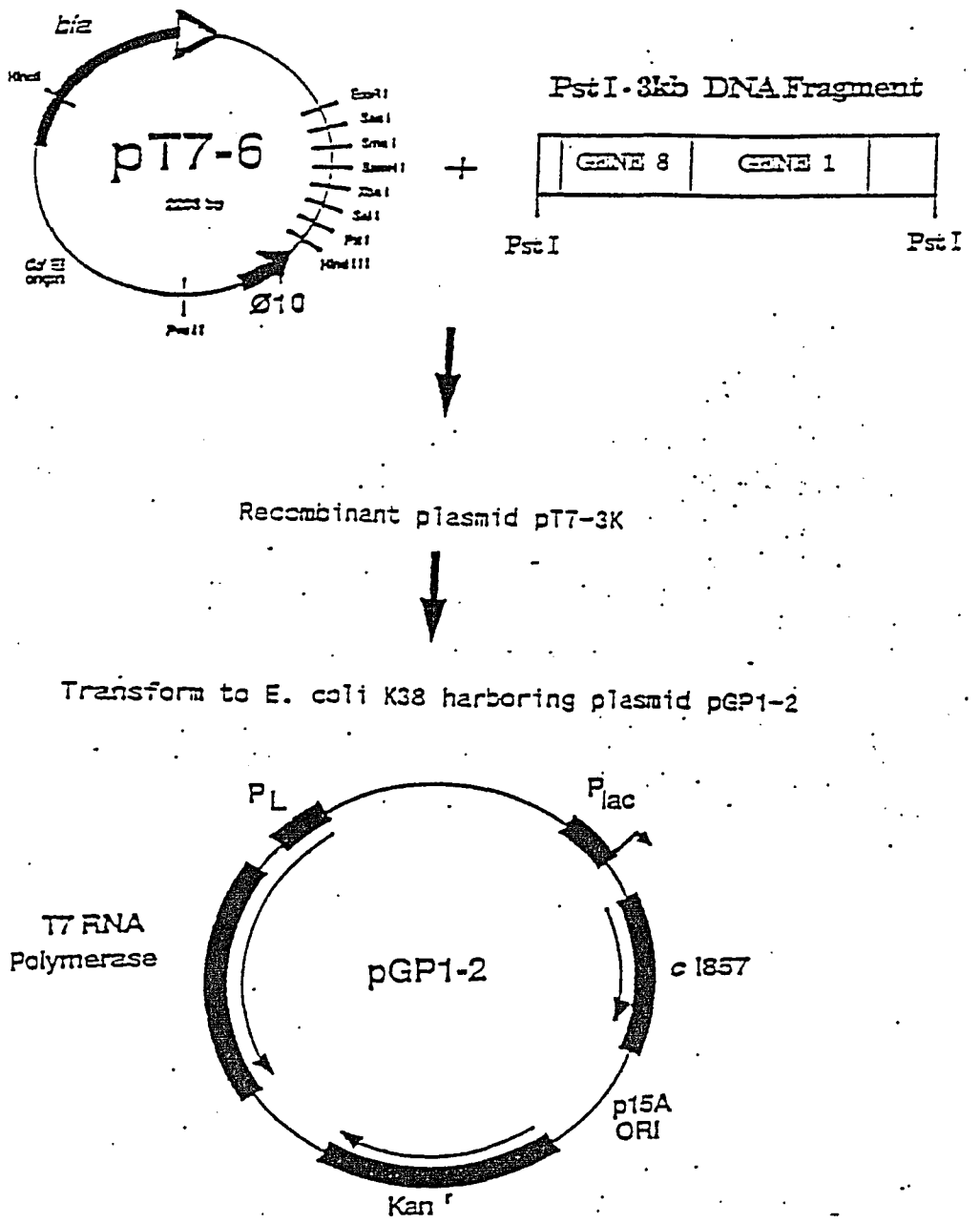


Figure 5. Cloning strategy in T7 RNA polymerase/promoter system.

Radioactive labeling of plasmid-encoded proteins

1 ml of cells was removed from cultures when the cells reached an OD_{590} =0.5. The cells were harvested and washed once with minimal M9 media. The cells were resuspended in 1 ml of M9 media containing 0.05% each of 18 amino acids (minus cysteine and methionine). The cells were grown with shaking at 30°C for 60 min. Expression of inserted genes was induced by raising the temperature to 42°C for 20 min. Rifampicin (20 mg/ml in methanol) was added to a final concentration of 200 ug/ml, and the cells incubated at 42°C for an additional 10 min. The temperature was then reduced to 30°C for 20 min. 1 ml of cells were radioactively labeled by the addition of 10 uCi of [35-S]methionine for 10 min. The cells were harvested by centrifugation for 30 s in a microcentrifuge and resuspended in 50 ul of sample buffer. The sample were heated at 95 °C for 2 min immediately prior to polyacrylamide gel electrophoresis.

Cloning of PRD1 terminal protein and DNA polymerase genes into phagemid expression vector pEMBLex3

A 3kb DNA fragment of phage PRD1 containing the terminal protein and DNA polymerase genes was digested from plasmid pLM3 with restriction enzyme PstI. This PstI-3kb DNA fragment was ligated into the PstI cloning site

of phagemid pEMBLex3 that had been digested using PstI and dephosphorylated with calf intestinal phosphatase. The ligation mixtures were transformed into E. coli NM522 and screened by restriction enzyme BglI analysis. The cloning strategy is shown in Figure 6. The recombinant clone was named "pEMBL 3K" and used for site-directed mutagenesis experiments.

Preparation of uracil-containing ss-DNA template

RZ1032 cells harboring recombinant pEMBL 3K plasmid, were grown at 30°C in 2X YT supplemented with 0.001% thiamine, 150 ug/ml ampicillin, and 0.25 ug/ml uridine. The early log phase culture was infected with helper phage M13KO7 at a M.O.I. of 20 and incubated at 37°C for 75 minutes with low shaking. Afterwards kanamycin was added to a final concentration of 70ug/ml to maintain the M13KO7 and the culture was incubated at 30°C with vigorous shaking. After 20 hours, 1 ml of cell culture was centrifuged for 5 minutes and 800 ul of supernant was transferred to a new eppendorf tube. 50 ul of 5 M NaCl and 50 ul of 40% PEG were then added, mixed and the mixture was placed at room temperature for approximately 5 minutes. The viral pellet was obtained after 5 minutes of centrifugation at 10k rpm.

The viral pellet was resuspended in 200 ul TE buffer. 50 ul of TE saturated phenol (PH7.0) was added and the suspension vortexed for 2 minutes, then

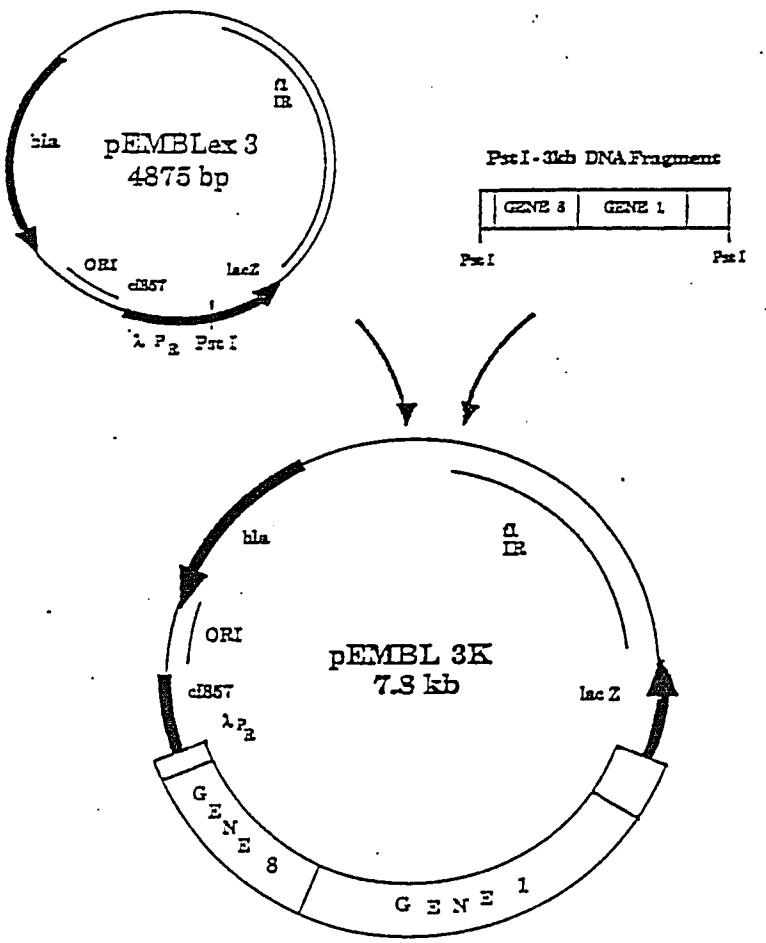


Figure 6. Schematic outline of the construction of recombinant plasmid pEMBL 3K.

spun 5 minutes at 10k rpm. The aqueous layer was removed to a fresh tube and phenol/chloroform extraction was repeated until the interphase was clear. Finally, the uracil-containing ssDNA template was precipitated with 0.5 volume of 7.5 M ammonim acetate and 2.5 volumes of 95% ethanol. The wild type ssDNA template was prepared in same manner as described above.

Oligonucleotide Synthesis

All synthetic oligonucleotides used for site-directed mutagenesis and DNA sequencing were synthesized on the Biosearch Cyclone DNA Synthesizer. Beta-amidite synthesis kits used for the synthesis were purphased from Biosearch company (San Rafael, CA). The oligonucleotides were removed from the CPG column by treatment with 30% ammonium hydroxide. This cleaves the oligonucleotides from the support and dissolves the released oligonucleotides. This solution was then incubated at 55^{o}C for 5 hours, and dried in a heated Savant Speed Vac Concentrator. The crude synthetic oligonucleotides were suspended in 0.5 ml TE buffer and purified by electrophoresis through a 20% acrylamide/8M urea gel, visualized by UV shadowing, excised, and eluted in TE buffer at room temperature overnight. Finally, the gel eluate was purified by passage through a SEP-PAK column to obtain the purified oligonucleotides. The synthetic oligonucleotides are shown in Table 3.

Table 3. Sequence and location of oligonucleotides used for site-specific mutation and sequencing

Oligonucleotide	Nucleotide location	Sequence(5'-3')
Y172Ochre	736-762	TTAACGGCAACTAATCCCGCCGTACCT
Y172N	733-761	CTTTTAACGGCAACAATTCCCGCCGTACC
S173A	736-764	TTAACGGCAACTATGCCCGCCGTACCTAT
R174G	739-767	ACGGCAACTATTCCGGCCGTACCTATACG
R174H	741-767	GGCAACTATTCCCACCGTACCTATACG
R174L	739-767	ACGGCAACTATTCCCTCCGTACCTATACG
R174S	739-767	ACGGCAACTATTCCAGCCGTACCTATACG
R175G	742-770	GCAACTATTCCCGCGGTACCTATACGTCA
R175H	743-769	CAACTATTCCCGCCATACCTATACGTC
T176A	745-773	ACTATTCCCGCCGTGCCTATACGTCATTT
Y177F	750-776	TCCCGCCGTACCTTTACGTCATTTGAT
Y190F	788-816	TAAATTTATGACGTTTGACATTATTATTG
Y226F	896-924	CGAAGCCAAGGCGTTTAACCGTAACCGTA
Y246F	956-984	GGCTAAAAAGAAATTCAAGCGCCGTCAAA
Y254F	981-1007	CAAAAACGCGGCTTTGGCAGCAAGGGG
PI	pEMBLex3	ATAATGGTTGCATGT
PII	570-587	CGCGGCGAAGTGGTGCGC
PIII	739-767	ACGGCAACTATTCCGGCCGTACCTATACG

Position +1 is the first nucleotide of 5' left end of PRD1 genome.

In Vitro Site-Directed Mutagenesis

1 ug uracil-containing ss pEMBL 3K DNA was mixed with 10 ng phosphorylated mutagenic oligonucleotides, 1 ul 10X synthesis buffer, 1 ul T4 DNA ligase (2-5 units), and 1 ul T4 DNA polymerase (1 unit/ul). The mixture was incubated at 37°C for 90 min to synthesize the mutated strand. The uracil-containing double stranded pEMBL 3K DNA then was transformed into E. coli NM522 competent cells. Colonies were screened directly by DNA sequencing.

Preparation of Cell-Free Extract

Cells were grown at 30°C in 500 ml of 2YT medium to an OD_{590}=0.45, cells were then heat induced at 42°C for about 2 hours and then harvested at 5°C. The pellet was suspended in 6 ml of 50 mM Tris-HCl, ph7.6/10%(wt/vol) sucrose/ 1 mM DTT. Lysozyme and EDTA were added to final concentration of 300 ug/ml and 1 mM respectively. The mixture was incubated at 0°C for 5 minutes. Non-ionic detergent Brij 58 (final concentration of 0.1%) was then added and the mixture was incubated at 0°C for 40 minutes. The mixture was twice frozen in dry ice and thawed at 30°C. The lysate was adjusted to 0.8 M NaCl and centrifugated at 40,000 rpm for 90 minutes in a Beckman SW 50.1 rotor. The supernatant was adjusted to 50% saturation in ammonium sulfate. After slowly stirring for 30 minutes, the precipitate was collected and suspended

in the desired buffer. Further, the cell extract was treated with streptomycin sulfate or on a DEAE column to remove DNA. The cell extract was stored at -70°C until used. In general, the protein concentration of the cell extract was about 15-20 mg/ml.

Assay for PRD1 terminal protein-dGMP complex formation

The reaction mixture (50 ul) contained 50 mM hepes (PH 7.5), 5 mM MgCl , 3 mM DTT, 2 mM ATP, 0.5 uM [a-32P]dGTP (3000Ci/mmole), 1 ug DNA-protein complex, and 100 ug of cell extracts. After incubation at 30°C for 30 minutes, 5 units RQ1 DNase (Promega) was added to the reaction mixture. The mixture was incubated at 37°C for 30 minutes. The reaction was terminated by the addition of 50 ul stop solution (0.18 M sodium pyrophosphate, 0.05 M EDTA and 20% trichloroacetic acid) and incubated on ice for about 20 minutes. After centrifugation, the supernatant was discarded and the pellet dissolved in 30 ul of sample buffer (0.1 M tris-HCl, PH6.8; 2% SDS; 20% glycerol; 10% beta-mercaptethanol and 0.005% bromophenol blue). The reaction products were examined on a 15% SDS/PAGE followed by autoradiography.

