The Endonuclease Domain of Bacteriophage Terminases Belongs to the Resolvase/Integrase/Ribonuclease H Superfamily

A BIOINFORMATICS ANALYSIS VALIDATED BY A FUNCTIONAL STUDY ON BACTERIOPHAGE T5*

Received for publication, November 2, 2005, and in revised form, December 22, 2005 Published, JBC Papers in Press, December 22, 2005, DOI 10.1074/jbc.M511817200

Luc Ponchon†1, Pascale Boulanger1, Gilles Labesse§, and Lucienne Letellier‡2

From the †Institut de Biochimie et Biophysique Moléculaire et Cellulaire, Unité Mixte de Recherche CNRS 8619, Bâtiment 430, Université de Paris-Sud, 91405 Orsay, France and §Centre de Biochimie Structurale, INSERM U554, CNRS UMR5048, Université Montpellier I, 29 Rue de Navacelles, 34060 Montpellier, France

Bacteriophage terminases are essential molecular motors involved in the encapsidation of viral DNA. They are hetero-multimers whose large subunit encodes both ATPase and endonuclease activities. Although the ATPase domain is well characterized from sequence and functional analysis, the C-terminal region remains poorly defined. We describe sequence-structure comparisons of the endonuclease region of various bacteriophages that revealed new sequence similarities shared by this region and the Holliday junction resolvase RuvC and to a lesser extent the HIV integrase and the ribonuclease H. Extensive sequence comparison and motif refinement led to a common signature of terminases and resolvases with three conserved acidic residues engaged in catalytic activity. Sequence analyses were validated by in vivo and in vitro functional assays showing that the nuclease activity of the endonuclease domain of bacteriophage T5 terminase was abolished by mutation of any of the three predicted catalytic aspartates. Overall, these data suggest that the endonuclease domains of terminases operate autonomously and that they adopt a fold similar to that of resolvases and share the same divergent cation-dependent enzymatic mechanism.

Packaging of the double-stranded DNA of bacteriophages into the viral capsid relies on a molecular machine consisting of several proteins: the portal pore, which is located at one vertex of the capsid, and the terminases. The latter are implicated, first, in bringing the DNA concatemers to the empty prohead through interaction with the portal protein, then in active ATP-driven translocation of the DNA into the capsid, and, finally, in cutting the concatemer to generate the mature viral DNA (1). Terminases are generally described as hetero-multimeric structures consisting of two subunits. The small regulatory subunit binds to the viral DNA, and the large catalytic subunit carries both an endonuclease activity for DNA cleavage and an ATPase activity required for DNA packaging (2–5). Such general organization is well conserved among tailed phages and herpes viruses (6). The functional characterization of these proteins and the interplay of the different domains is the object of intensive investigation, specially in the case of phage λ (1, 7–9), T4 (10, 11), and SPP1 (12, 13). The terminase large subunit is a two-domain protein. Its N terminus is well characterized both from sequence and functional analysis. It consists of a conserved ATPase catalytic center with clear ATP binding motifs (named Walker A and B) (14). The C-terminal region is less well defined (10, 13, 15, 16). Mutagenesis studies have shown that the endonuclease activity of phage λ is located within the C-terminal half of the terminase large subunit (gpA; 73 kDa) (17). Several mutations affecting this region, among which one was located on Asp-401, were shown to inactivate the endonuclease activity (16). Phage T4 large subunit terminase (gp17, 70 kDa) exhibits an in vivo endonuclease activity (18). Extensive site-directed mutagenesis has revealed a cluster of conserved residues (Asp-401, Glu-404, Gly-405, and Asp-409) within the C-terminal half of gp17 that are critical for function (15). Further biochemical analyses revealed only aspartate Asp-401 as an essential residue for T4 phage nuclease activity (15). Expression of the C-terminal half of gp17 was sufficient for nuclease activity (10).

