Nucleic Acids Research, Vol. 19, No. 14 3805–3810

# Mapping of the DNA linking tyrosine residue of the PRD1 terminal protein

Shih-Yi Shiue, Jui-Cheng Hsieh and Junetsu Ito*
Department of Microbiology and Immunology, College of Medicine, The University of Arizona, Tucson, AZ 85724, USA

Received May 28, 1991; Revised and Accepted June 19, 1991

## ABSTRACT

**DNA replication of PRD1, a lipid-containing phage, is initiated by a protein-priming mechanism. The terminal protein encoded by gene 8 acts as a protein primer in DNA synthesis by forming an initiation complex with the 5′-terminal nucleotide, dGMP. The linkage between the terminal protein and the 5′ terminal nucleotide is a tyrosylphosphodiester bond. The PRD1 terminal protein contains 13 tyrosine residues in a total of 259 amino acids. By site-directed mutagenesis of cloned PRD1 gene 8, we replaced 12 of the 13 tyrosine residues in the terminal protein with phenylalanine and the other tyrosine residue with asparagine. Functional analysis of these mutant terminal proteins suggested that tyrosine-190 is the linking amino acid that forms a covalent bond with dGMP. Cyanogen bromide cleavage studies also implicated tyrosine-190 as the DNA-linking amino acid residue of the PRD1 terminal protein. Our results further show that tyrosine residues at both the amino-terminal and the carboxyl-terminal regions are important for the initiation complex forming activity. Predicted secondary structures for the regions around the DNA linking amino acid residues were compared in three terminal proteins (ϕ29, adenovirus-2, and PRD1). While the linking amino acids serine-232 (ϕ29) and serine-577 (adenovirus-2) are found in β-turns in hydrophilic regions, the linking tyrosine-190 of the PRD1 terminal protein is found in a β-sheet in a hydrophobic region.**

## INTRODUCTION

A key protein involved in protein-primed DNA replication is the DNA-terminal protein (terminal protein). The terminal protein serves as a primer for initiation of linear duplex DNA replication (1). In several viral systems, the terminal protein remains bound covalently to genomic DNA and is encapsulated during viral morphogenesis. Upon infection, the terminal protein, together with genomic DNA, is injected into host cells. The terminal protein can also be isolated as a DNA-protein complex from viral particles. The linkage between the terminal protein and genomic DNA is a phosphodiester bond between the hydroxyl group of the linking amino acid residue and the 5′-alphaphosphate of the terminal nucleotide of the genome (2,3). The linking amino acid could have been any one of the following hydroxy amino acids: serine (2,3), threonine (4), or tyrosine (5).

Bacteriophage PRD1 is a prototype of small lipid-containing phages that infect a wide variety of gram-negative bacteria ranging from *E. coli* and *Salmonella typhimurium* to *Pseudomonas aeruginosa* (6). PRD1's genome consists of a linear, double-stranded DNA approximately 14.7 kilobases (Kb) long and covalently bound to a 28,000 dalton terminal protein at its 5′ end (6). The linkage between the PRD1 terminal protein and genomic DNA has been identified as a tyrosylphosphodiester bond (5). The amino acid sequence of the PRD1 terminal protein deduced from the nucleotide sequence of the PRD1 genome indicates 13 tyrosine residues out of 259 amino acids (7,8). To precisely map the linking tyrosine residue, we applied site-directed mutagenesis to the cloned PRD1 terminal protein gene (9). We replaced 12 tyrosine residues with phenylalanine and one tyrosine with asparagine. Since the substitution of a phenylalanine for a tyrosine is structurally conservative, this approach provided a powerful means to determine the DNA linking site of the PRD1 terminal protein. Our functional analyses of the mutant terminal protein suggested that the tyrosine at position 190 of the PRD1 terminal protein is the DNA linking site and the results of cyanogen bromide cleavage studies of the terminal protein-dGMP complex agreed with that conclusion.

## MATERIALS AND METHODS

### Bacterial, phage and plasmids

*Escherichia coli* RZ 1032 [Hfr KL16 pro/45 {lys A (61-61)}, dut1, ung1, thi1, relA1], kindly provided by Dr. C.M.Joyce, was used to prepare the uracil DNA template. *E. coli* NM522 (lac-proAB), thi, hsd5, supE, [$F^-$, proAB $lacI^q\Delta M15$] (10) was used for the transformation of the uracil-containing plasmid DNA. *E. coli* HLB3 (pLM2, pLM3), which carries PRD1 genes 1 and 8, was kindly provided by Dr. L.Mindich (11). The phagemid expression vector, pEMBLex3 was obtained from Dr. G.Cesarevi (12). The helper phage M13K07 (13) was purchased from International Biotechnologies, Inc.