Doubling Time Measurement

All identified mutant plasmids were transfered into E. coli NM522. Each

mutant culture was adjusted to about 12 kletts from a fresh overnight culture, and then allowed to grow at 30°C. The number of klett units was measured every hour using a KLETT-SUMMERSON photoelectric colorimeter.

CHAPTER 3

RESULTS

The Terminal Protein Is Required for Phage PRD1 Transfection

In preliminary experiments, we were interested in knowing whether or not the PRD1 terminal protein is required for transfection. Two different PRD1 DNA templates were prepared for these experiments. One was treated with proteinase K to remove 5' terminal proteins, while the other was the intact terminal protein-DNA template. After transfection, the numbers of plaques were counted. The result clearly showed that the proteinase-K-treated DNA template was not transfectious (Figure 7). This observation was similar to those of Bacillus phages such as ϕ29, M2 and GA-1, which terminal proteins were also required for their DNA templates transfection (Hirokawa, 1972; Kawamura and Ito, 1977; Arwert and Venema, 1974).

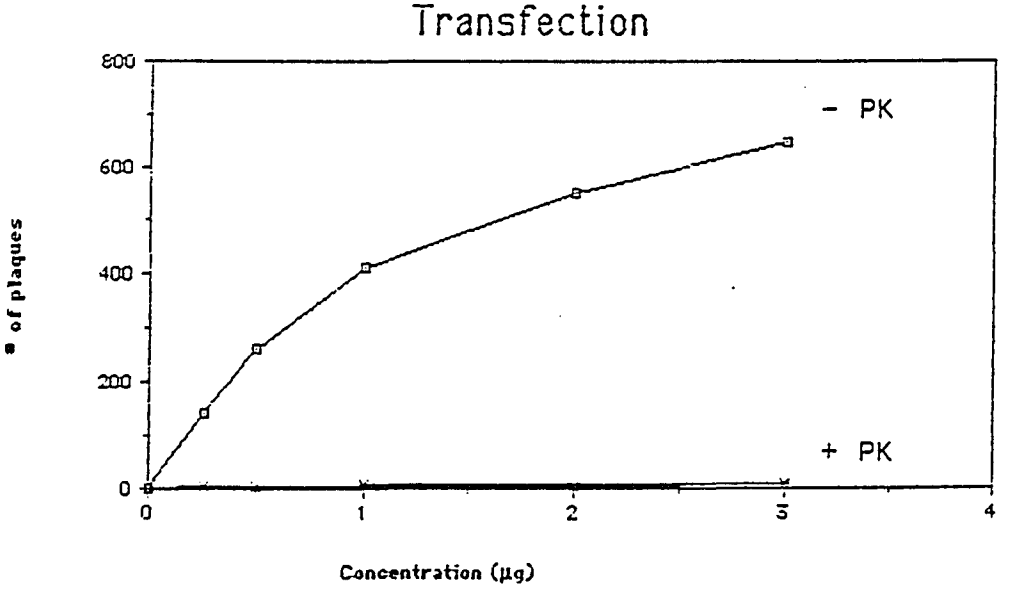


Figure 7. Effect of proteinase K (PK) on the transfection of PRD1 genome.

Nucleotide Sequence of PRD1 Terminal Protein Gene

The PRD1 gene coding for the DNA terminal protein (gene 8) is mapped at the left end of the PRD1 genome, as shown in Figure 8. The direction of early transcription has been determined to be left to right (Savilahti and Bamford, 1987). The nucleotide sequence of the left end of the PRD1 DNA (1,046 bp) and the deduced amino acid sequence are diagrammed in Figure 9. The first 150 bp of the DNA sequence, which contains the 111 bp terminal inverted repeat previously described (Gerendasy and Ito, 1987), is included here for clarity.

In agreement with genetic studies by Mindich and coworkers (McGraw et al., 1983; Mindich and McGraw, 1983), nucleotide sequence data reveals that there is only one open reading frame that is of sufficient length to encode the entire PRD1 terminal protein. This open reading frame is 780 nucleotides long and ends in an ochre termination codon TAA. A protein encoded by this open reading frame is 259 amino acids in length and has a calculated molecular weight of 29,326 daltons, a value which is in good agreement with molecular weight estimates from SDS/polyacrylamide gel (Mindich et al., 1982). The initiation codon of the PRD1 terminal protein is preceded by an identifiable

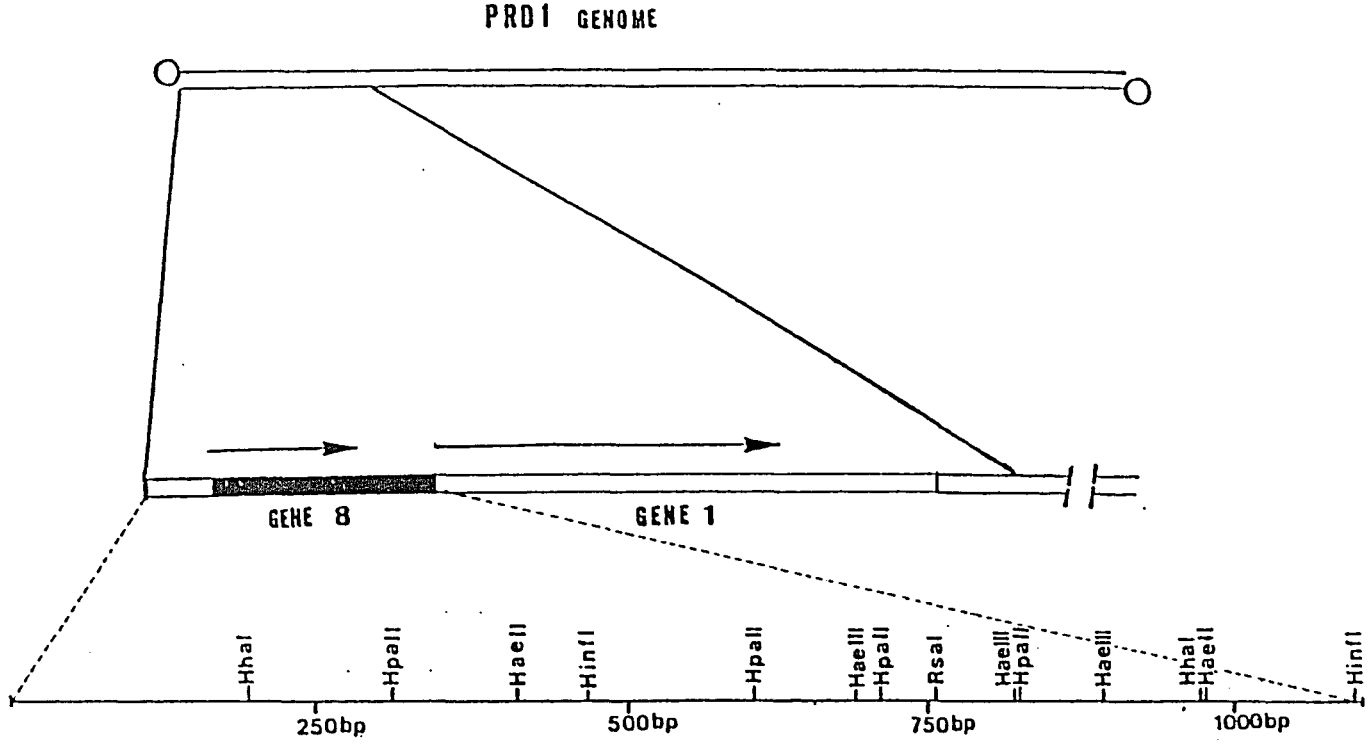


Figure 8. Genetic and physical maps of the region of PRD1 which contains the genes for terminal protein (gene 8) and for DNA polymerase (gene 1). The genetic map was adapted from that of Mindich et al., (1983). Arrows indicate the direction of translation. Open circles at both 5' ends of the genome indicate terminal protein.

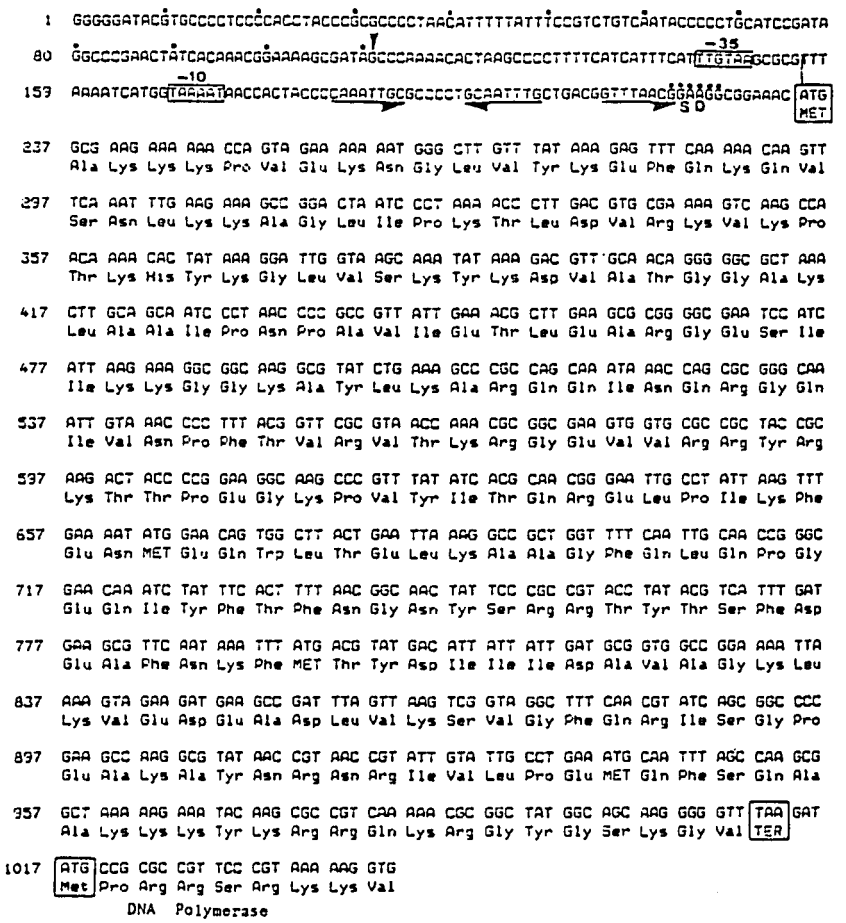


Figure 9. The nucleotide sequence and predicted amino acid sequence of the PRD1 terminal protein gene. Initiation and termination codons for the PRD1 terminal protein and initiation codon for the PRD1 DNA polymerase are boxed. Putative promoter sequences "-35 region" and "-10 region" are also boxed. Shine-Dalgarno sequences are dotted. The vertical arrow point is the end nucleotide of the inverted terminal repeat sequences. The horizonal arrows indicate the unique repeat sequences. Numbers indicate nucleotide positions from the left end of the PRD1 genome.

Shine-Dalgarno ribosome binding site. A presumptive RNA polymerase recognition sequence "-35 region" can be found at nucleotides 145-150 (TTGTAA) and a "-10 region" at nucleotides 169-174 (TAAAAT). Because PRD1 infects a wide variety of gram-negative bacteria, it was of interest to find out what kind of promoter sequences the phage might have. Although these putative promoter sequences must be confirmed by S1 mapping experiments, these nucleotide sequences resemble consensus E. coli promoter sequences (Pribnow, D., 1975).

Between the "-10 region" and the initation codon, there are two palindromic sequences that can form various secondary structures (Figure 10). The biological significance of these sequences is not known, but they may represent transcriptional or translational regulatory signals. Since there is neither an obvious transcriptional termination site at the end of the terminal protein gene, nor promoter sequences for the DNA polymerase gene, it seems likely that a polycistronic mRNA specifies both PRD1 terminal protein and DNA polymerase. It will be interesting to elucidate the regulatory mechanism by which the synthesis of these two proteins is controlled.

An unusual feature of the nucleotide sequence of the PRD1 terminal protein gene is the consecutive runs of the single base adenine. There are 7A, 6A, 5A, 4A and 3A sequences. These poly A sequences run 1, 3, 6, and 20

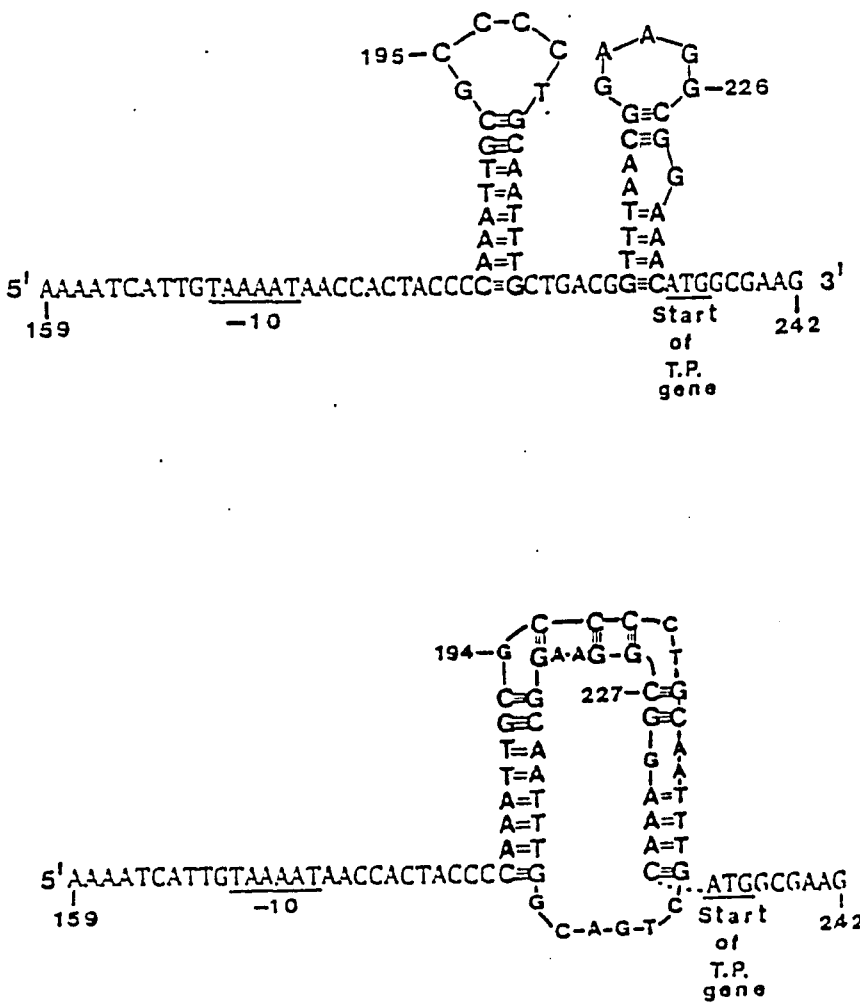


Figure 10. Possible secondary structures at the 5' flanking nucleotide sequence of the PRD1 terminal protein gene.

times, respectively. These sequences are attributed to 27 lysine residues in the PRD1 terminal protein.