The C-terminal regions of terminases characterized so far share neither a sequence motif nor sequence similarities with any other known protein family. This sequence divergence makes difficult a clear delineation of the endonuclease domain and a definition of its active site. No relationship with endonucleases already characterized at the structural level was described yet. Indeed, the numerous endonucleases studied so far adopt a great variety of folds ranging from all α to all β, including various α/β and α + β folds, but show no common sequence motifs (19). The recent sequencing of the bacteriophage T5 genome (GenBank® accession numbers AY692264, AY734508, and AY543070) prompted us to search for new sequence structure-function relationships in terminases.

The T5 genome is a linear double-stranded DNA of 121,750 bp that contains large direct terminal repeats of ~10,000 bp (20, 21). Its replication proceeds via concatemeric intermediates (22), but the processing pathway of T5 concatemers as well as the mechanism of DNA encapsidation remain unknown. The presence of unusually extensive terminal redundancies and of site-specific single-stranded interruptions (23) in the mature as well as in the concatemeric DNA (22) makes T5 an interesting model of DNA packaging.

Here we report an extended bioinformatics analysis of the endonuclease region of double-stranded DNA phage terminases that reveals new sequence similarities shared by this region and the Holliday junction resolvase RuvC and to a lesser extent by the HIV integrase and the ribonuclease H. This analysis is further validated by in vivo and in vitro...
Bioinformatics and Functional Study of Terminases

EXPERIMENTAL PROCEDURES

Bioinformatic Procedures—The PSI-BLAST program (24) was applied with standard parameters to search for homologous proteins in the SPTrEMBL and NCBI non-redundant sequence databases. Protein sequences were retrieved from the NCBI Entrez data base (www.ncbi.nlm.nih.gov/entrez). Their accession numbers are given in the legend of Fig. 2A. Fold compatibility for the full-length and truncated sequences of the terminases SPP1, T4, and T5 was searched using 3DPSSM (25), FUGUE (26), mGenThreader (27), PDB-Blast (bioinformatics.burnham-inst.org/pdb_blast/), SAM-T99 (28) via the metaserver @TOME (29). Automatic modeling were performed using TITO (30), SCWRL (31), and MODELLER (32), and the validity of the resulting models was evaluated with PROSA (33) and Verify3D (34). Sequence-structure alignments including various terminases and PDB1HRJ (see alignment in Fig. 2A) were manually refined with the help of the program ViTO (35). Improved three-dimensional models were built for the T5 terminase using MODELLER 6.2 with the loop optimization procedure. Secondary structure assignments were performed using DSSP (36).

Cloning Procedure—The gene encoding the terminase large subunit was identified from the complete genome sequence (GenBankTM accession number AY692264). The Endo-Term3 domain, delimited between amino acids 191 and 438, was amplified by PCR using T5 DNA as a template. The oligonucleotides used for PCR were complementary to the coding sequence of the Endo-Term gene fragment over 6 codons at both 5′- and 3′-ends and contained additional bases to create a XhoI and a BamH1 restriction site at the 5′- and 3′-ends of the gene fragment, respectively. The amplified Endo-Term PCR fragment was cloned into the vector pET15b (Novagen), between the XhoI and BamH1 restriction sites of the multicloning site, to generate an N-terminal hexahistidine-tagged Endo-Term domain. Three mutated Endo-Term gene fragments were constructed with the following changes in the sequence: D286N, D342N, and D425N. Site-directed mutagenesis was performed by using the kit “Expand High Fidelity PCR System” from Roche Applied Science care) pre-equilibrated with 20 mM Tris-HCl, pH 8, 300 mM NaCl, 5 mM MgCl2, and 20 μg of T5 DNA. The purified proteins (200 μg) were then added. The fluorescence intensity was monitored at 30 °C as a function of time on a Varian fluorimeter with excitation and emission wavelengths set at 491 and 509 nm, respectively.