* To whom correspondence should be addressed

## Oligonucleotides

All the synthetic oligonucleotides used for site-directed mutagenesis and DNA sequencing were synthesized on a Cyclone DNA synthesizer (Milligen/Biosearch). Beta-amidite synthesis kits used for the syntheses were purchased from Biosearch company (San Rafael, CA). The oligonucleotides were removed from the CPG column by treatment with 30% ammonium hydroxide, which cleaves the oligonucleotides from the support and dissolves the released oligonucleotides. These were then incubated at 55°C for 5 hours, and dried in a heated Savant Speed Vac Concentrator. The crude synthetic oligonucleotides were suspended in 0.5 ml of 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (TE buffer) and purified by electrophoresis through a 20% acrylamide/7M urea gel, visualized by UV shadowing, excised, and eluted in TE buffer at room temperature overnight. The gel elutes were cleaned by passage through a SEP-PAK column (Waters and Associates) to obtain the purified oligonucleotides. Table 1 gives the mutagenic oligonucleotide sequences with corresponding mutant positions.

## Cloning Strategy

A 3kb DNA fragment containing the PRD1 terminal protein and DNA polymerase genes and the adjacent 90% of the lysin gene, was obtained from a recombinant plasmid pLM3 (11) by cleaving it with restriction endonuclease *Pst*I. The 3kb PRD1 DNA fragment was then inserted into a *Pst*I site of the phagemid pEMBLex3. Phagemid pEMBLex3 carries the N-terminal portion of the lacZ gene. Insertion of the *Pst*I fragment into the cloning sites of this vector resulted in the loss of beta-galactosidase activity. Therefore, a recombinant plasmid can be directly screened on plates containing the substrate X-gal (12). The orientation of the insert was verified by DNA sequence analysis.

## In vitro site-directed mutagenesis

Uracil-containing, single-stranded pEMBL3k DNA was prepared according to the protocol of Kunkel et al. (14). Uracil-containing DNA (1 μg) was mixed with phosphorylated mutagenic oligonucleotides (10 ng), T4 DNA polymerase (1 unit) and T4 DNA ligase (2−5 units). The mixture was incubated at 37°C for 90 min. to synthesize the mutated strand. The resulting double-stranded DNA was used to transform *E. coli* NM522 competent cells. Colonies were selected at random and directly sequenced to screen for mutants.

Table 1. Sequences of synthetic oligonucleotides used for site-directed mutagenesis

| Oligonucleotide | Sequence | Location |
|---|---|---|
| 1 (Y14F) | 5'-AATGGGCTTGTTTTTAAAGAGTTTCAA-3' | 261-287 |
| 2 (Y45F) | 5'-CCAACAAAACACTTTAAAGGAT-3' | 354-375 |
| 3 (Y52F) | 5'-AAGCAAATTTAAAGACG-3' | 380-396 |
| 4 (Y89F) | 5'-CAAGGCGTTTCTGAAAG-3' | 491-507 |
| 5 (Y120F) | 5'-GCGCCGCTTCCGCAAGA-3' | 584-600 |
| 6 (Y131F) | 5'-GCCCGTTTTTATCACGC-3' | 617-633 |
| 7 (Y165F) | 5'-ACAAATCTTTTTCACTTTTAAC-3' | 719-740 |
| 8 (Y172N) | 5'-CTTTTAACGGCAAC**A**ATTCCCGCCGTACC-3' | 733-761 |
| 9 (Y177F) | 5'-TCCCGCCGTACCTTTACGTCATTTGAT-3' | 750-776 |
| 10 (Y190F) | 5'-TAAATTTATGACGTTTGACATTATTATTG-3' | 788-816 |
| 11 (Y226F) | 5'-CGAAGCCAAGGCGTTTAACCGTAACCGTA-3' | 896-924 |
| 12 (Y246F) | 5'-GGCTAAAAAGAAATTCAAGCGCCGTCAAA-3' | 956-984 |
| 13 (Y254F) | 5'-TCAAAAACGCGGCTTTGGCAGCAAGGGGG-3' | 980-1008 |

Location numbers indicate distance from the left end of PRD1 genome. The mutagenic oligonucleotide corresponds to respective mutants shown in parentheses(). The position of point mutation is marked with **boldface.**