Phosphate Box

There are several E. coli genes involed in the transport and metabolism of phosphate which are expressed coordinately when the cells are starved for inorganic phosphate. These include phoA, phoS, phoE and unpB, which encode alkaline phosphatase, phosphate-binding protein, outer membrane porin protein, and sn-glycerol-3-phosphate binding protein, respectively (Makino et al., 1986). Recently, a consensus "phosphate box" nucleotide sequence was identified at the upstream regions of the phoA, phoB, phoE and phoS genes with the phoB gene encoding a positive regulator for these genes (Makino et al., 1986). The phosphate box is probably the recognition site for the phoB protein. In comparison with the upstream nucleotide sequence of the PRD1 terminal protein gene, there is a similar sequence with 86% homology to the phosphate box located between the -52 and -38 regions (Figure 11). The phosphate box-like sequence may also exist in other terminal protein genes. The exact function of the phosphate box in the PRD1 genome is not known at present. It remains to be seen as to what extent this homology is biologocally important.

PRD1 terminal protein gene	(-52)	TTTTCAT*CAT*T*TCAT (-38)
Consensus sequence of phosphate box		CTTTCATAAATCTGTCAT G T A C
E. coli phoA gene	(-41)	CTGTCATAAAGTTGTCAC (-22)
E. coli phoS1 gene	(-63)	CTGTCATAAAACTGTCAT (-44)
E. coli phoS2 gene	(-41)	CTTACATATAACTGTCAT (-22)
E. coli phoE gene	(-41)	CTGTAATATATCTTTAAC (-22)
E. coli phoB gene	(-41)	TTTTCATAAATCTGTCAT (-22)
Zymomonas mobilis phoC gene	(-40)	TTGTCTTATTATAGCCAC (-23)

Figure 11. Comparison of the phosphate box sequences of phoA, phoS, phoE, phoB, phoC and PRD1 terminal protein genes. The position of the transcription initiation site of each gene denotes +1.

Amino Acid Sequence of The PRD1 Terminal Protein

When the deduced amino acid sequence of the PRD1 terminal protein was compared with those of terminal proteins from ϕ29, Nf and adenovirus, no overall amino acid sequence homology was found. On the basis of their different hosts, Gram negative (PRD1) vs. Gram positive (ϕ29), it seemed of interest to determine differences between PRD1 and ϕ29 terminal proteins. Comparison of the amino acid composition of the terminal proteins of PRD1 and ϕ29 are shown in Table 4. The ϕ29 terminal protein contains 7 additional amino acid residues (Shih et al., 1984). The most significant difference is the presence of 22 additional acidic amino acid residues (Asp, Glu) in the ϕ29 terminal protein. On the other hand, the PRD1 terminal protein contains 10 additional basic amino acid residues. Clearly, the PRD1 terminal protein has a greater net positive charge. It is remarkable that more than 60% of the aspartic acid and 46% of the glutamic acid residues are present in the amide form in the PRD1 terminal protein. This is in contrast to the ϕ29 terminal protein in which only 34% of the glutamic acid residues are present in the amide form. Neither protein contains cysteine residues.

The distribution of the amino acid residues over the polypeptide chain of the PRD1 terminal protein exhibits some interesting features (Figure 13). The PRD1 terminal protein has two strongly basic regions. At the amino-

Table 4. Predicted amino acid composition of terminal protein

Amino acid	Comparison Numbers in PRD1		Numbers in ϕ29	
Basic side chains				
Lys	36		25	
Arg	18	(21.2%)	18	(16.9%)
His	1		2	
Acidic side chains				
Asp	7		19	
Glu	17	(9.2%)	27	(17.3%)
Uncharged polar side chains				
Gly	20		11	
Asn	11		17	
Gln	15		14	
Cys	0	(31.6%)	0	(31.2%)
Ser	9		20	
Thr	14		10	
Tyr	13		11	
Nonpolar side chains				
Ala	20		14	
Val	20		12	
Leu	15		17	
Ile	15		18	
Pro	12	(37.8%)	8	(34.5%)
Phe	11		15	
Met	4		6	
Trp	1		2	
Tolal	259		266	

terminus, there are 17 positively charged residues (Arg, Lys, His) in the first 53 residues (about 32%). At the carboxyl-terminus, there are 9 positively charged amino acids in the last 17 residues (about 53%). The rest of the polypeptide chain contains about 13% basic amino acid residues. The calculated pI value of the PRD1 terminal protein is 9. This value makes it the most basic terminal protein yet identified.

Codon Usage

Codon frequencies in the terminal protein genes of the PRD1 and ϕ29 are summarized in Table 5. Comparison of codon usage reveals some striking differences between the terminal proteins. Most notable is the use of the glycine codon GGC by PRD1. Of the 20 glycine codons present in the PRD1 terminal protein, GGC is used 12 times (60%). On the other hand, ϕ29 terminal protein does not use GGC at all, even though it contains 11 glycine residues. This can be explained by the fact that the ϕ29 DNA does not contain the GGCC sequence perhaps by evolutionary avoidance of the restriction site (GGCC) of Bsu I endonucleases which are produced by its host, Bacillus subtilis (Ito and Roberts, 1975). In addition, the PRD1 terminal protein gene shows a strong bias toward GCC as its arginine codon, and CAA as its glutamine codon.

Table 5. Codon usage in terminal protein genes

	PRD1	Ø29		PRD1	Ø29		PRD1	Ø29		PRD1	Ø29
TTT/Phe	9	9	TCT/Ser	0	1	TAT/Tyr	11	9	TGT/Cys	0	0
TTC/Phe	2	6	TCC/Ser	2	2	TAC/Tyr	2	4	TGC/Cys	0	0
TTA/Leu	3	5	TCA/Ser	2	4	TAA/TER	1	0	TGA/TER	0	0
TTG/Leu	5	4	TCG/Ser	1	0	TAG/TER	0	1	TGG/Trp	1	0
CTT/Leu	5	2	CCT/Pro	4	5	CAT/His	0	1	CGT/Arg	5	5
CTC/Leu	0	1	CCC/Pro	4	1	CAC/His	1	1	CGC/Arg	10	1
CTA/Leu	1	2	CCA/Pro	2	1	CAA/Gln	12	3	CGA/Arg	1	4
CTG/Leu	1	2	CCG/Pro	2	1	CAG/Gln	3	7	CGG/Arg	2	3
ATT/Ile	8	7	ACT/Thr	3	1	AAT/Asn	4	8	AGT/Ser	0	5
ATC/Ile	6	2	ACC/Thr	4	2	AAC/Asn	7	5	AGC/Ser	4	6
ATA/Ile	1	9	ACA/Thr	2	8	AAA/Lys	22	11	AGA/Arg	0	5
ATG/MET	4	7	ACG/Thr	5	1	AAG/Lys	14	16	AGG/Arg	0	2
GTT/Val	8	3	GCT/Ala	3	8	GAT/Asp	4	10	GGT/Gly	1	6
GTC/Val	1	1	GCC/Ala	7	2	GAC/Asp	3	8	GGC/Gly	12	0
GTA/Val	7	4	GCA/Ala	3	4	GAA/Glu	16	19	GGA/Gly	3	3
GTG/Val	4	4	GCG/Ala	7	5	GAG/Glu	1	8	GGG/Gly	4	2

Numbers indicate the frequency of codons used in the terminal protein genes.

Hydropathy

It is known that linkage of PRD1 terminal protein to the DNA is a phosphodiester bond between tyrosine and a dGMPof the DNA (Bamford and Mindich, 1984). Nucleotide sequence data predicts the existence of 13 tyrosine residues in the PRD1 terminal protein. To obtain information on which tyrosine residue would likely be the linking amino acid to the DNA, we determined the hydropathic profile of the PRD1 terminal protein following the algorithm of Kyte and Doolittle (1982). When the hydropathic profiles of PRD1, ϕ29, and Nf terminal proteins were compared, there seemed to be no similarity at first glance. However, when 7 residue gaps were introduced into the PRD1 terminal protein, the three profiles were superimposable at the C-terminal portion (residues 210-260) of the terminal protein (Figure 12). This similarity suggests that the carboxyl termini of the three terminal proteins are functionally homologous.

Secondary Structure

The secondary structure of PRD1 terminal protein was predicted by the empirical method of Chou and Fasman (1978). PRD1 terminal protein has a 37% α-helical, 24% ß-pleated sheet, 14% ß-turn and 25% random coil structure (Figure 13). Distribution of the α-helical and ß-sheet structures is not

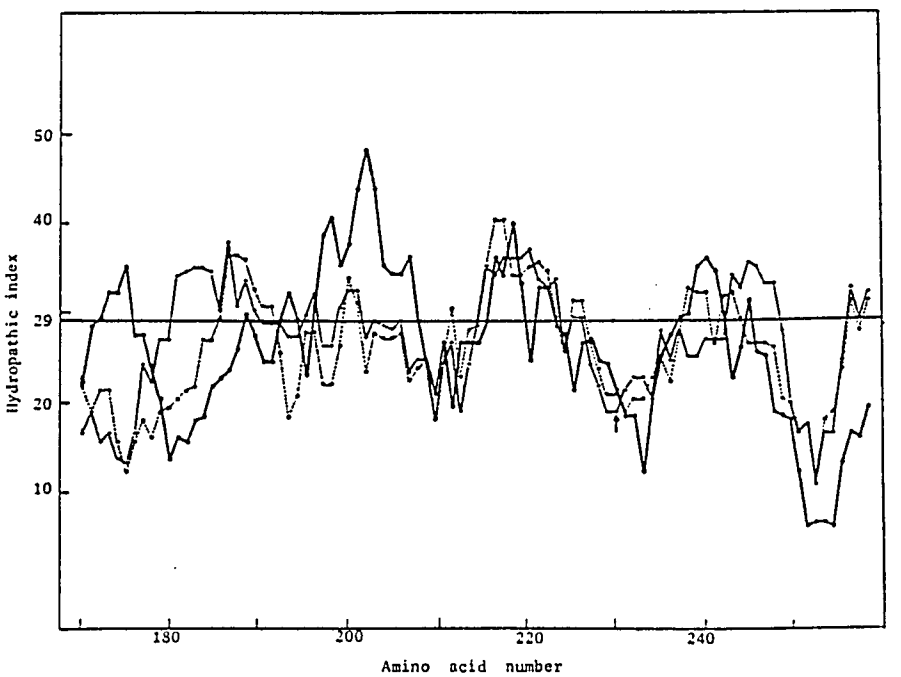


Figure 12. Hydropathic profiles of C-terminal regions of the PRD1, Φ29 and Nf terminal proteins. PRD1 (———), Φ29 (———), and Nf (----). The arrow is the linking site of Φ29 terminal protein.

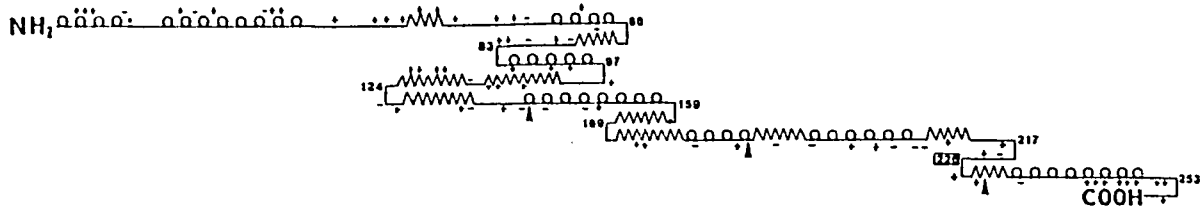


Figure 13. The predicted secondary structure of the PRD1 terminal protein. ( ), ( ), and (—) denote alpha-helical, beta-sheet, and random-coil residues, respectively. Beta-turns are indicated by chain reversal. The + and - signs refer to charged residues, and the numbers refer to residues in the terminal protein.

symmetrical: the N-terminal one-third of the protein, is predicted to contain about 37% α-helical and 10% ß-sheet, while the C-terminal one-third contains 44% α-helix and 20% ß-sheet.

The predicted secondary structure of the PRD1 terminal protein contains two antiparallel ß-ß dimer structures at residues 116-136 and 163-180. The ß-ß dimers have been suggested to form a polypeptide double helix that could play a role in nucleic acid binding (Carter and Kraut, 1974; Chou and Fasman, 1978).

Possible Evolutionary Relationship Between Poliovirus and PRD1 Terminal Protein

The poliovirus terminal protein is a short peptide of 22 amino acid residues (Kitamura et al., 1981). When the amino acid sequences of PRD1 and poliovirus terminal proteins were compared, a conserved sequence was noted around nucleotide-linking site of poliovirus terminal protein. To determine if this conserved sequence, Ala-Tyr-$(A.A.)_7$-Pro-$(A.A.)_7$-Ala-Lys, occurs randomly in other proteins, more than 4,750 protein sequences were searched using a FASTP microcomputer program. Only one other protein has this conserved sequence. This protein is poliovirus type 3 terminal protein (Figure 14). Considering the rapid evolution of RNA viruses, such high conservation is

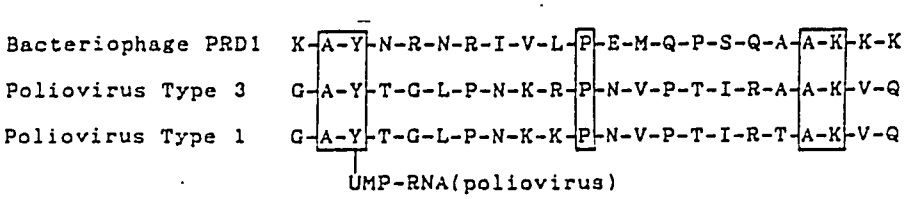


Figure 14. The conserved sequences of PRD1 terminal protein and VPgs of polioviruses. Homologous amino acid sequences are boxed.

remarkable. This could be the product of strong selective pressure on this region of these proteins during evolution.

Highly Conserved Sequence, YSRLRT, on All Available Terminal Proteins and DHBcAg

As described above (page 52), the deduced amino acid sequence of phage PRD1 terminal protein revealed no overall homology with other reported terminal proteins or related proteins. However, closer examination revealed a short patches of six identical amino acid sequences, YSRLRT, present in all identified terminal proteins such as PRD1, Nf, PZA, ϕ29 and adenovirus (Figure 15). In addition, this short patch sequence can also be found in the core antigen of duck hepatitis B virus (DHBcAg).