RESULTS

Bioinformatic Analysis of Terminator Sequences—The T5 genome encodes a putative 49-kDa large subunit terminase, and the latter was used as a representative member of the terminase family together with those of bacteriophages SPP1 and T4 (10, 13). A sequence similarity search using the program PSI-BLAST (24) and the large subunit of the T5 terminase as a query only hit several terminase sequences at convergence. This search also confirmed a higher level of homology at the N terminus (residues 1–165) and a higher divergence at the C-terminal part (residues 170–438). We further searched for structural similarities using the full-length terminases of bacteriophages SPP1, T4, and T5 as queries. Among the 5-fold recognition servers used, four highlighted strong similarities (>95% certainty) to ATP-dependent helicases and related ATPases. Highly significant scores were computed by mGenThreader and FUGUE for phages T4 and T5 terminases (see supplementary materials at www.infobiosud.cnrs.fr/bioserver/TER/suppl.html). However, the detected similarities covered only roughly one half of the terminase sequences. No significant hit corresponded to the C-terminal half. However, analysis of the full-size protein sequence could hide other structural and functional relationships that the C-terminal part might share with other proteins. Delimitation of the C- and N-terminal regions was therefore undertaken starting from the above similarity searches and from additional fold-compatibility searches. To further confirm and refine the proposed structural alignments, sequence analyses were performed with the N terminus and the C-terminal sequences, separately.

According to sequence comparisons, the N-terminal region corresponds to an ATPase domain. It possesses characteristic Walker A (Motif, gXXXGKs where “X” stands for any amino acid and small and capital letters represent mostly present and strictly conserved amino acids). The His-tagged proteins were eluted with 1M imidazole and the supernatant was loaded onto a Ni-NTA Superflow affinity resin was identified from the complete genome sequence (GenBankTM accession number AY692264). The Endo-Term gene constructs were verified by sequence analysis separated by electrophoresis on SDS-PAGE (12%) and stained with Coomassie blue. N-terminal amino acid sequencing was performed after transfer of the protein bands to a polyvinylidene difluoride membrane. The concentration of the proteins was determined using a value of the molar extinction coefficient of 39070 M−1 cm−1.

In Vivo Nuclease Assay—Cells containing the various pET15b constructs were grown at 37 °C in Luria Bertani medium supplemented with ampicillin (100 μg/ml) and chloramphenicol (20 μg/ml). Overproduction of the wild type and the mutated endonuclease—D-galactopyranoside (1 mM). Purification of the proteins was conducted at room temperature as described above. 1-ml aliquots of the cultures were withdrawn prior to induction and 60, 120, and 180 min after induction. Each aliquot was centrifuged at 3000 × g for 5 min, and the pellet was used to isolate the DNA by the alkaline lysis procedure. The DNA was further purified using the Quick purification kit Nucleospin; Macherey-Nagel. 10 μl of DNA were run on a 1% (w/v) agarose gel.

In Vitro Nuclease Assays—All assays were performed after elution of the purified proteins from the gel filtration column to avoid their proteolysis. Purified proteins (36 μl; 0.2 mg/ml) were incubated at 37 °C with purified phage T5 DNA (4 μl; 0.32 mg/ml) in the presence of MgCl2 (5 mM). 10-μl aliquots of the reaction mixture were withdrawn 30, 60, 120, and 180 min after the beginning of the reaction. Nuclease activity was stopped by addition of EDTA (10 mM). DNA cleavage was followed on a 1% (w/v) agarose gel.

The nuclease activity of the purified proteins was also assessed from fluorescence measurements using the DNA intercalant YO-PRO-1 (38) (1 mM in dimethyl sulfoxide; Molecular Probes). The dye was 1000-fold diluted in a 1-ml cuvette containing 20 mM Tris-HCl, pH 8, 300 mM NaCl, 5 mM MgCl2, and 20 μg of T5 DNA. The purified proteins (200 μg) were then added. The fluorescence intensity was monitored at 30 °C as a function of time on a Varian fluorimeter with excitation and emission wavelengths set at 491 and 509 nm, respectively.
acids, respectively) and Walker B (Motif, ///DE where “/” stand for a hydrophobic residue) Mg-ATP binding motifs and an ATPase-coupling motif (motif III). These motifs are located at positions 62–68, 150–155, and 184–186 in the T5 terminase (Fig. 1). This signature is present in most terminases (11, 39, 40) and is related to the ATP binding signature of helicases (41) (notably the particular Walker B motif ///DEad).
FIGURE 2. A, multiple sequence alignment of terminases and RuvC resolvases. The alignment was refined with the help of the program ViTO from pairwise alignments obtained from the metaserver @TOME (see "Experimental Procedures"). Structure code is from the Protein Data Bank (www.rcsb.org/pdb/); resolvases and terminases are named according to the organism or bacteriophage to which they belong. Secondary structure assignments were performed using DSSP (36). Positions occupied mostly by hydrophobic or small residues are highlighted in yellow and green, respectively. The conserved and putative catalytic residues are shown in red and purple, respectively. The sequence pattern specifically derived using Bioinformatics and Functional Study of Terminases