## Preparation of Cell Extracts

Cells were grown at 30°C in 500 ml of 2× YT medium supplemented with ampicillin (150μg/ml). The cultures were shifted to 42°C at A590 = 0.45. After 2 hours, cells were harvested by centrifugation at 5°C. Cell extracts were prepared essentially as described by Watabe et al. (15). The cells were resuspended in Tris-HCl buffer (50 mM, pH 7.6) containing 10% (w/v) sucrose and 1 mM dithiothreitol. Cells were lysed by treating with lysozyme (300 μg/ml), EDTA (1 mM) and Brij 58 (0.1%). The resulting lysate was quickly frozen in a liquid nitrogen bath and thawed rapidly at 30°C. The lysate was adjusted to 0.8M NaCl and centrifuged at 220,000 ×g for 90 min at 4°C in a Beckman SW50.1 rotor. To the supernatant, solid ammonium sulfate was added to adjust to 50% saturation. The precipitates were collected by centrifugation. The pellets were resuspended in buffer A (50 mM Tris-HCl, pH 7.6, 1 mM EDTA, 1 mM dithiothreitol). The solution was passed through a DEAE-Cellulose column equilibrated with buffer A. Proteins were eluted by the same buffer containing 0.4 M NaCl. The proteins were precipitated by ammonium sulfate and stored at −80°C until used. The pellets were suspended in buffer A before use. The protein determinations were made by the method of Bradford (16).

## Assay for PRD1 terminal protein-dGMP complex formation

The standard reaction mixture in 50 μl contained 50 mM Hepes-NaOH (pH 7.5), 5 mM $MgCl_2$, 3 mM DTT, 2 mM ATP, 0.13 μM [α-$P^{32}$]dGTP (3000 Ci/mmol), 1 μg PRD1 DNA-protein complex and 100 μg of cell extract. After incubating at 30°C for 30 min., samples were processed as described previously (15). The samples were treated often with RQ1 DNAse (10 units, Promega) for 30 min. at 37°C. The reaction was terminated by adding 50 μl of stopping solution (0.18 M sodium pyrophosphate, 0.05 M EDTA and 20% of Trichloroacetic acid). After centrifugation, the supernatant was discarded and the pellet was dissolved in 30 μl of sample buffer (0.1M Tris-HCl (pH 6.8); 2% SDS; 20% glycerol, 10% β-mercaptoethanol and 0.005% bromophenol blue). The PRD1 terminal protein-dGMP complex was visualized by autoradiography after electrophoresis of the reaction production through a 12% SDS-polyacrylamide gel as described (9).

## Cyanogen bromide cleavage of the PRD1 terminal protein-dGMP complex

The PRD1 terminal protein-dGMP complex was isolated by electroelution from the SDS-polyacrylamide gel and the complex was precipitated by the method of Wessel (17). The cyanogen bromide cleavage of the PRD1 terminal protein-dGMP complex was carried out according to the method of Matsudaria (18).

## Computer-assisted secondary structure analysis

The secondary structure and hydrophilicity of the terminal proteins were predicted using the Genetics Computer Group programs from the University of Wisconsin (19).

## RESULTS AND DISCUSSION

### Site-directed mutagenesis of the PRD1 terminal protein gene identifies tyrosine-190 as the DNA linking amino acid

The nucleotide sequence data predict the existence of 13 tyrosine residues in the PRD1 terminal protein, which consists of 259 amino acid residues (7,8). To determine the position in the PRD1 terminal protein of the tyrosine residue that links to the 5′ terminus of DNA (linking amino acid), we replaced all the tyrosine residues with phenylalanine or asparagine (Figure 1). A series of 17-mer to 29-mer mutagenic oligonucleotides (Table 1) were used to generate mutant PRD1 terminal proteins. After the DNA sequence of each mutant was confirmed, we prepared soluble extracts from each mutant clone after heat induction. Analysis of soluble extracts by SDS-polyacrylamide gel electrophoresis indicated that all mutant clones encoding an amino acid substitution produced a 28,000-dalton polypeptide that co-migrated with wild type PRD1 terminal protein. This polypeptide band was missing from an extract of a clone containing the ochre nonsense mutation in the PRD1 terminal protein gene (9). Extracts were assayed for covalent complex formation between the terminal protein and [$\alpha$-$P^{32}$]dGMP. As shown in Figure 2 (A and B), mutant Y190F abolished completely the complex forming activity with dGMP. Mutant