Another interesting feature is that they are located at approximately the same position in the functionally related terminal proteins. The conserved spatial locations of YSRLRT sequences of all known terminal proteins and DHBcAg are given in Figure 16. All are similarly spaced on the central portion of the whole molecule, except for the precursor of adenovirus terminal protein, which is located close to the amino terminal region (residues 101-107). The probability of such a chance occurence of the conserved sequence YSRLRT would be $(1/20)^{6} = 1.56 \times 10^{-8}$. The significance of these sequences of YSRLRT is

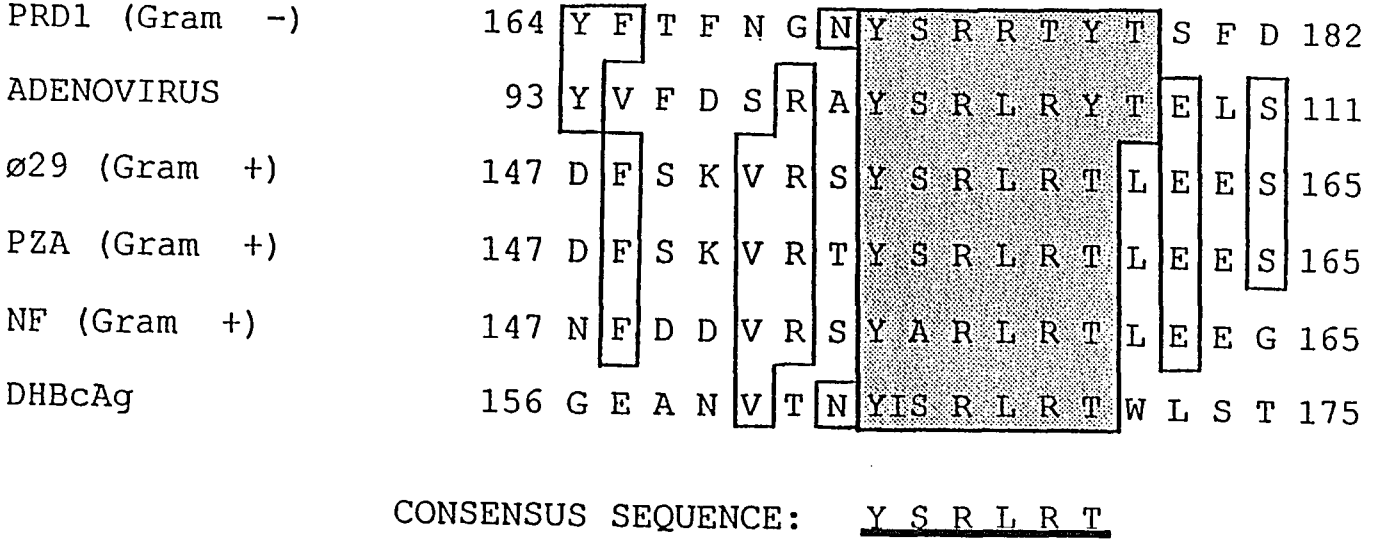


Figure 15. Conserved regions (shaded areas) of terminal proteins and DHBcAg.

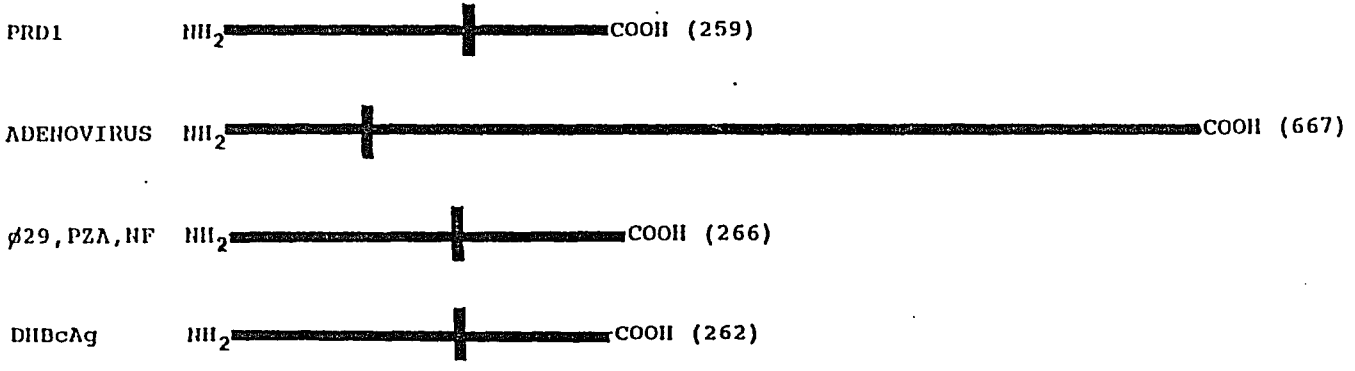


Figure 16. Spatial relationship of the conserved sequence YSRLRT among the terminal proteins and DHBcAg. The conserved sequence of each polypeptide is indicated by a vertical line. Terminal protein and DHBcAg sizes are proportional to the length of each line. The total amino acid residues in each polypeptide is shown in parentheses.

suggested by (i) its conservative spatial location on each respective polypeptide and (ii) its high degree of homology across species. Although functional significance of this YSRLRT sequence is not known, it can be surmised that such sequence conservation is a consequence of their contribution to the terminal protein function. Therefore, it is reasonable to propose that these conserved sequences may be somehow involved in priming function of linear DNA replication.

Cloning and Overexpression of The PRD1 Terminal Protein and DNA Polymerase Genes

Thus far, the unique PRD1 terminal protein alone has never been successfully cloned into expression vectors. However, a 3kb DNA fragment that contains PRD1 terminal protein and DNA polymerase genes was able to cloned into expression vectors, pT7-6 and pEMBLex3. The expression vector pT7-6 is a part of the T7 RNA polymerase/promoter system which contains two compatible plasmids. The first plasmid pT7-6 was used for cloning the 3kb DNA fragment, while the second plasmid pGP1-2 was used to facilitate the expression of inserted genes. Plasmid pGP1-2 contains the T7 RNA polymerase (under the control of the inducible lamda pL promoter), and the temperature-sensitive cI857 repressor. The cloning strategy is shown in Figure 5. This system

exploits the unique specificity of T7 RNA polymerase for its own promoter, ϕ10, to control the expression of inserted genes. The exclusive expression of inserted genes is achieved after heat induction of T7 RNA polymerase of pGP1-2 and addition of rifampicin to shut off the host E. coli RNA polymerase transcription. Afterwards, the addition of [35-S]methionine results in specific labeling of plasmid-encoded proteins. An autoradiogram of a gel electrophoresis of this experiment is shown in Figure 17. This T7 RNA polymerase/promoter system has been utilized for the purification of the PRD1 terminal protein and for the DNase I footprinting experiments.

In order to gain insight into the functional role of the conserved sequence YSRLRT in PRD1 terminal protein, site-directed mutagenesis was carried out to generate mutants. An expression vector which contains an f1 origin for making ss-DNA and an inducible strong promoter were used for the following experiments. The expression vector, pEMBLex3, was chosen and the PstI-3kb DNA fragment was cloned into the PstI cloning site. The pEMBLex3 carries a temperature-sensitive repressor, cI857, so that transcription from lamda pR is repressed at low temperature (30°C) and can be activated at high temperature (42°C). Cells containing the recombinant plasmid pEMBL 3K were grown at 30 °C to 0.45 OD_{590} and then shifted rapidly to 42°C. After a 2 hour induction, the cells were labeled with [35-S]methionine and analyzed by SDS/Polyacrylamide

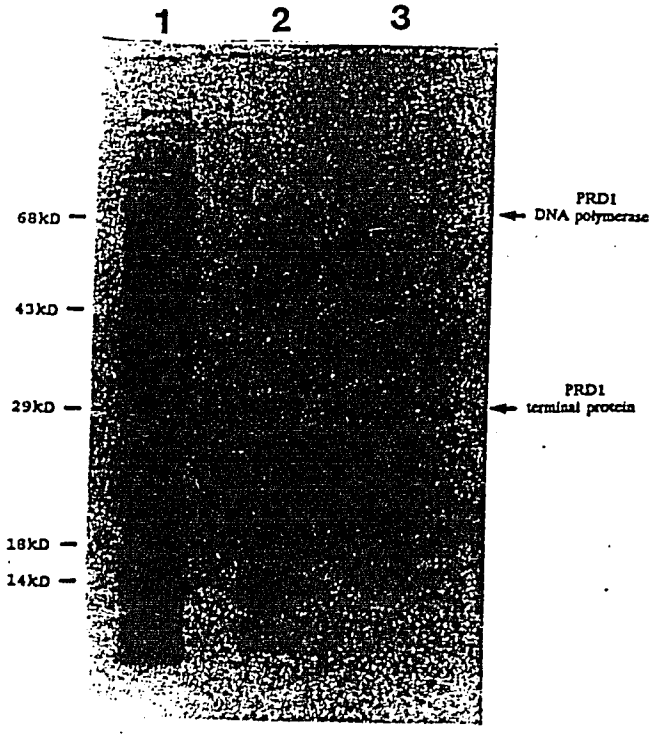


Figure 17. Inclusive expression of the PRD1 terminal protein and DNA polymerase genes from a T7 RNA polymerase/promoter. Lane 1 the cells were induced by shifting the temperature to 42 C for 30 min. In lane 2 rifampicin was added after induction for 10 min at 42 C, followed by 20 min at 30 C, prior to pulse labeling. Lane 3 the T7 expression plasmid (pT7-6) without insert DNA is used as a negative control.

gel. The comparison of uninduced and induced PRD1 terminal protein and DNA polymerase from pEMBL 3K was shown in Figure 18. The cell extract from the induced pEMBL 3K clone was prepared for the assay of complex formation activity. Comparison of complex formation activity between the recombinant pEMBL 3K and non-recombinant pEMBLex3 clones is shown in Figure 19. The pEMBL 3K clone exhibited a 29kd P8-dGMP complex band. These results suggest that the gene products of the PstI-3kb inserted fragment are functional in vitro. This functional pEMBL 3K clone was used for in vitro site-directed mutagenesis.

Site-Directed Mutagenesis on The Conserved Sequence YSRRT of PRD1 Terminal Protein

The conserved sequence in PRD1 terminal protein is YSRRT, which is one leucine residue less than the typical consensus sequence YSRLRT. Each amino acid residue of the YSRRT sequence was mutated using the Kunkel's method. Totally, seven 29-mer and three 27-mer synthetic oligonucleotide primers were used in experiments to change 747(T→A), 749(T→A), 750(T→G), 753(C→G,A), 754(G→A,T), 756(C→G), 757(G→A), and 759(A→G) of PRD1 terminal protein gene, converting tyrosine-172 to asparagine and ochre termination codon, serine-173 to alanine, arginine-174 to glycine, histidine,

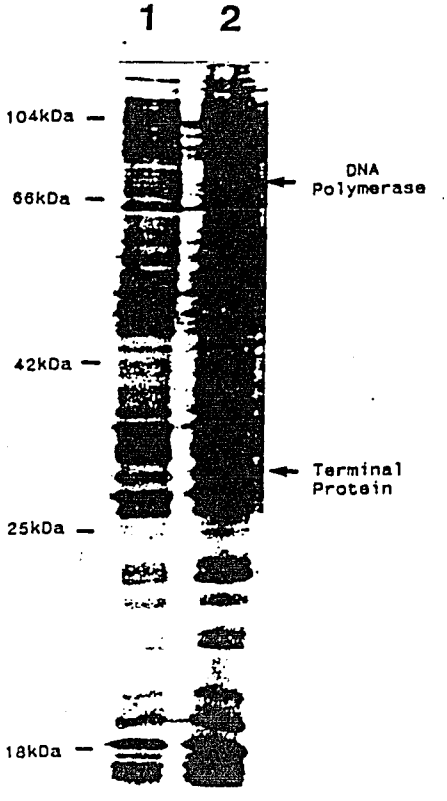


Figure 18. Autoradiogram of an SDS/Polyacrylamide gel of 35-S pulse-labeled proteins synthesized from the pEMBL 3K uninduced (lane 1) and induced cells (lane 2).

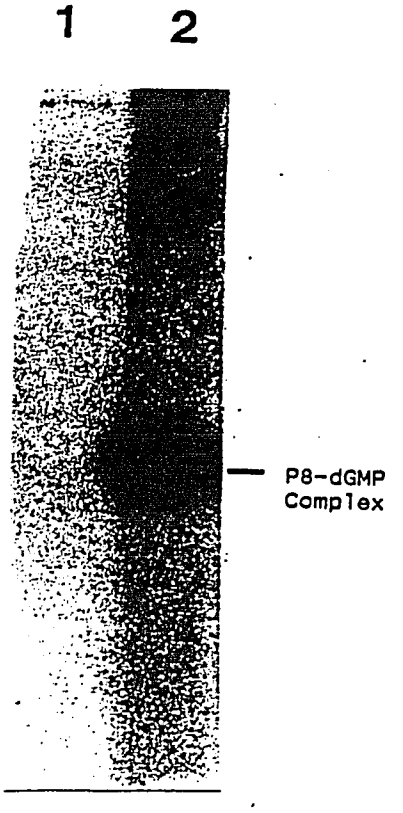


Figure 19. In vitro formation of the P8-dGMP initiation complex from the pEMBLex 3 clone (lane 1) and the pEMBL 3K clone (lane 2).

leucine and serine, arginine-175 to glycine and histidine, and threonine-176 to alanine, respectively. This uracil template method produced PRD1 terminal protein mutants with a frequency of about 75%-100%. These terminal protein mutants were denoted as Y172N, Y172O, S173A, R174G, R174H, R174L, R174S, R175G, R175H, and T176A, respectively. The DNA sequences of most these mutants are shown in Figure 20-25. All of these mutant clones were induced and their cell extracts were prepared for examining priming complex formation.