5832 JOURNAL OF BIOLOGICAL CHEMISTRY
VOLUME 281 • NUMBER 9 • MARCH 3, 2006
Bioinformatics and Functional Study of Terminases

To further confirm this new relationship, the sequence motifs containing two putatively catalytic aspartates were refined using the program PATTINPROT (43). Extensive sequence comparison and motif refinement led us to a specific sequence signature (see alignment in Fig. 2A) of the resolvase-terminase subfamily with a highly significant e-value of 4.4e-11. This signature contains three well conserved acidic residues (aspartates 286, 342, and 425 according to the T5 terminase numbering). They are equivalent to the residues of the Holliday junction resolvase (PDB 1HJR) that form a catalytic triad requiring divalent cations (Mg\(^{2+}\) or Mn\(^{2+}\)) for activity (44, 45). This local sequence and overall fold conservation suggested that the terminases possess a similar endonuclease domain sharing a common enzymatic mechanism with resolvases. To validate this proposal, we expressed the wild type T5 endonuclease domain and three recombinant proteins in which each of the predicted catalytic aspartates was substituted by an asparagine and measured their endonuclease activity.

**Expression of Endo-term, the Endonuclease Domain of Phage T5 Terminase: Mutations of the Three Predicted Catalytic Aspartates Abolish in Vivo Endonuclease Activity—**The constructs used (respectively, named WT, D286N, D342N, D425N Endo-term) were delimited between amino acids 191 and 438 and contained an additional N-terminal hexahistidine sequence. The recombinant proteins were expressed in *E. coli* BL21 (DE3) pLysS to lower their basal expression and subsequent toxicity. Isopropyl-1-thio-β-D-galactopyranoside induction resulted in an immediate arrest of growth of the bacteria carrying the WT Endo-term plasmid but only slightly affected the growth of those carrying the D286N, D342N, D425N mutations. Aliquots of the cells were taken at different times after induction and tested for protein expression. Although the level of expression of WT Endo-term appeared significantly lower than that of the mutated proteins (data not shown), its nuclease activity was higher. Indeed, WT Endo-term expression resulted in cleavage of the corresponding pET15b-derived plasmid and of the genomic DNA (see below), generating a complex mixture of DNA fragments that migrated as a characteristic smear on agarose gels (Fig. 3, lanes 2 and 3). The DNA was totally cut at t = 180 min (lane 4). On the contrary, the plasmids harboring each of the conservative mutations D286N, D342N, or D425N showed no sign of degradation over 180 min (lanes 6–8, 10–12, 14–16). These results suggested that mutation of the conserved aspartate residues strongly affected the *in vivo* endonuclease activity of the proteins. To validate these results and to determine how the mutations affected the nuclease activity, we purified the four proteins and determined their *in vitro* endonuclease activity.

**Purification and in Vitro Endonuclease Activity of the WT, D286N, D342N, and D425N Endo-term Proteins—**All four proteins, expressed as described above, were found in the supernatant of cells centrifuged after a French pressure treatment. They were purified to near homogeneity by successive passage on an affinity column and on a gel filtration column as described. WT Endo-term migrated at an apparent molecular mass of 31 kDa, corresponding to that expected for the cloned DNA fragment (Fig. 4, lanes 3 and 5). The protein underwent limited proteolysis with time. A single product with molecular mass of 25 kDa was obtained after 72 h (Fig. 4, lane 7). Sequencing indicated that the protein

---

**PATTINPROT for both terminases and resolvases RuvC is underlined in the alignment using asterisks and is as follows (with catalytic residues in bold): [SAGCTVILMFYRW]-[ACTVILM]