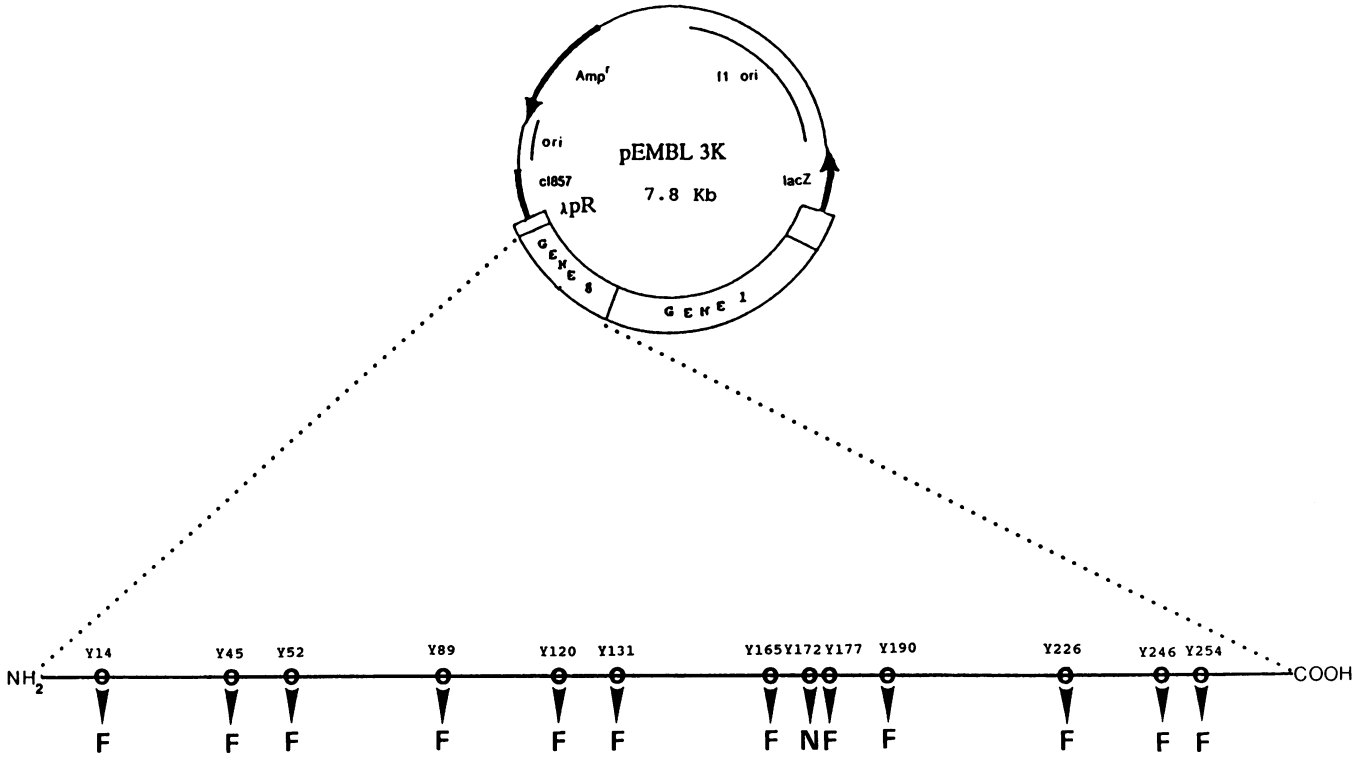


**Fig 1.** (Upper) Diagram of pEMBL3k used for *in vitro* mutagenesis of the PRD1 terminal protein gene; gene 8 encodes the terminal protein; gene 1 encodes PRD1 DNA polymerase. (Lower) The enlarged portion is the PRD1 terminal protein molecule. Numbers indicate the positions of all 13 tyrosine residues in the terminal protein. All the tyrosine residues were changed to phenylalanine (F), except for Y172, which was changed to asparagine (N).