In the initial stage, seven PRD1 terminal protein mutant clones (i.e., Y172N, S173A, R174G, R174L, R174S, R175G and T176A) were tested. The significant finding was that the arginine-174 lacking mutant clones abolished the priming complex formation. The Y172N and S173A mutant clones exhibit reduced activities, whereas T176A clone shows normal activity in complex formation. The R175G clone, in contrast with other mutant clones, has an increased activity (Figure 26). These different responses of amino acid residue changes on the YSRRT conserved sequence of the PRD1 terminal protein may be due to the change of original catalytic group, causing perturbation of the overall structure of PRD1 terminal protein molecule. However, the results have demonstrated that the arginine-174 of the conserved sequence YSRRT is essential and important for performing the native function of PRD1 terminal

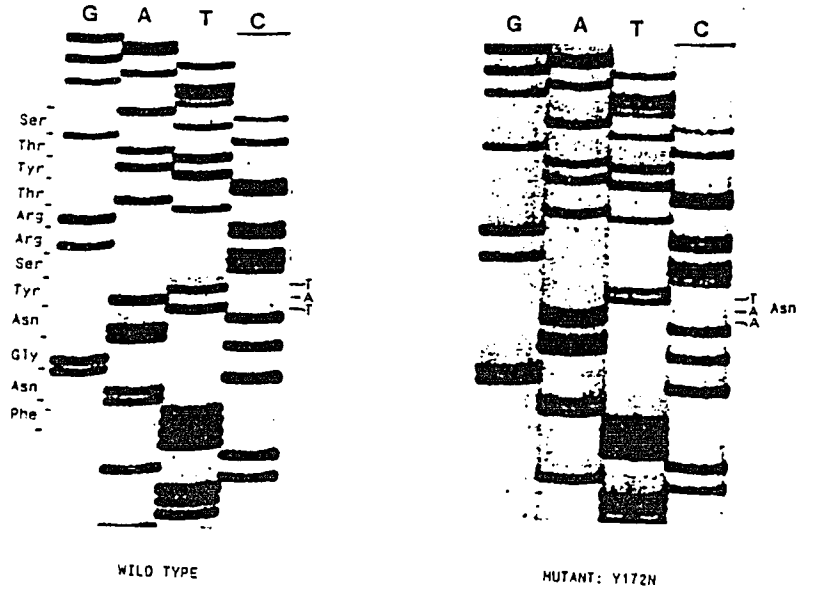


Figure 20. DNA sequence of terminal protein mutant (Y172N) and wild-type clones in the mutated region.

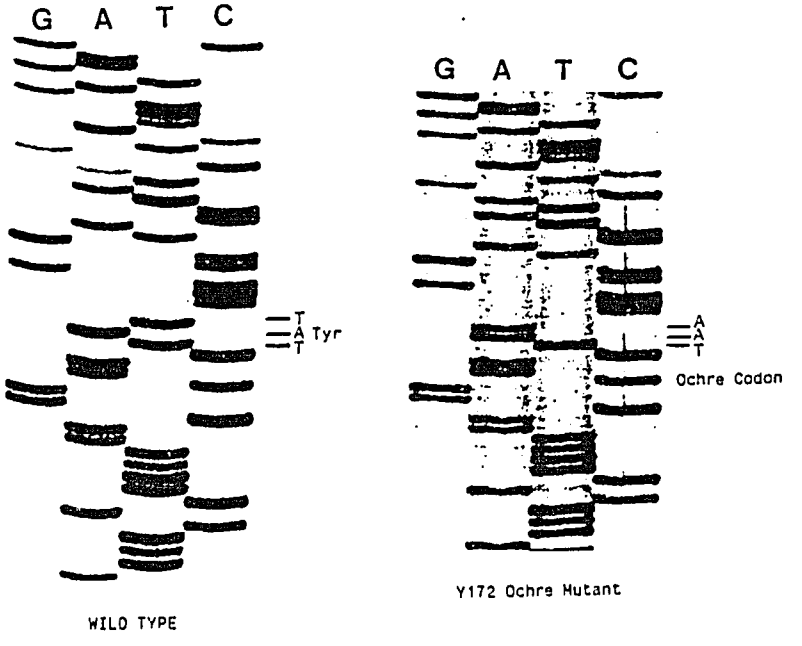


Figure 21. DNA sequence of terminal protein mutant (Y172Ochre) wild-type clones in the mutated region.

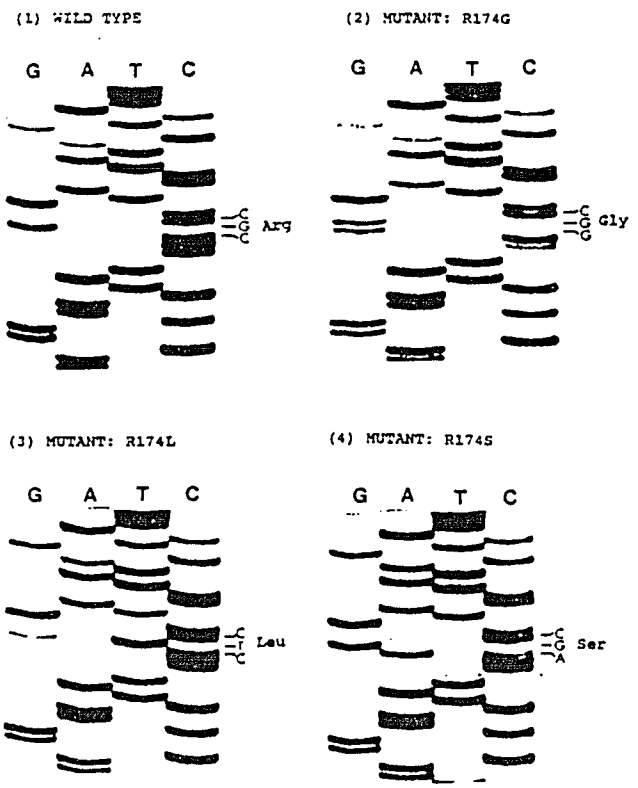


Figure 22. DNA sequence of terminal protein mutant (R174G, R174L and R174S) and wild-type clones in the mutated region.

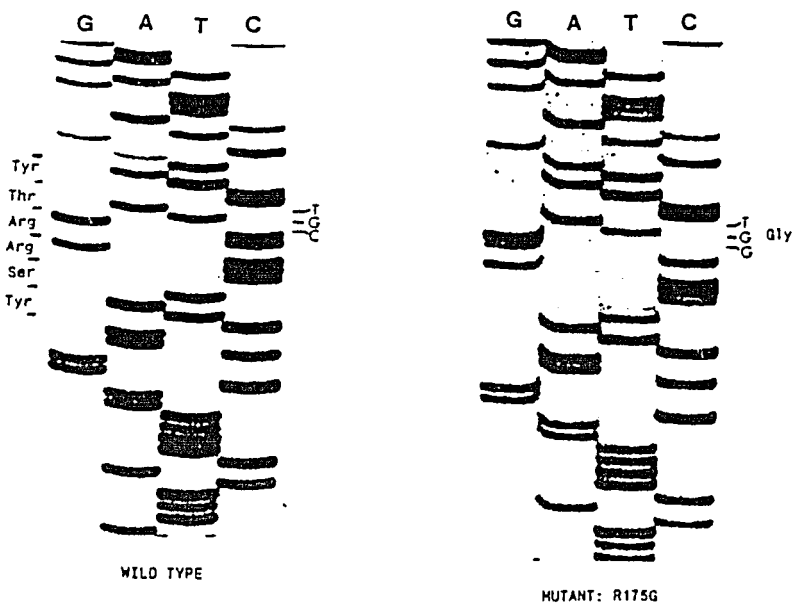


Figure 23. DNA sequence of terminal protein mutant (R175G) and wild-type clones in mutated region.

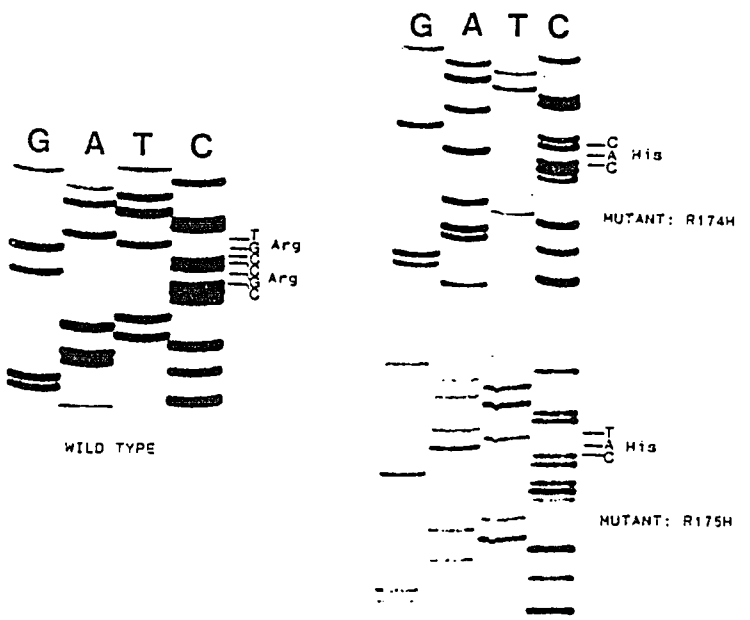


Figure 24. DNA sequence of terminal protein mutant (R174H and R175H) and wild-type clones in the mutated region.

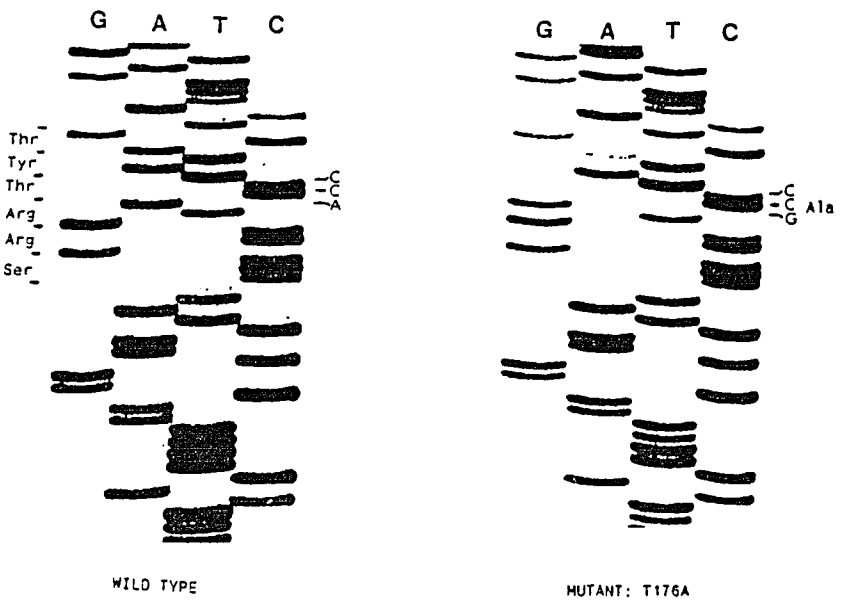


Figure 25. DNA sequence of terminal protein mutant (T176A) and wild-type clones in mutated region.

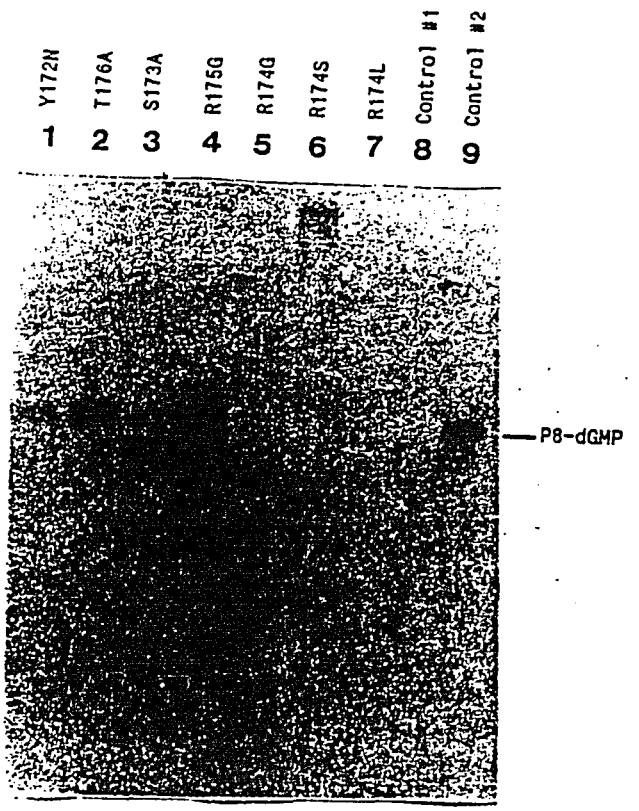


Figure 26. P8-dGMP complex-forming activity of PRD1 terminal protein mutant and wild-type clones. Reactions were processed as in MATERIALS and METHODS. Lanes: 8, plasmid pEMBLex3 without 3kb insert (negative control); 9, plasmid pEMBL 3K with 3kb insert (positive control).

protein.

Another interesting finding is that R174G, R174L and R174S inhibit host cell growth. The measured doubling time for each mutant clone is shown in Table 6. The reason for causing the longer doubling time for the arginine-174 lacking clones is not understood.

In order to further understand whether the positive charge of position 174 of PRD1 terminal protein molecule might be critical to the PRD1 terminal protein function, the arginine-174 was replaced with histidine. In addition, another PRD1 terminal protein mutant, R175H, was constructed to examine the effect at position 175 of the molecule. The results of priming complex formation of R174H and R175H are shown in Figure 27. The R175H exhibits normal activity, whereas the R174H shows a reduced priming activity. The results strongly demonstrated that the positive charge residue of arginine-174 plays a critical role in PRD1 terminal protein function, but not the positive charge residue of arginine-175. Finally, an ochre mutant at tyrosine-172 position was constructed to see whether the YSRRT conserved region is resposible for inhibiting host cell growth. The doubling time of the Y172Ochre clone was normal. Hence, it is clear that the arginine-174 residue is responsible for the inhibition of host cells. The characteristics of the PRD1 terminal protein mutant clones are shown in Table 7.

Table 6. Effect of various mutants of PRD1 terminal protein on the growth of E. coli NM522

Strain	Doubling time (min)
TPY172Ochre	60
TPY172N	51
TPS173A	51
TPR174G	112
TPR174L	108
TPR174S	114
TPR174H	67
TPR175G	68
TPR175H	54
TPT176A	57
TPY177F	60
TPY190F	52
TPY226F	51
TPY246F	51
TPY254F	53
Control 1(TP+DP)	51
Control 2(no TP+DP)	54
Control 3(DP only)	54

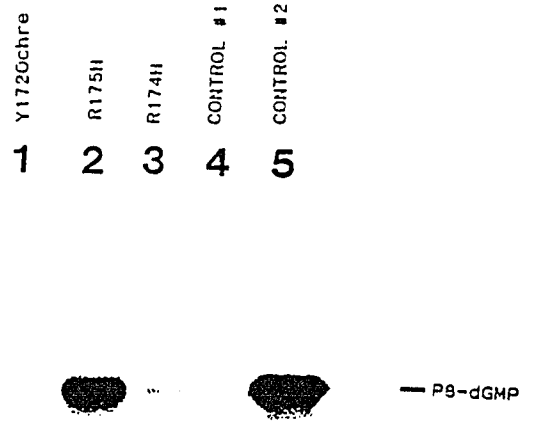


Figure 27. P8-dGMP complex-forming activity of PRD1 terminal protein mutant and wild-type clones. Reactions were processed as in MATERIALS & METHODS. Lanes: 4, plasmid pEMBLex3 without 3kb insert (negative control); 5, plasmid pEMBL 3K with 3kb insert (positive control).