---

**MARCH 3, 2006•VOLUME 281•NUMBER 9**

**JOURNAL OF BIOLOGICAL CHEMISTRY**

**5833**

---

**minimal ATase region should encompass these motifs, and sequence-structure comparisons using the full-length terminases suggested that the N-terminal domain is roughly 160 residues in length.** Fold-compatibility searches using the derived N-terminal region revealed the same trends: highly significant similarities (e-values below e-4 corresponding to >95% certainties) were observed with various helicases despite a rather low level of sequence identity (~15%). Molecular modeling using MODELLER (32) was performed to analyze the proposed structural alignment and to precisely delimit the helicase-like ATP-binding domain of the terminases (data not shown). It suggested that the ~160 residues of the N-terminal domain could form a stable and functional ATP-binding domain autonomously. Fold-recognition searches were then performed using the C-terminal region from distinct terminases.

No significant score for any known structure was observed in the various runs performed using PSI-BLAST, in agreement with previous studies. On the contrary, fold-recognition using the metaserver @TOME suggested distant relationships with various structures, among which were some endonucleases (see supplementary materials at www.infobiosud.cnrs.fr/bioserver/TER/suppl.html). However, the compatibility scores were rather weak, and most structural alignments seemed to match only to the very C terminus (residues 270–438). The searches were resumed with shorter sequences according to the primary matches. In addition, the 14 top-scoring sequence-structure alignments were systematically tested using the automatic procedure available on the @TOME server. Promising three-dimensional models were observed for the C-terminal domain of the T5 and T4 phage terminases with scores ranging between ~0.45 and ~0.65 in PROSA and 0.25 and 0.35 in Verify3D. However, the overall sequence identity was low (~10% over 150 residues). In most cases the models were built using three structurally related templates: a resolvase RuvC (PDB1HJR), a ribonuclease H (PDB1I02), and two retroviral integrases (PDB1A5V and PDB1B9D). These four proteins possess a similar catalytic domain sharing a common topology (α/β) and the same enzymatic mechanism (42). Among these enzymes, resolvase RuvC appeared to be the most closely related to the terminases (data not shown). Resolvases were used for further alignment refinement and validation of the structural comparison.

The first sequence-structure alignments including various terminases and the resolvase RuvC (PDB: 1HJR, see alignment in Fig. 2A) were manually refined and used to build new three-dimensional models for the T5 terminase. The resulting model (Fig. 2B) appeared significantly better and more satisfactory, taking into account the low level of sequence identity (PROSA score below ~1.0 versus ~1.2 for the template PDB 1HJR; Verify3D score above 0.32 and 0.35, respectively). These results supported the proposed structural alignment of the catalytic domain of resolvases with the C-terminal endonuclease domain of the terminases. Furthermore, two sequence motifs (g/D/g and dx/D) corresponding to the active site of RuvC resolvases (residues 5–10 and 139–148, respectively) are conserved in the terminases (in T5, residues 283–288 and 422–431, respectively). This suggested a potential functional homology and also enabled us to delimit more precisely the C-terminal domain (residues 281–438).
was missing its N-terminal hexahistidine tag. Because proteolysis resulted in a shift of the molecular mass from 31 to 25 kDa this indicated that the cleavage had essentially occurred in the C-terminal region. The release of these 6 kDa results in a loss of the critical aspartate 425 and is also likely to affect the protein fold, explaining why the proteolysed protein showed no nuclease activity when assayed by agarose gel electrophoresis (data not shown). The purified D286N, D342N, and D425N Endo-term proteins showed a pattern of time-dependent proteolysis similar to that of WT Endo-term (data not shown). Further functional assays were therefore only performed on freshly prepared proteins that were retained on the Ni-NTA column.