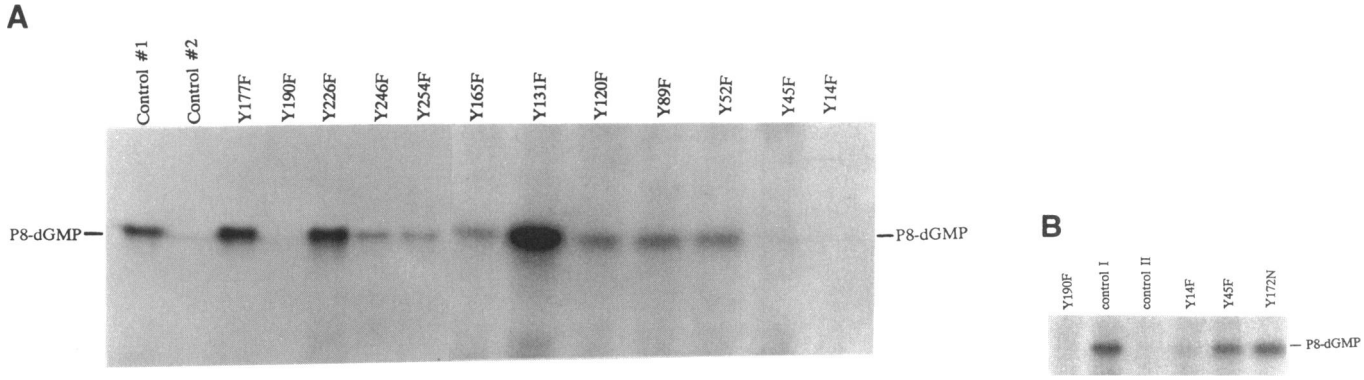


**Fig 2.** Mutational effects of PRD1 terminal protein-dGMP complex formation. Complex formation was carried out under the standard reaction conditions as described in Materials and Methods. (A) Lanes: control #1, wild type PRD1 terminal protein (pEMBL3k); control #2, without 3kb PRD1 terminal protein (pEMBLex3); and 12 Y to F mutant terminal proteins. Mutant Y172N (not shown) was described previously (9). (B) Lanes: control I, wild type PRD1 terminal protein (pEMBL3k); control II, without 3kb PRD1 terminal protein (pEMBLex3); and 4 Y to F mutant terminal proteins.

clones Y14F and Y45F, which are located at the amino terminal region of the PRD1 terminal protein, also exhibited reduced complex forming activities. In particular, mutant Y14F showed very little complex forming activity (Fig. 2B). Y246F and Y254F, which are mapped at the carboxyl region of the PRD1 terminal protein showed somewhat reduced activities. On the other hand, the mutant clones located in the middle portion of the protein displayed normal complex forming activities, except for Y131F, which showed significantly increased activity (Fig. 2A). These data suggest that Tyr-190 is most likely to be the linking site. To confirm Tyr-190 as the DNA-linking tyrosine residue, we applied a CNBr cleavage method to the PRD1 terminal protein-dGMP complex. The PRD1 terminal protein contains four methionine and no cysteine (7). Since one of four methionine residues is at the N-terminal, the CNBr treatment of the PRD1 terminal protein should generate four peptide fragments as shown in the Figure 3A. The molecular weight of the fragment A peptide that contains tyrosine-14 and tyrosine-45 is 16.13 kDa, and fragment B containing tyrosine-190 is 5.3 kDa. The other fragment peptides, C and D are 5.2 kDa and 2.7 kDa, respectively (Figure 3A).

After electrophoresis in SDS-polyacrylamide gel, the radioactive terminal protein-dGMP complex was isolated by electroelution and precipitated by chloroform/methanol. The radioactive complex was then treated with cyanogen bromide according to the method of Matsudaria (18). The CNBr cleaved peptides were analyzed by a Tricine-SDS-Page system (20) and by autoradiography. Only the peptide that contains the DNA-linking tyrosine residue will be radioactively labeled. As shown in Figure 3B, two major labeled bands had approximate molecular weights of 8 kDa and 5 kDa. It is most likely that the 8 kDa peptide is an incomplete cleavage product containing the B and D fragment peptides. The 5 kDa fragment is very close to the calculated molecular weight value (5.3 kDa) of fragment B, containing the tyrosine-190 residue. There are two other, faint bands, with molecular weights of approximately 12 kDa and 27 kDa. We suggest that the 12 kDa fragment is a partial cleavage product containing fragments B, C and D and that the 27 kDa