Table 7. The characteristics of PRD1 terminal protein mutant clones.

	168 175 182	P8-dGMP complex formation	Host Cell Growth
WILD TYPE	F N G N Y S R R T Y T S F D E	++	Normal
Y172Ochre	- - - - * - - - - - - - - - -	-	Normal
Y172N	- - - - N - - - - - - - - - -	+	Normal
S173A	- - - - - A - - - - - - - - -	+	Normal
R174G	- - - - - - - G - - - - - - -	-	Inhibited
R174L	- - - - - - - L - - - - - - -	-	Inhibited
R174S	- - - - - - - S - - - - - - -	-	Inhibited
R174H	- - - - - - - H - - - - - - -	+	Normal
R175G	- - - - - - - - G - - - - - -	+++	Normal
R175H	- - - - - - - - H - - - - - -	++	Normal
T176A	- - - - - - - - - A - - - - -	++	Normal
Y177FS179T	- - - - - - - - - - F - T - - -	++	Normal

Location of the Tyrosine Residue Involved in the Linkage Between the PRD1 Terminal Protein and DNA

In the cases of the phage ϕ29 and adenovirus systems, the linking residues (serine) are located at the carboxyl terminal portion (Hermoso et al., 1985; Smart and Stillman, 1982). By analogy, the tyrosine residues located at the carboxyl terminal portion of PRD1 terminal protein molecule could be a candidate for the linking site. The determination of linking sites of ϕ29 and adenovirus terminal proteins were performed by direct sequencing of the 125-I tryptic peptides. Here, we used site-directed mutagenesis to determine the linking site of PRD1 terminal protein. This approach involved replacing the tyrosine residues with phenylalanine residues in the carboxyl terminal area. Three 29-mer and two 27-mer synthetic oligonucleotides were used in site-directed mutagenesis to change tyrosine codons (TAT, TAC) to phenylalanine codons (TTT, TTC). These PRD1 terminal protein mutants are denoted as Y177F, Y190F, Y226F, Y246F, and Y254F, respectively. The DNA sequences of mutants Y177F, Y226F and Y246F are shown in Figure 28, 29 and 30.

All terminal protein mutant clones were grown and cell extracts were prepared to examine their complex forming activities. In contrast to the previous prediction, the Y190F clone abolished the complex formation activity (Figure 31). This implies that tyrosine-190 is the linking residue of the PRD1 terminal

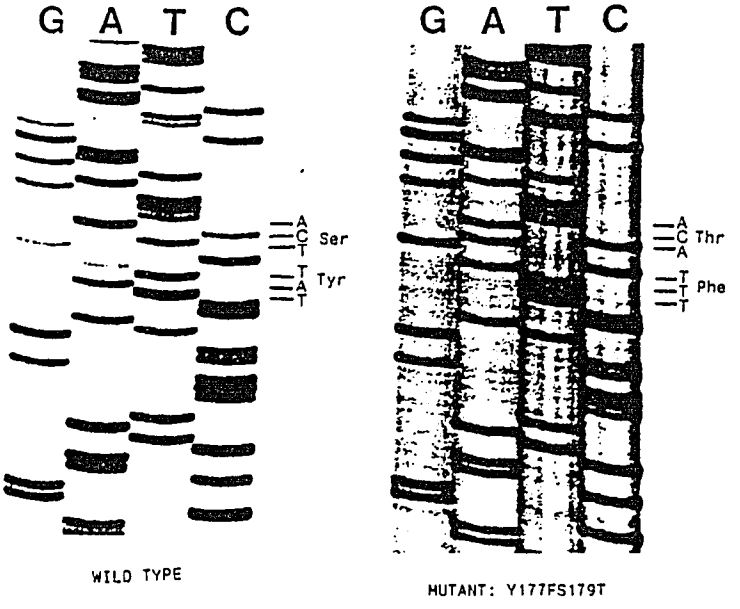


Figure 28. DNA sequence of PRD1 terminal protein mutant (Y177F) and wild-type clones in mutated region.

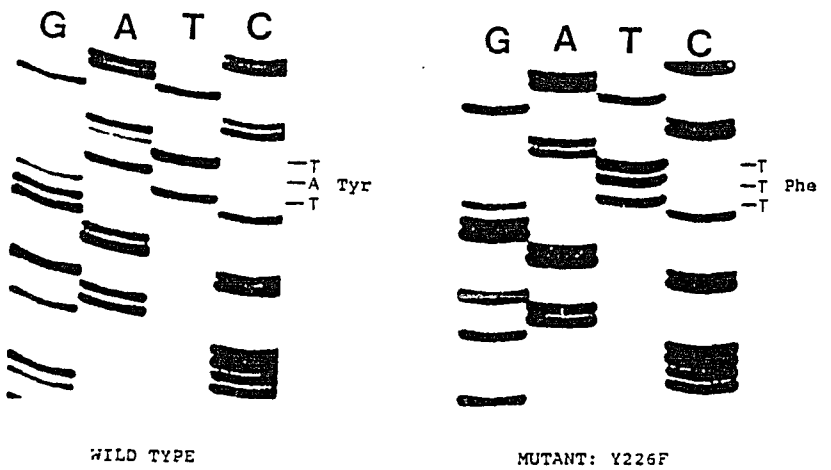


Figure 29. DNA sequence of PRD1 terminal protein mutant (Y226F) and wild-type clones in mutated region.

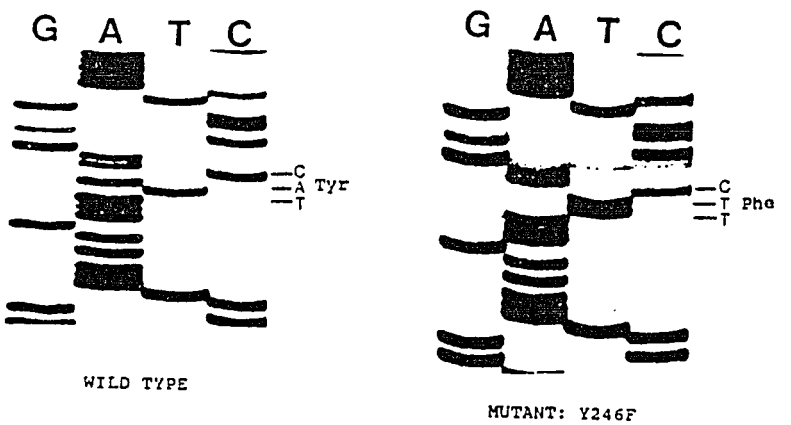


Figure 30. DNA sequence of PRD1 terminal protein mutant (Y246F) and wild-type clones in mutated region.

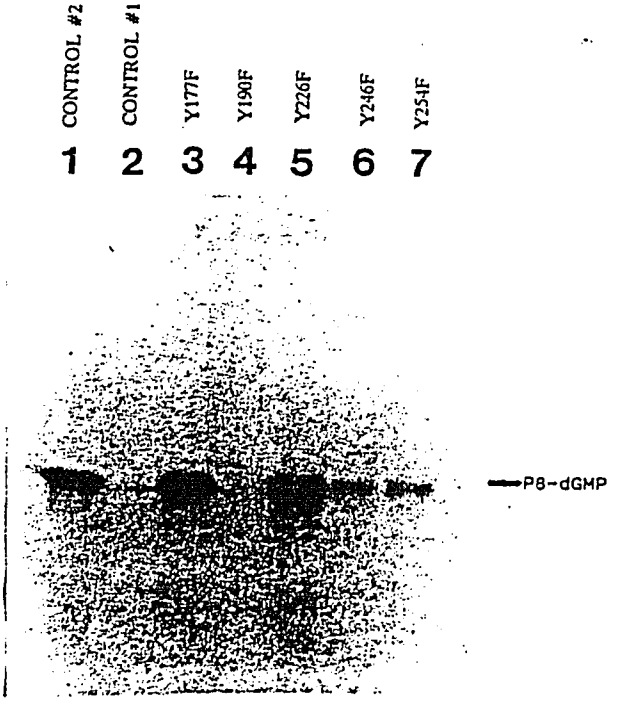


Figure 31. P8-dGMP complex-forming activity of PRD1 terminal protein mutant and wild-type clones. Reactions were processed as in MATERIALS and METHODS. Lanes: 2, plasmid pEMBLex3 without 3kb insert (negative control); 1, plasmid pEMBL 3K with 3kb insert (positive control).

protein to DNA. The Y177F and Y226F mutant clones exhibit normal activities, whereas the Y246F and Y254F clones showed the somewhat decreasing activity. From these results, we concluded that tyrosine-190 is the most likely linking residue of the PRD1 terminal protein.

So far, the linking site has been characterized only in three cases, ϕ29 (Hermoso et al., 1985), adenovirus (Smart and Stillman, 1982) and PRD1 (this dissertation). Comparison of the hydropathic profiles of these three linking site enviroment is shown in Figure 32. Both the linking sites of ϕ29 and adenovirus terminal proteins are located at a hydrophilic region and also probably on the surface of the molecules (Kyte and Doolittle, 1982).

In contrast, the linking site of PRD1 terminal protein is located in a hydrophobic area. This implies that the structural topology of the linking site of PRD1 terminal protein is different from those of ϕ29 and adenovirus terminal proteins. The linking sites of ϕ29 and adenovirus terminal proteins are similarly located at a ß-turn which is preceded by α-helical region, while the linking site of PRD1 terminal protein is located at ß-sheet which also is preceded by α-helical region (Figure 33).

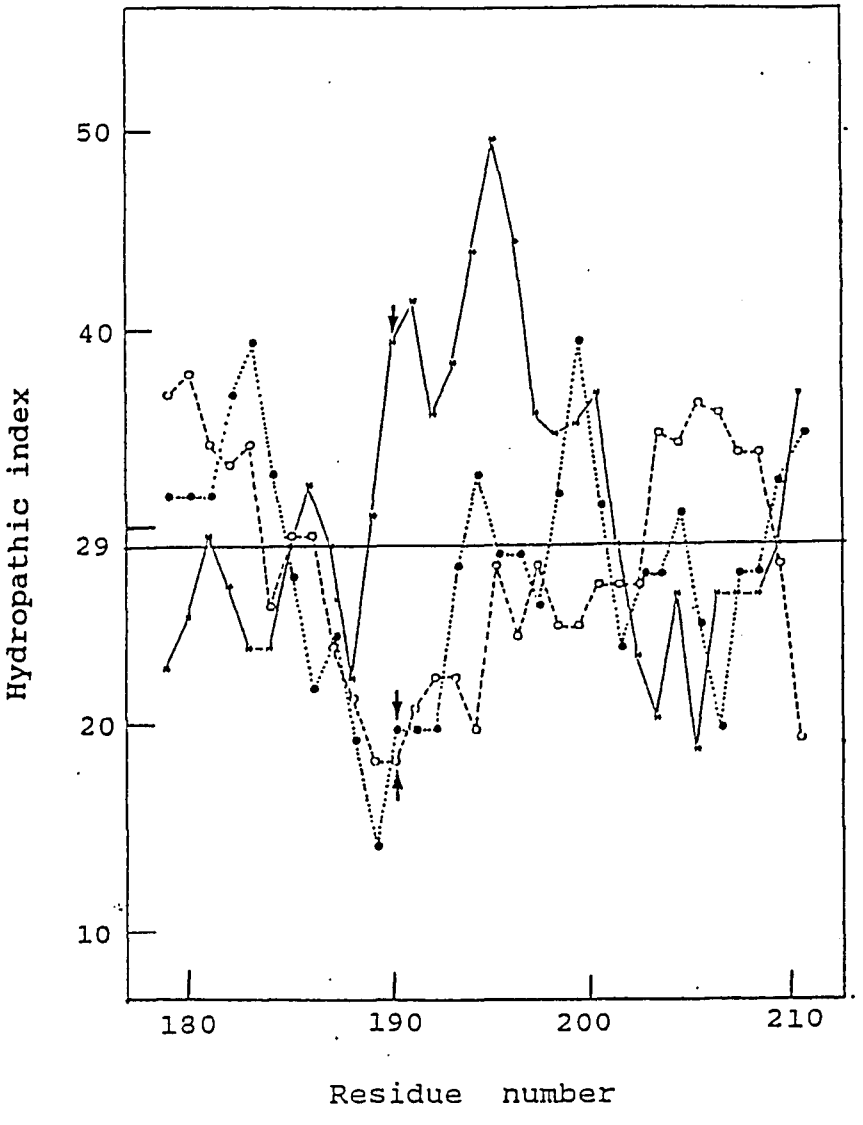


Figure 32. Hydropathic index of a 32 residues long region including the linking site, using a span setting of seven according to Kyte and Doolittle (1982). *—*, PRD1 terminal protein; ●·····●, adenovirus terminal protein; o----o, ϕ29 terminal protein. The arrows indicate the linking site positions.

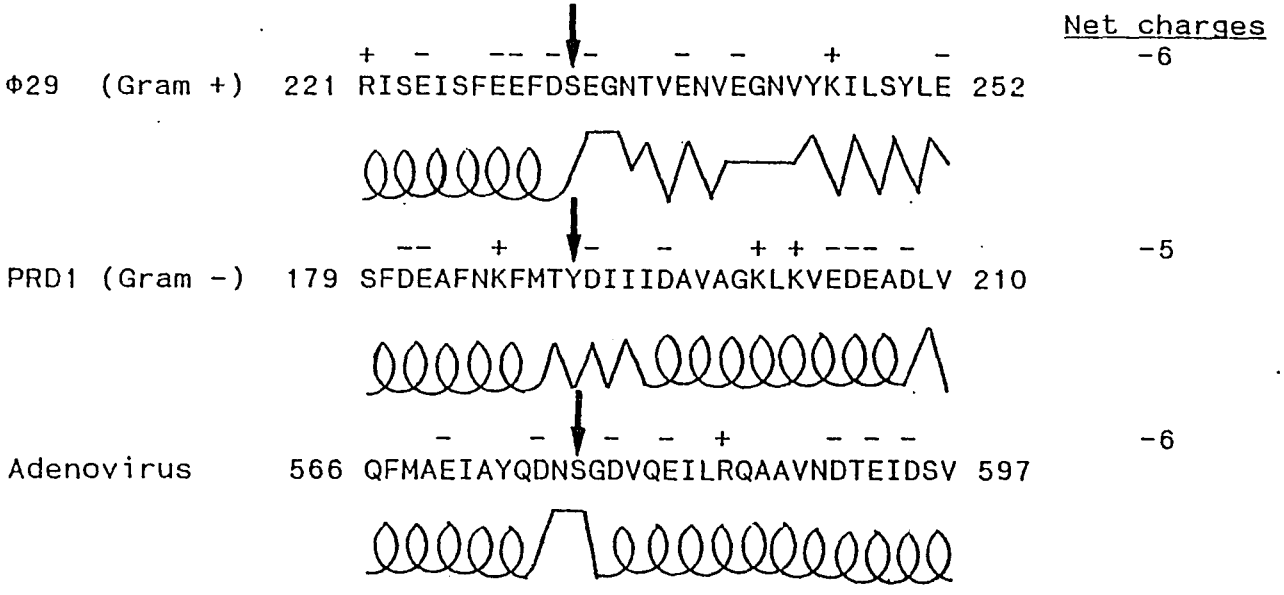


Figure 33. Secondary structure prediction of regions surrounding the linking site in PRD1, ϕ29 and adenovirus terminal proteins. Secondary structure using the Chou and Fasman method (1978) was predicted for a 32 residues long region. (Ω), (/\ ), (/\ ) and (—) denote α-helical, ß-turn, ß-sheet, and random-coil residues, respectively. The arrows indicate the linking site positions. The + and - signs refer to charged residues, and the numbers refer to residues in the respective terminal protein molecule.