Each of the four proteins was mixed with phage T5 DNA in a one:one protein:nucleotide molar ratio. Aliquots were taken at given times, and DNA cleavage was followed on agarose gel. Fig. 5A, lanes 5–8, shows that WT Endo-term efficiently cut T5 DNA in less than 60 min. DNA cleavage by the three mutants was slower: D286N and D425N Endo-term showed the same activity and cut the DNA in less than 180 min (lanes 9–12 and 17–20). D342N Endo-term showed the lowest nuclease activity, and the DNA was only slightly cut after 180 min (lanes 13–16). In all cases was the cut nonspecific, confirming that the smear observed in the in vivo assay corresponded to small pieces of E. coli genomic DNA. Endonuclease activity had an absolute requirement for divalent cations (MgCl₂ or MnCl₂) because it was not observed in the absence of these cations or arrested by addition of an excess of EDTA (data not shown). The decreased activity of the mutated proteins compared with the wild type protein was not due to a misfolding of the proteins because all four proteins retained the same content of secondary structure as judged by their circular dichroism spectrum (data not shown).

Endonuclease activity was further quantified by measuring the dependence with time of the DNA intercalant YO-PRO1 fluorescence, which is enhanced upon binding to DNA and decreased when the DNA is cut by nucleases (38). Fig. 5B highlights the difference in nuclease activity of the four proteins. The t₁/₂ of endonuclease activity was 14 and 50 min for the WT and D342N Endo-term proteins, respectively, and 30 min for both the D286N and D425N Endo-term proteins. This unambiguously demonstrated that the aspartate mutations affected nuclease activity and that all three substitutions were not equivalent.

**DISCUSSION**

In this study we have used a combination of bioinformatics and biochemical analysis to precisely define the catalytic residues involved in the endonuclease activity of the large subunit of tailed phage terminases.

**FIGURE 3.** Aspartate 286, 342, and 425 mutants lost in vivo endonuclease activity. The different constructs were transferred into E. coli BL21(DE3) pLys-S. Cultures were grown to an A₆₀₀ of 0.4 and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside. Aliquots were taken before induction (−) and 60, 120, and 180 min following induction (+). A portion from each aliquot was used to make miniprep plasmid DNA. These DNA samples were electrophoresed on a 1% (w/v) agarose gel. Smeared appearance of the DNA revealed DNA activity, and the DNA was only slightly cut after 180 min (lanes 1–4), wild type Endo-term (lanes 5–8, D286N; lanes 9–12, D342N; lanes 13–16, D425N; lanes 17–21, control experiment with the terminase helicase N-terminal domain (residues 1–190) showing no DNA degradation. Right lane (M), size markers.

**FIGURE 4.** Expression, purification, and proteolytic cleavage of WT Endo-term. SDS-PAGE analysis of His-tagged WT Endo-term (residues 191–438) expressed in E. coli BL21(DE3), lanes 1, 4, and 6, molecular mass markers. Lanes 2 and 3, supernatant of lysed bacteria before (−) and after (+) isopropyl-1-thio-β-D-galactopyranoside induction. Lane 5, purified Endo-term obtained after successive passage on Ni-NTA and superose 12 columns. The protein migrates at the expected molecular mass of 31 kDa. Lane 7, purified Endo-term 72 h after purification showing the proteolyzed form migrating at a MM of 25 kDa.