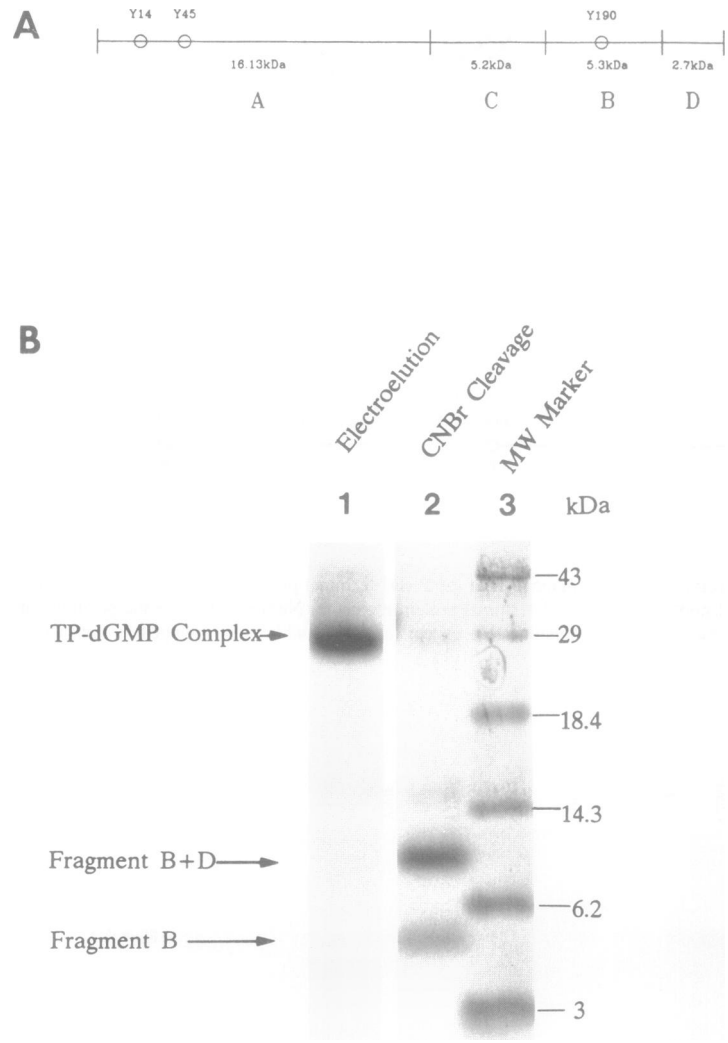


**Fig. 3.** Cyanogen bromide cleavage of the PRD1 terminal protein-dGMP complex. (A). CNBr cleavage map of the PRD1 terminal protein. ○ indicates the location of the tyrosine residues possibly linked to the DNA. Y14, Y45, and Y190 represent mutant clones. (B). Electrophoretic patterns of the CNBr cleaved PRD1 terminal protein fragments. Lane 1, $^{32}$P-labeled PRD1 terminal protein-dGMP complex; Lane 2, the CNBr cleaved fragments of the PRD1 terminal protein-dGMP complex; Lane 3, $^{14}$C-labeled protein molecular weight markers: (Ovalbumin (43 kDa), Carbonic anhydrase (29 kDa), β-lactoglobulin (18.4 kDa), lysozyme (14.3 kDa), Bovine trypsin inhibitor (6.2 kDa), and Insulin (α and β chains) (3.0 kDa)).

represents uncleaved material. It is possible that this incomplete cleavage is due to the presence of some methionine sulfoxide, which is resistant to attack by cyanogen bromide (21). Since fragment A is not labeled, the possibility that tyrosine-14 is the DNA-linking residue of the PRD1 terminal protein is excluded. Therefore, we concluded that the tyrosine residue at position 190 is the DNA-linking amino acid residue of the PRD1 terminal protein. It should be noted that these results are not unique to the PRD1 terminal protein. It has been reported that in both the adenovirus (22,23) and the *Bacillus* phage $\phi$29 (24,25) systems, the amino-terminal and carboxyl-terminal regions of the terminal proteins are important for terminal protein function. Recently, Zaballos and Salas (26) identified the amino-terminal region of the $\phi$29 terminal protein to be associated with DNA, and that the carboxyl-terminal region and an internal region near the amino-terminal region are important for association with the DNA polymerase.

### Comparison of secondary structures around the DNA linking sites among the different terminal proteins

Now that the linking sites of the terminal protein have been determined in three different systems, it is of interest to compare the structure-function relationships among them. These three systems are adenovirus (27), Bacillus phage $\phi$29 (28), and PRD1 (this paper). The primary sequences of these three terminal proteins do not appear to be strongly related (7,8). However, there seems to be some weak homology, in the regions containing the linking sites, between the terminal proteins of PRD1 and adenovirus (Figure 4). Figure 5 illustrates the secondary structure of the PRD1 terminal protein predicted by the Garnier-


**Fig 4.** Alignment of amino acid sequences around the DNA linking sites of the PRD1 and the adenovirus-2 terminal proteins. Arrows indicate the DNA linking amino acids, tyrosine-190 (PRD1) and serine-577 (adenovirus-2).