CHAPTER 4

DISCUSSION

Sequence Analysis of Terminal Protein and Adjacent Genes

There are three open-reading frames located on the left 3kb region of the phage PRD1 genome which encode terminal protein, DNA polymerase, and lysin, respectively. The first open reading frame (ORF I) begins at nucleotides 234-236 ATG and terminates with a TAA at nucleotides 1011-1013. This ORF I encodes a polypeptide of MW=30 kDa which is in good agreement with previous determinations. The second open reading frame (ORF II) begins at nucleotides 1017-1019 ATG and is followed by a 1662 nucleotides long ORF which terminates with a TAA at nucleotides 2676-2678. ORF II encodes a polypeptide with 553 amino acids. The third open reading frame (ORF III) starts at nucleotide 2679 and ends at nucleotide 3126. ORF III encodes a protein of 149 amino acids with a MW=17.3 kDa which is roughly consistent with 13 kDa as determined by SDS/polyacrylamide gel. The DNA sequence of the left 3.2kb region has been completely confirmed by our laboratory and

Bamford's laboratory in Finland.

In an overview of the 3kb early region, several features seem to be unique: (i) In the 5'-terminal noncoding region, there exist inverted terminal repeat sequences, plasmid pBR322 and ColE1 origin-like sequences, a phosphate box, and two stem-loop structures. Though difficult to interpret their exact functions at present, they may be involved in DNA replication, or gene regulation at the transcriptional or translational level. (ii) Only one obvious promoter exists for the terminal protein gene and no other promoters for DNA polymerase and lysin genes are discernable. This suggests that the three genes are transcribed as a polycistronic mRNA and form an operon. (iii) The compact organization of the PRD1 genome is illustrated by the fact that only very short distances exist between a stop codon for one gene and subsequent initiator codon. This distance is only 3 bp between the terminal protein gene and DNA polymerase gene and 1 bp between the DNA polymerase gene and lysin gene. The significance of this gene organization is at present unclear, but may indicate that these three genes are coordinately regulated at both the level of transcription and translation. (iv) There is a potential secondary structure located between nucleotides 2459-2503. This secondary structure has a high free energy value of about -23.8 kcal. Whether this secondary structure plays any role in the regulation of translation and transcription from the promoter of the

terminal protein gene in the expression of which is required early in infection remains to be determined.

Primary Structure of the PRD1 Terminal Protein

The DNA sequence of the PRD1 terminal protein gene has been determined. The coding region is 780 base pairs long and encodes 259 amino acids. The predicted amino acid sequence of the PRD1 terminal protein revealed no overall appreciable homology with other reported terminal proteins or related proteins. On the otherhand, it has been shown recently that the DNA polymerases which use proteins as primers share significant amino acid sequence homologies. Moreover, these DNA polymerases also have partial, but significant amino acid sequence homologies with DNA polymerases which use RNAs as primers (Larder et al., 1987; Jung et al., 1987). These findings strongly suggest that both types of DNA polymerases have evolved from a common ancestor. In view of these findings, it is of great interest to investigate evolutionary origins of the terminal proteins.

On the basis of sharing a common function (i.e., as primer for linear DNA replication), the terminal proteins would be expected to have conserved amino acid sequences or three-dimensional structures required for priming function, and/or association with DNA polymerase. In order to search for conserved

sequences among various terminal proteins, I first focused on the amino and carboxyl terminal areas because of existing evidence for their involvement in priming complex formation (Zaballos et al., 1986, 1988). No conserved sequence was observed within these regions. Next I switched my attention to the internal portions of the molecules. Finally, a highly conserved sequence, YSRLRT, could be found on all identified terminal proteins (Figure 15). Moreover, the spatial location of the YSRLRT sequence on each polypeptide is similar as well. These findings suggest that the conserved YSRLRT domain may play a prominent role in the mechanism of linear DNA replication. From the codon usage analysis of PRD1, ϕ29 and adenovirus terminal proteins, the percursor of adenovirus terminal protein exhibits a considerable bias for codons ending with C (NNC; 49%) compared to codons of the types NNG (28%), NNA (12%) and NNU (11%). This is consistent with the preference for codons ending with C or G nucleotide in eukaryotic DNA (Alestrom et al., 1982). The codon frequencies used in PRD1 terminal protein are NNC (25%), NNG (18%), NNA (32%) and NNU (25%); and in ϕ29 terminal protein are NNC (16%), NNG (24%), NNA (30%) and NNU (30%). This uneven distribution of codon usage in respective terminal proteins may result from different resident host systems.

Cloning and Overexpression of the PRD1 Terminal Protein, DNA Polymerase and Lysin Genes

As described previously that the PRD1 left end genome contains three left early proteins: terminal protein, DNA polymerase, and lysin enzyme. In order to examine whether these proteins are expressed in a T7 RNA polymerase/promoter system, proteins were radioactively labeled using the addition of rifampicin and [35-S]methionine. The results of autoradiography of these labeled proteins are consistent with the expected pattern. There are 3 labeled bands from pT7-6 plasmid-encoded proteins which are shown in the gel in Figure 17). According to expression quantity, the lysin protein was expressed in an unusually high yield. This was suprising because of the lack of a clearly strong promoter for the lysin gene. If the lysin gene shares the same promoter with terminal protein and DNA polymerase genes, the translational efficiency should gradually decrease in the polycistronic mRNA from beginning to end, due to rapid mRNA degradation. To examine the unusual high expression of gene 15, the autoradiography film was exposed for a short time. The new film revealed that the original lysin band is comprised of two protein bands; one is for lysin protein and the other one probably comes from an unknown internal open reading frame of plasmid or 3kb insert.

The Characterization and Predicted Function of the Conserved Sequence "YSRLRT"

Many aspects of the molecular biology of terminal proteins are not understood. Detailed physical and functional studies of terminal protein have been hampered by difficulty in obtaining significant quantities of the purified terminal protein. In addition, the lack of a visible conserved sequence which exist among them. We were interested in characterizing whether or not the short patch of identical amino acid sequence, YSRLRT, play some role in the priming activity of the PRD1 terminal protein. To this end, site-directed mutagenesis was performed to examine the role of each amino acid residue of the YSRRT domain in the PRD1 terminal protein function. Eleven terminal protein mutants were constructed and all of these clones were examined for the ability to form terminal protein-dGMP complex. The results reflect that the YSRRT domain of the PRD1 terminal protein plays an important role in priming complex formation, with arginine-174 residue essential for YSRRT domain function. Further, the positive charge of position 174 is indispensable for PRD1 terminal protein to perform its function. One possibility for this observation is that the arginine-174 residue plays a critical role in maintaining the charge balance of the PRD1 terminal protein. Once arginine-174 is mutated

to a non-charged residue, it could disturb the charge distribution of the overall molecule. This then would result in a change in secondary structure of the PRD1 terminal protein molecule, which may affect the configurations of interactions with dGTP, DNA polymerase, genomic template, or related proteins. Similarly, structural perturbation of the PRD1 terminal protein molecule might induce an unstable state. This unstable PRD1 terminal protein could interfere with either host DNA replication or cellular division via its carboxyl terminal region, which is rich in basic amino acid residues. This might explain why the arginine-174 lacking terminal protein inhibits host cell growth.

At this stage, it is difficult for us to define the exact function of the conserved sequence YSRRT in the PRD1 terminal protein molecule. A hypothetical function of the conserved domain, YSRLRT, could be; 1) as an active priming center (i.e., the center interacts directly with the all substrates for complex formation, 2) a unique association site for DNA polymerase, 3) a required structural signal for the regulation between the free priming terminal protein molecule and parental genomic terminal protein molecule, 4) an essential conformation for priming activity, 5) a recognition site for interaction with dGMP or DNA template, or 6) as an active site for some unknown enzymatic function. Further support for the importance of the conserved domain YSRLRT was shown in the precursor of adenovirus terminal protein. Freimuth

et al., (1986) demonstrated that a mutant was created by the insertion of two codons between YSRLR and YT sequences, resulting in a lethal mutation for the viral replication. This suggests that the contact structure of the conserved sequence YSRLRYT of adenovirus terminal protein precursor is required for adenovirus DNA replication. Purification of mutant terminal protein will be necessary for further determinal of the role of the conserved sequence in protein-primed DNA synthesis.

It is interesting to note that the E. coli primase, dnaG, also shows some homology to the YSRLRT sequence between amino acid residues 194-199 (Smiley et al., 1982). Comparison of the amino acid sequence between dnaG and phage Nf terminal protein in this region reveals the maximum homology (i.e., RSY*R*R*) (Figure 34). Although dnaG protein does not contain the complete YSRLRT consensus sequence, it conserves the sequence Y*R*R*. Specifically, dnaG protein still conserves the most important arginine-174 residue. The spatial location of the Y*R*R* sequence of dnaG protein is similar to that of phage Nf terminal protein. Whether the Y*R*R* sequence of dnaG protein is responsible for priming activity in the E. coli system is at present unknown.

Evolution of Terminal Protein

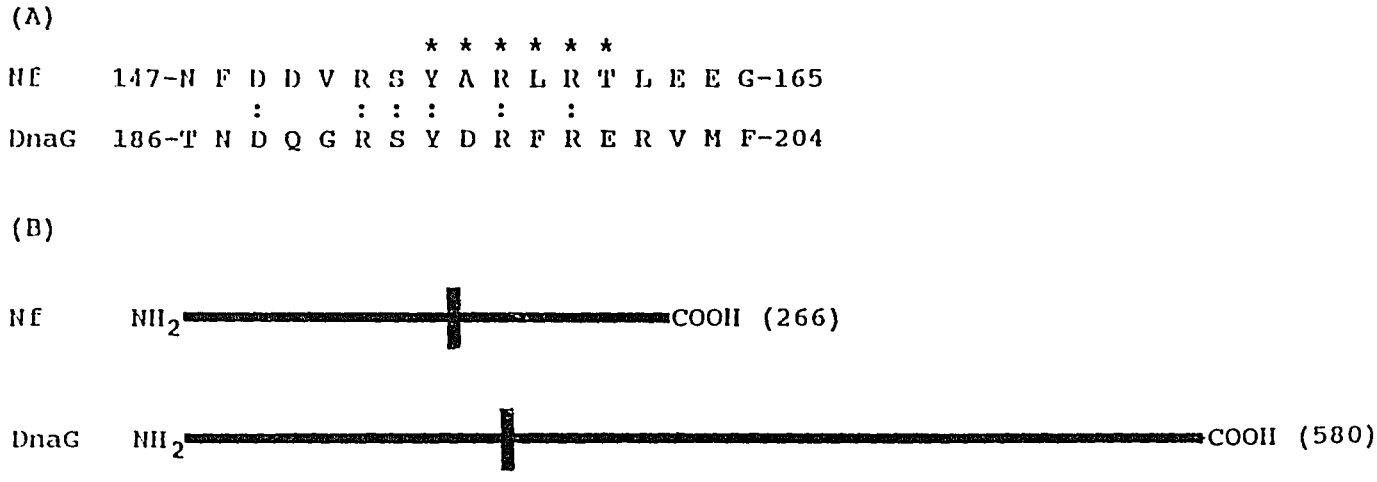


Figure 34. (A) Conserved region of phage Nf terminal protein and E. coli dnaG primase. The star signs indicate the conserved sequence present in all identified terminal proteins. (B) Relative linear spatial location of the conserved region (vertical line) of phage Nf terminal protein and E. coli dnaG primase.

Given the highly conserved amino acid sequence, YSRLRT, and its similar spatial location on all known terminal proteins, it seems likely that these terminal proteins have a common evolutionary origin. Two conclusions can be drawn from the analysis of these conserved features among terminal proteins. First, the conserved sequence YSRLRT must be the most stringently selected region of the whole molecule. It probably avoids complete sequence divergence from an ancestral gene and thereby maintains its original structure-function role. Second, this conserved sequence is an internal homologous region, implying that the terminal protein gene family was generated from the duplication of a small ancestral gene, and gene fusion during evolution.

Can DHBcAg Act As Terminal Protein of Duck Hepatitis B Virus?

Duck hepatitis B virus (DHBV) has a small, circular, partially double-stranded DNA genome. The minus strand is about 3 kilobases with a terminal protein covalently linked to its 5' end. The plus strand is incomplete and varies in length (Sprengel et al., 1985; Mandart et al., 1984). Molnar-Kimber et al., (1983) demonstrated that terminal protein is covalently bound to nascent minus-strand DNA intermediates of DHBV. This indicates that the terminal protein may serve as a primer for DNA synthesis of DHBV minus-strand. Although the DNA has been sequenced, the genes of DHBV have not been confirmed

because of a lack of available data about virus-specific transcripts and proteins. Hence, the exact location of the terminal protein gene is still an open question.

As shown in this thesis, all known terminal proteins and the core antigen of duck hepatitis B virus (DHBcAg) contain short patchs of amino acid sequence (YSRLRT). Not only is the YSRLRT sequence conserved, but the spatial location of the YSRLRT sequence is similar as well. The work of Radziwill et al., (1988), Sprengel et al., (1985) and the present data imply that DHBcAg is the most likely candidate for being the terminal protein because: (i) DHBcAg possesses the conserved sequence, YSRLRT, which is unique to all identified terminal proteins (i.e., PRD1, Nf, PZA, ϕ29 and adenovirus); (ii) The spatial location of the YSRLRT sequence in DHBcAg is quite similar to those of the PRD1, Nf, PZA and ϕ29 terminal proteins (all are located in the central portion of the proteins); (iii) The size of DHBcAg is about 31 kDa, which is close to PRD1, Nf, PZA and ϕ29 terminal proteins; (iv) DHBV DNA polymerase activity is tighly associated with DHBcAg. This observation was also shown for PRD1 and ϕ29 terminal protein and adenovirus pre-terminal protein as they were purified (Enomato et al., 1981; Watabe et al., 1984); (v) Strongly basic amino acid residues clustered in the carboxyl terminal region is similar to that seen in PRD1 terminal protein. These basic amino acid residues could be involved in DNA binding (Pasek et al., 1979). In principle, we can not rule out

the possibility that the final gene product of DHBV terminal protein generated in vivo could be a new shorter polypeptide through the posttranslational processing of DHBcAg as is seen with adenovirus terminal protein.