Extensive sequence comparisons led to the conclusion that, despite primary sequence divergence, terminases including those of phage T7, P1, A, P22, SPP1, T1, P2, T4, and T5 contain three well conserved aspartate residues that are equivalent to the residues of the Holliday junction resolvases and, to a lesser extent, the HIV integrase and ribonuclease H. The role of these residues in the endonuclease activity of the terminase large subunits was attested by functional assays performed on phage T5. From sequence comparisons and from the known location of the functional domains of phage T4 (10) and λ terminases (16, 46), we could delimit the endonuclease domain (Endo-Term) of the T5 terminase. Expression of the plasmid-encoded corresponding protein resulted in in vivo cutting of both plasmid and genomic DNA. Such in vivo nuclease activity was considerably reduced by mutating each of the critical conserved aspartate residues (respectively D286N, D342N, and D425N). Further purification of the corresponding expressed proteins confirmed the role played by the aspartate residues and allowed determination of the relative nuclease activity of the proteins: WT Endo-term showed the highest nuclease activity, and the activity of D286N and D425 was about half that of WT Endo-term. Finally, D342 showed the lowest nuclease activity. The defect in DNA cleavage activity was not due to misfolding,
because all four proteins retained the same content of secondary structure. The endonuclease activity of the proteins had an absolute requirement for divalent cations (mainly Mg\(^{2+}\) or Mn\(^{2+}\)) in agreement with that found for the nuclease domain of phage \(\lambda\) terminase (1) and for Holliday junction resolvases, among which is the \(E.\ coli\) protein RuvC (47). A divalent metal cation is also a cofactor required for the catalytic activity of the proteins belonging to the integrase/ribonuclease-H superfamily (45). All these proteins have similar active sites containing three to four catalytic carboxylate-bearing residues (44). It is therefore likely that phage endonucleases and these proteins share a common catalytic mechanism in which the acidic residues are involved in the coordination of the metal required for the cleavage of the phosphodiester bond. Does such a mechanism explain the differences observed in the nuclease activity of the three mutated proteins? According to the model shown in Fig. 2B, Asp-286 and D425N are located in the vicinity of the aspartate 291 residue so that chelation of the metal might occur via this third aspartate if any of the two others is mutated. As a result, some of the nuclease activity would be restored. On the other hand, the nuclease activity would not be restored by mutation of Asp-342 because this residue lies too far away from any other aspartate to permit chelation.

Exhaustive mutagenesis experiments have allowed critical amino acids involved in the endonuclease activity of phage T4 and \(\lambda\) terminases to be defined (15–17). In particular, it was shown that residue Asp-401 (which corresponds to Asp-286 in T5) from phage T4 and \(\lambda\) terminases was absolutely required for \textit{in vivo} and \textit{in vitro} DNA cutting activity, but not for \textit{in vitro} packaging of the phage genome. On the basis of the T5 data and given the high conservation of the three-aspartate motif (among which is Asp-401 from \(\lambda\) and T4) (see Fig. 2A), we propose that the aspartate residues are required, and sufficient, to promote the endonucleotidic activity of phage T4 and \(\lambda\) terminases. Therefore phage endonucleases share the same catalytic activity independently of whether DNA cleavage is sequence specific as in \(\lambda\) or nonspecific as in T4.

Altogether, our results suggest that bacteriophage-encoded terminases share a conserved endonuclease domain that is distantly related to the resolvase/integrase/ribonuclease-H superfamily and that behaves as a properly folded and independent domain. The interplay between the...
Bioinformatics and Functional Study of Terminases

nuclease and packaging domains starts to be understood in the case of phage T4 (10) and A (7, 8). The recent sequencing of the T5 genome has allowed identification of the genes encoding the different partners of the encapsidation machinery. The unusual characteristics of the T5 genome (large terminal repeats, presence of genetically defined single-stranded interruptions) and of its mechanism of transfer into E. coli cells (recently reviewed in Ref. 48) makes it a new and interesting challenge to define the rules, and exceptions, that govern phage DNA packaging.

Acknowledgments—We thank M. Santamaria for assistance in the terminase constructs, P. Decottignies for protein sequencing, Chantal Janmot for technical assistance, and G. Craescu for access to the circular dichroism apparatus.