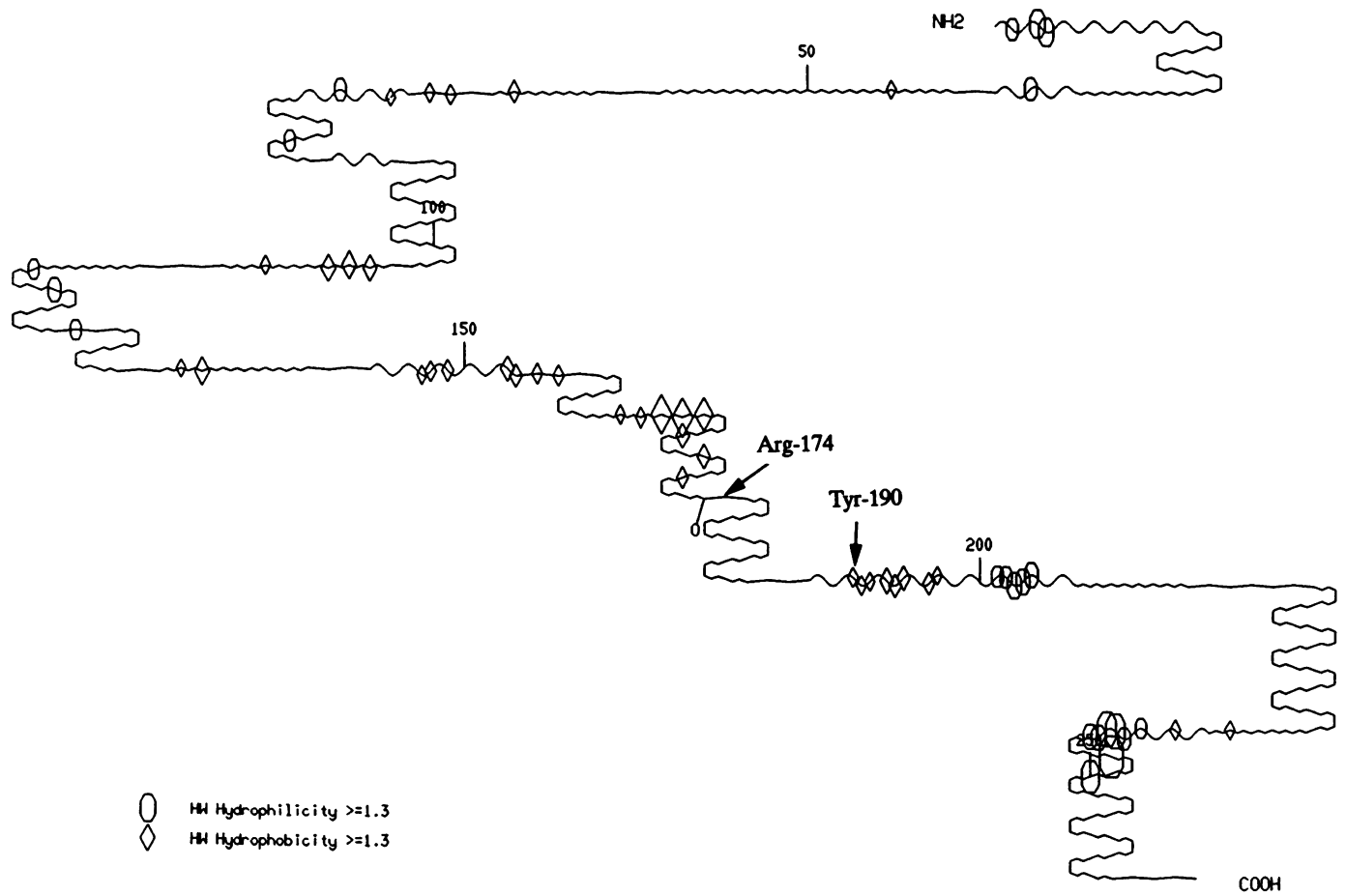


**Fig 5.** Predicted secondary structure and hydrophilicity of the PRD1 terminal protein. The secondary structure of the PRD1 terminal protein was predicted and the figure created by GCG computer programs (19) according to the rules devised by Garnier et al. (29). Open oval circles indicate hydrophilic areas and diamonds indicate hydrophobic areas.