Evolutionary Relationship Between Terminal Protein and DNA polymerase

It is interesting in comparing the relationship between the terminal protein and DNA polymerase in PRD1, ϕ29 and adenovirus systems. Until now, a highly conserved amino acid sequence YSRLRT (this dissertation), a conserved 3'-5' exonuclease active site (Bernad et al., 1989) and another three conserved regions (Jung et al., 1987), have been found to be present in terminal proteins and DNA polymerase, respectively. Moreover, these 3 systems also display a similar gene order and transcription direction for the terminal protein and DNA polymerase genes (Bamford et al., 1987; Gingeras et al., 1982; Kawamura and Ito, 1977; Davison et al., 1980) (Figure 35). These observations reflect a high evolutionary conservation of the terminal protein and DNA polymerase genes organization. This implies that the terminal protein and DNA polymerase always duplicate together as a unit during evolution. In addition, it is possible that the duplication of the terminal protein and DNA polymerase genes was accompanied by the duplication of a large chrosomal origin in protein-priming systems. The fairly high conservation of amino acid sequences, the structural

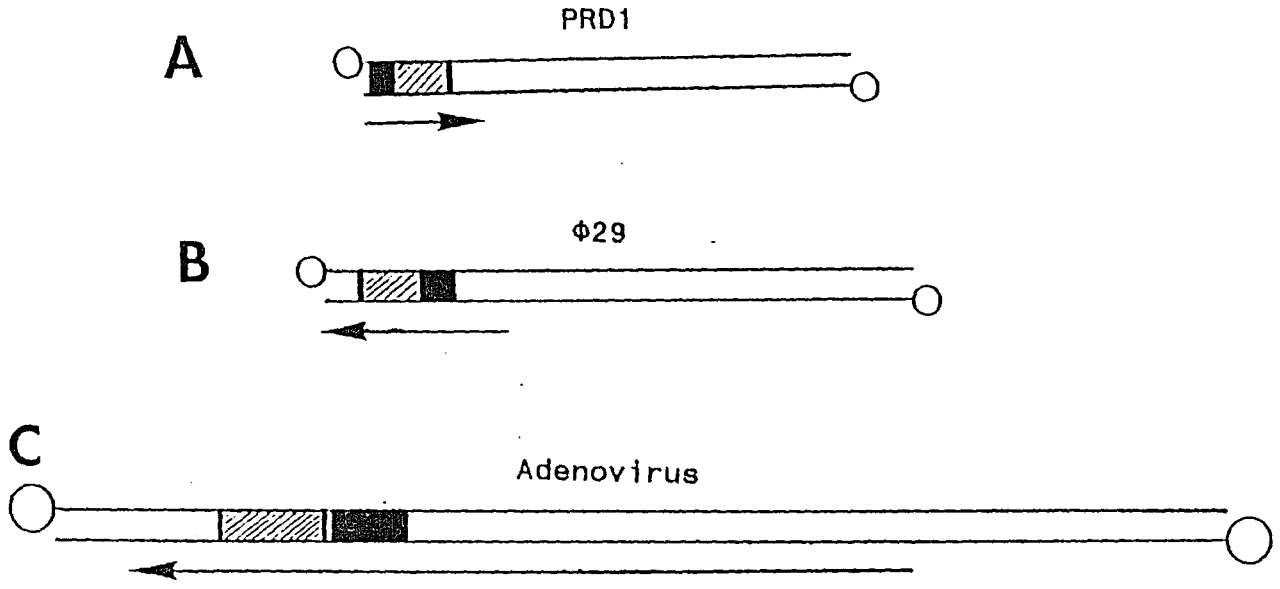


Figure 35. Genetic map of the terminal protein and DNA polymerase genes of PRD1, ϕ29 and adenovirus. Representation of the respective genomes are drawn to scale. The closed box and cross-hatch indicate the positions of terminal protein gene and DNA polymerase gene, respectively. Arrows indicate the direction of transcription. (A) was adapted from McGraw et al. (1983), (B) from Geidnscheck and Ito (1982) and (C) from Kelly (1984).

organization, and the transcriptional direction of terminal protein and DNA polymerase genes provides a valuable tool to identify the terminal protein and DNA polymerase in different species and to analyze phylogenetic relationship among them.

Linking Site, Tyrosine-190, of Phage PRD1 Terminal Protein

The linking site, tyrosine-190, of the PRD1 terminal protein has been determined (this dissertation) and display quite different characteristics from the linking sites of ϕ29 and adenovirus such as hydropathic index and secondary structure (Figure 32 and 33). There is similarity of the hydropathic index and the location of secondary structure between the ϕ29 and adenovirus linking sites, but it is quite different with that of the PRD1 linking site. These enviromental difference of the linking sites could reflect the speciality of the interaction between the linking residue and 5'-terminal nucleotide of their genomes. However, a comparison of the linking site enviroments among PRD1, ϕ29 and adenovirus systems shows a high density of negative charge residues. It seems likely that the high negative charges region can provide a cleft via strong electrostatic repulsion for interaction with the 5'-terminal nucleotide of the genome. Afterwards, the linking residue can bind to the 5'-terminal nucleotide of genome with the help of unknown factors. In addition, the

location of these three linking sites is always preceded by a long α-helix structure. This observation implies that the α-helix structure probably is required for the maintenance of a suitable tertiary conformation for priming complex formation.

LIST OF REFERENCES

Alestrom, P. Akusjarvi, G., Pettersson, M. and Pettersson, U. (1982) J. Biol. Chem. 257, 13492-13498.

Andrews, B. J., Proteau, G. A., Beatty, L. G. and Sadowski, P. D. (1985) Cell 40, 795-803.

Bamford, D. H. and Mindich, L. (1984) J. Virol. 50, 309-315.

Bamford, D. H., Rouhiainen, L., Takkinen, K. and Soderlund, H. (1981) J. Gen. Virol. 57, 365-373.

Bjornsti, M. A., Reilly, B. E. and Anderson, D. C. (1984) J. Virol. 50, 766-772.

Bradley, D. E. (1974) Biochem. and Biophy. Res. Comm. 57, 893-900.

Carter, C. W. and Krant, J. (1974) Proc. Natl. Acad. Sci. USA 71, 283-287.

Cavalier-Smith, T. (1974) Nature 250, 467-470.

Challberg, M. D., Desiderio, S. V. and Kelly, T. J. Jr. (1980) Proc. Natl. Acad. Sci. USA 77, 5105-5109.

Challberg, M. D. and Kelly, T. J. (1979) J. Mol. Biol. 135, 999-1012.

Challberg, M. D. and Kelly, T. J. (1982) Annu. Rev. Biochem. 51, 901-934.

Champoux, J. J. (1981) J. Biol. Chem. 256, 4805-4809.

Chou, P. and Fasman, G. (1978) Adv. Enzymol. 47, 45-148.

Craig, N. L. and Nash, H. A. (1983) Cell 35, 795-803.

Dale, R. M. K., McClure, B. A. and Houchins, J. P. (1985) Plasmid 13, 31-40.

Daniell, E. (1976) J. Virol. 19, 685-708.

Davison, B. L., Murray, C. L. and Rabinowitz, J. C. (1980) J. Biol. Chem. 255, 8319-8330.

Eckard, W. (1982) Cell 28, 199-201.

Enomoto, T., Lichy, J. H., Ikeda, J. E. and Hurwitz, J. (1981) Proc. Natl. Acad. Sci. USA 78, 6779-6783.

Esteban, M., Flores, L. and Holowczak, J. A. (1977) Virology 83, 467-473.

Freimuth, P. I. and Ginsberg, H. S. (1986) Proc. Natl. Acad. Sci. USA 83, 7816-7820.

Gellert, M. (1981) Ann. Rev. Biochem. 50, 879-910.

Gerendasy, D. D. and Ito, J. (1987) J. Virol. 61, 594-596.

Gingeras, T. R., Sciaky, D., Gelinas, R. E., Jiang, B. D., Yen, C. E., Kelly, M. M., Bullock, P. A., Parsons, B. L., O'Neill, K. E. and Roberts R. J. (1982) J. Bioch. Chem. 257, 13475-13491.

Gronostajski, R. M. and Sadowski, P. D. (1985) Mol. Cell Biol. 5, 3247-3279.

Harding, N. E. and Ito, J. (1980) Virology 104, 323-338.

Hirokawa, H. (1972) Proc. Natl. Acad. Sci. USA 69, 1555-1559.

Horwitz, M. S., Kaplan, L. M., Abbound, M., Maritato, J., Chow, L. T. and Broker, T. R. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 769-780.

Hermoso, J. M., Mendez, E., Soriano, F. and Salas, M. (1985) Nucleic Acids Res. 13, 7715-7728.

Hoss, R. H. and Abremski, K. (1985) J. Mol. Biol. 181, 351-362.

Inciarte, M. R., Salas, M. and Sogo, J. M. (1980) J. Virol. 34, 187-199.

Ito, J., Harding, N. E. and Saigo, K. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 525-536.

Ito, J. and Roberts, R. J. (1979) Gene 5, 1-7.

Kawamura, F. and Ito, J. (1977) J. Virol. 23, 562-577.

Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. USA 82, 488-492.

Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.

Larder, B. A., Kemp, S. D. and Darby, G, (1987) EMBO 6, 169-175.

Lechner, R. L. and Kelly, T. J., Jr. (1977) Cell 12, 1007-1020.

Lichy, J. H., Horwitz, M. S. and Hurwitz, J. (1981) Proc. Natl. Acad. Sci. USA 77, 5105-5109.

Lipman, D. J. and Pearson, W. R. (1985) Science 227, 1435-1441.

Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) J. Mol. Biol. 192, 549-556.

Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) 190, 37-44.

Mandart, E., Kay, A. and Galibert, F. (1984) J. Virol. 49, 782-792.

Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.

McGraw, T., Yang, H. and Mindich, L. (1983) Mol. Gen. Genet. 190, 237-244.

Mindich, L., Bamford, D., Goldthwate, C., Laverty, M. and Mackenzie, G. (1982) J. Virol. 44, 1013-1020.

Molnar-Kimber, K. L., Summer, J., Taylor, J. M. and Mason, W. S. (1983) 45, 165-172.

Mount, D. W. and Conard, B. (1984) Nucleic Acids Res. 12, 811-817.

Olsen, R. H., Siak, J. S. and Gray, R. H. (1974) J. Virol. 14, 689-699.

Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., Mackay, P., Leadbetter, G. and Murray, K. (1979) Nature 282, 575-579.

Penalva, M. A. and Salas, M. (1982) Proc. Natl. Acad. Sci. USA 79, 5522-5526.

Pribnow, D. (1975) J. Mol. Biol. 99, 419-443.

Radziwill, G., Zentgraf, H., Schaller, H. and Bosch, V. (1988) Virology 163, 123-132.

Reed, R. R. and Grindley, M. (1981) Cell 25, 721-728.

Reed, R. R. and Moser, C. D. (1984) Cold Spring Harbor Symp. Quant. Biol. 49, 245-249.

Rehosh, D. M. K., Russel, W. C., Bellet, A. J. D. and Robinson, A. J. (1977) Cell 11, 283-295.

Roth, M. J., Brown, D. R. and Hurwitz, J. (1984) J. Biol. Chem. 259, 10556-10567.

Rowe, T. C., Tewey, K. M. and Liu, L. F. (1984) J. Biol. Chem. 259, 9177-9181.

Sadowski, P. (1986) J. Bact. 165, 341-347.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 82, 6783-6787.

Sanhueza, S. and Eisenberg, S. (1984) Proc. Natl. Acad. Sci. USA 81, 4285-4289.

Savilahti, H. and Bamford, D. H. (1986) Gene 49, 199-205.

Shih, M., Watabe, K. and Ito, J. (1982) Biochem. Biophys. REs. Commun. 105, 1031-1036.

Smart, J. E. and Stillman, B. W. (1982) J. Biol. Chem. 257, 13499-13506.

Smiley, B. L., Lupski, J. R., Svec, P. S., McMacken, R. and Godson, G. N. (1982) Proc. Natl. Acad. Sci. USA 79, 4550-4554.

Sollazzo, M., Frank, R. and Cesareni, G. (1985) Gene 37, 199-206.

Sprengel, R., Kuhn, C., Will, H. and Schaller, H. (1985) J. Med. Virol. 15, 323-333.

Stow, N. D. (1982) Nucleic Acids Res. 10, 5105-5119.

Sugino, A., Peebles, C. L., Kreuzer, K. N. and Cozzarelli, N. R. (1977) Proc. Natl. Acad. Sci. USA 74, 4767-4771.

Tabor, S. and Richardson, C. C. (1985) Proc. Natl. Acad. Sci. USA 82, 1074-1078.

Takeda, N., Kuhn, R. J., Yang, C. F., Takegami, T. and Wimmer, E. (1986) J. Virol. 60, 43-51.

Topal, M. D. and Fresco, J. R. (1976) Nature 236, 285-289.

Tse, Y. C., Kirkegard, K. and Wang, J. C. (1980) J. Biol. Chem. 255, 5560-5565.

Vartapetian, A. B., Kooniw, E. V., Agol, V. I. and Bogdanov, A. A. (1984) EMBO 3, 2593-2598.

Vieira, J. and Messing, J. (1987) Methods Enzymol. 153, 3-11.

Watabe, K., Leusch, M. and Ito, J. (1984) Proc. Natl. Acad. Sci. USA 81, 5374-5378.

Watabe, K., Shih, M. F. and Ito, J. (1983) Proc. Natl. Acad. Sci. USA 80, 4248-4252.

Watson, J. D. (1972) Nature 239, 197-201.

Wold, M. S. and McMacken, R. (1982) Proc. Natl. Acad. Sci. USA 79, 4907-4911.

Yoo, S. K. and Ito, J. (1989) Virology 170, 442-449.

Zaballos, A., Mellado, R. P. and Salas, M. (1988) Gene 63, 113-121.

Zaballos, A., Salas, M. and Mellado, R. P. (1986) Gene 43, 103-110.