REFERENCES

1. Feiss, M., and Catalano, C. (2005) in Viral Genome Packaging Machines (Catalano, C., ed) pp. 5–39, Kluwer Academic/Plenum Publishers, New York
2. Earnshaw, W. C., and Casjens, S. R. (1980) Annu. Rev. Microbiol. 34, 267–292
3. Catalano, C. E., Cue, D., and Feiss, M. (1995) J. Virol. 69, 3069–3077
4. Koonin, E. V. (1993) Cell. Mol. Life Sci. 50, 14216–14224
5. Catalano, C. E. (2000) Bioinformatics 16, 72–79
6. Parris, W., Rubinchik, S., Yang, Y. C., and Gold, M. (1994) J. Biol. Chem. 269, 34988–34995
7. Saito, A., Iwasaki, H., Ariyoshi, M., Morikawa, K., and Shinagawa, H. (1995) J. Biol. Chem. 270, 13564–13574
8. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
9. Rentas, F. J., and Rao, V. B. (2003) J. Biol. Chem. 278, 35510–35517
10. Weng, Y., Catalano, C. E., and Feiss, M. (1996) Biochemistry 35, 7296–2803
11. Bhattacharyya, S. P., and Rao, V. B. (1993) Virology 196, 34–44
12. Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Bao, Q., Kong, H., and Hu, S. (2005) Virology 332, 45–65
13. Rentas, F. J., and Rao, V. B. (2003) J. Biol. Chem. 278, 35510–35517
14. Bhattacharyya, S. P., and Rao, V. B. (1993) Virology 196, 34–44
15. Davidson, A. R., and Gold, M. (1992) J. Mol. Biol. 69, 187–200
16. Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Bao, Q., Kong, H., and Hu, S. (2005) Virology 332, 45–65
17. Hwang, Y., Catalano, C. E., and Feiss, M. (1996) Biochemistry 35, 7296–2803
18. Bhattacharyya, S. P., and Rao, V. B. (1993) Virology 196, 34–44
19. Nishino, T., and Morikawa, K. (2002) Oncogene 21, 9022–9032
20. Rhoades, M., and Rhoades, E. A. (1972) J. Mol. Biol. 69, 187–200
21. Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Bao, Q., Kong, H., and Hu, S. (2005) Virology 332, 45–65
22. Everett, R. D. (1981) J. Gen. Virol. 52, 25–38
23. McCorduck, J. D., and Warner, H. R. (1988) in The Viruses (Calendar, R., ed) Vol. 1, pp. 439–476, Plenum Press, New York
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
25. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) J. Mol. Biol. 299, 499–520
26. Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) J. Mol. Biol. 310, 243–257
27. McGuillin, L. J., Bryson, K., and Jones, D. T. (2000) Bioinformatics 16, 404–405
28. Karplus, K., Barrett, C., and Hughey, R. (1998) Bioinformatics 14, 846–856
29. Douguet, D., and Labesse, G. (2001) Bioinformatics 17, 752–753
30. Labesse, G., and Morisson, J. (1998) Bioinformatics 14, 206–211
31. Dunbrack, R. L., Jr., and Karplus, M. (1993) J. Mol. Biol. 230, 543–574
32. Sali, A., and Blundell, T. L. (1993) J. Mol. Biol. 234, 779–815
33. Sippl, M. J. (1993) Proteins 17, 355–362
34. Eisenberg, D., Lutby, R., and Bowie, J. U. (1997) Methods Enzymol. 277, 396–404
35. Catherine, V., and Labesse, G. (2004) Bioinformatics 20, 3694–3696
36. Kabsch, W., and Sander, C. (1983) Biopolymers 22, 2577–2637
37. Ansaldi, M., Lepelletier, M., and Mejean, V. (1996) Anal. Biochem. 243, 110–111
38. Boulanger, P., le Maire, M., Bornhivers, M., Duboix, S., Desmadril, M., and Letellier, L. (1996) Biochemistry 35, 14216–14224
39. Mitchell, M. S., Matsuizaki, S., Imai, S., and Rao, V. B. (2002) Nucleic Acids Res. 30, 4009–4021
40. Mitchell, M. S., and Rao, V. B. (2004) Virology 321, 217–221
41. Parris, W., Rubinchik, S., Yang, Y. C., and Gold, M. (1994) J. Biol. Chem. 269, 13564–13574
42. Yang, W., and Steitz, T. A. (1995) Structure 3, 131–134
43.Combat, C., Blanchet, C., Geourjon, C., and Delege, G. (2000) Trends Biochem. Sci. 25, 147–150
44. Rice, P., Craigie, R., and Davies, D. R. (1996) Curr. Opin. Struct. Biol. 6, 76–83
45. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005) Cell 121, 1005–1016
46. Hwang, Y., and Feiss, M. (1996) J. Mol. Biol. 261, 524–535
47. Saito, A., Iwasaki, H., Ariyoshi, M., Morikawa, K., and Shinagawa, H. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 7470–7474
48. Ponchon, L., Mangenot, S., Boulanger, P., and Letellier, L. (2005) Biochim. Biophys. Acta 1724, 255–261