Osguthorpe-Robson method (29). A similar secondary structure was also predicted by the Chou-Fasman method (30). The linking tyrosine-190 is found in a $\beta$-sheet after an $\alpha$-helix in a hydrophobic area. Interestingly, a residue close to tyrosine-190, arginine-174, has been proposed to be the binding site for the phosphate group of dGTP (9). Hermoso et al. (28) suggested previously that the DNA-linking serine residues of the $\phi$29 and adenovirus terminal proteins lie in $\beta$-turns located after $\alpha$-helix regions. Furthermore, the hydropathic profiles around the linking amino acid residues suggested that serine-232 ($\phi$29) and serine-577 (adenovirus) are located in hydrophilic areas. The significance of the topological differences between the linking amino acids of PRD1, $\phi$29, and adenovirus terminal proteins are not understood at present. Presumably, the terminal proteins function as a complex with the respective DNA polymerases. Therefore, it is possible that the secondary and tertiary structures of the free terminal proteins are quite different from those of the functional complexes.

## ACKNOWLEDGEMENTS

We gratefully acknowledge Mr. Dan K.Braithwaite for his computer analysis and for his assistance in preparing this manuscript. This research was supported by grant GM28013 from the National Institutes of Health and by grant NP-704 from the American Cancer Society.

## REFERENCES

1. Rekosh, D.M.K., Russell, W.C., Bellet, A.J.D., and Robinson, A.J. (1977) *Cell* **11**, 283–295.
2. Hermoso, J.M. and Salas, M. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6425–6428.
3. Desiderio, S.V. and Kelly, T.J. Jr. (1981) *J. Mol. Biol.* **145**, 319–337.
4. Garćia, P., Hermoso, J.M., Garćia, J.A., Garćia, E., Lopez, R.E. and Salas, M. (1986) *J. Virol.* **58**, 31–35.
5. Bamford, D.H. and Mindich, L. (1984) *J. Virol.* **50**, 309–315.
6. Mindich, L. and Bamford, D.H. (1988) In Calendar, R. (ed.), The Bacteriophages. Plenum, New York, Vol. 2, pp. 475–520.
7. Hsieh, J.-C., Jung, G., Leavitt, M.C. and Ito, J. (1987) *Nucl. Acids Res.* **15**, 8999–9009.
8. Savilahti, H. and Bamford, D.H. (1987) *Gene* **57**, 121–130.
9. Hsieh, J.-C., Yoo, S. and Ito, J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8665–8669.
10. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
11. Mindich, L. and McGraw, T. (1983) *Mol. Gen. Genet.* **190**, 233–236.
12. Sollazzo, M., Frank, R. and Cesareni, G. (1985) *Gene* **37**, 199–206.
13. Vieira, J. and Messing, J. (1987) In Wu and Grossman (eds.), Methods in Enzymol.—Recombinant DNA. Academic Press. New York, Vol. 153, pp. 3–11.
14. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) In Wu and Grossman (eds.), Methods in Enzymol.—Recombinant DNA. Academic Press, New York, Vol. 154, pp. 367–382.
15. Watabe, K., Shih, M. and Ito, J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4248–4252.
16. Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248–254.
17. Wessel, D. and Flügge, W.I. (1984) *Anal. Biochem.* **138**, 141–143.
18. Matsudaria, P. (1990) In Deutscher (ed.), Methods in Enzymol.—Guide to Protein Purification. Academic Press, New York, Vol. 182, pp. 602–613.
19. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387–395.
20. Schägger, H. and Von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379.
21. Gross, E. (1967) In Hirs (ed.), Methods in Enzymol.—Enzyme Structure. Academic Press, New York, Vol. 11, pp. 238–255.
22. Pettit, S.C., Horwitz, M.S. and Engler, J.A. (1989) J. Virol. **63**, 5244–5250.
23. Freimuth, P.I. and Ginsberg, H.S. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7816–7820.
24. Zaballos, A., Salas, M. and Mellado, R.P. (1986) *Gene* **43**, 103–110.
25. Zaballos, A., Mellado, R.P., and Salas, M. (1988) *Gene* **63**, 113–121.
26. Zaballos, A. and Salas, M. (1989) *Nucl. Acids Res.* **24**, 10353–10366.
27. Smart, J.E. and Stillman, B.W. (1982) *J. Biol. Chem.* **257**, 13499–13506.
28. Hermoso, J.M., Méndez, E., Soriano, F. and Salas, M. (1985) *Nucl. Acids Res.* **13**, 7715–7728.
29. Garnier, J., Osguthorpe, D.J., and Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
30. Chou, P.Y. and Fasman, G.D. (1978) In Meister, A.E. (ed.), *Advances in Enzymol.* **47**, pp. 45–148.