© 1993 Oxford University Press
Nucleic Acids Research, 1993, Vol. 21, No. 21 **4867–4872**

Genetic and biochemical analysis of an endonuclease encoded by the IncN plasmid pKM101

Robert F.Pohlman+, Franklin Liu§, Lu Wang, Margret I.Moré and Stephen C.Winans*
Section of Microbiology, Cornell University, Ithaca, NY 14853, USA

Received August 16, 1993; Revised and Accepted September 13, 1993

ABSTRACT

The IncN plasmid pKM101 *nuc* gene encodes a periplasmically localized endonuclease. DNA sequence analysis indicates that this gene encodes a hydrophilic protein of about 19.5 kDa containing a hydrophobic signal sequence. *nuc* is homologous to a partially sequenced open reading frame adjacent to the *sog* gene of the plasmid ColIB-P9, a plasmid known to encode an endonuclease similar to that of pKM101. A partially sequenced *tra* gene directly upstream of *nuc* is homologous to the *virB11* gene of *Agrobacterium tumefaciens.* We have partially purified the pKM101 nuclease by osmotic shock and cation exchange chromatography, and used this enzyme preparation to sequence the protein's amino terminus. The first 13 amino acids of the mature protein match amino acids 23 to 35 of the predicted sequence, indicating that the protein is proteolytically processed to a molecular mass of approximately 17 kDa, probably during export to the periplasmic space. The enzyme was able to attack many sites along an end labelled duplex DNA substrate, but showed clearly preferred cleavage sites, and may cleave preferentially at purine-rich regions.

INTRODUCTION

Plasmids can affect the processing and metabolism of host nucleic acids in interesting and complex ways. In addition to directing their own vegetative and conjugal DNA replication, plasmids often encode restriction and modification systems, anti-restriction systems, site-specific recombinases, fertility inhibition proteins, and proteins involved in DNA repair (5). The IncN plasmid pKM101 provides many good examples. This plasmid contains approximately 15 genes involved in conjugal transfer, entry exclusion, or fertility inhibition (33, 34). In addition, pKM101 contains genes whose products strongly inhibit Type I restriction enzymes (3), and probably contains a site-specific recombinase involved in plasmid stability (unpublished observations). Furthermore, this plasmid also encodes two genes that decrease the killing effects of a variety of DNA damaging agents, while enhancing their mutagenic effects (30). The ability of pKM101 to increase so-called error-prone DNA repair has resulted in its inclusion in the Ames *Salmonella typhimurium* mutagen detection strains, where it plays a major role in the sensitivity of these strains to mutagenic agents (19).

pKM101 also encodes a periplasmically localized, ATP-independent endonuclease which degrades single-stranded and duplex DNA. This enzyme lacks detectable exonuclease activity. It is maximally active in the presence of divalent cations, and uses magnesium, calcium, zinc, or cobalt with approximately equal efficiency (17). However, detectable endonuclease activity remains in the presence of EDTA; this provides a convenient assay, as host nuclease activities are not detected under these conditions (35). The enzyme generates 3′-hydroxyl and 5′-phosphate termini (17). Similar EDTA-resistant nucleases were detected in strains containing IncH, IncI or IncM plasmids, but were not detected in plasmids from 9 other incompatibility groups (35). The role of this protein in the biology of these plasmids was not determined.

At least one other site nonspecific endonuclease, bovine pancreatic deoxyribonuclease (DNase I) has been used as a biochemical tool for a variety of biochemical assays (12). The crystal structure of DNase I and of DNase I-substrate co-crystals has been described (27, 28). DNase I contacts phosphate residues of both strands of duplex DNA over a 10 bp region. It appears not to make contact with the major groove, but to interact with DNA exclusively in the minor groove. This enzyme may have two active sites (27), and at least one of these sites may use a proton acceptor-donor chain involving the residues Glu-His-water, in a reaction reminiscent of the Asp-His-Ser chain found in serine proteases. According to this model, DNase I removes a proton from water, and the resulting nucleophile attacks phosphorous directly, without any protein-DNA covalent intermediate. Unlike the pKM101 endonuclease, DNase I does not efficiently attack single-stranded DNA. In an effort to learn more about the pKM101 endonuclease and its potential as a biochemical tool, we have determined the DNA sequence of the *nuc* gene, partially purified the protein, and used this preparation for a preliminary examination of the enzyme's substrate specificity by using it to degrade an end-labelled DNA fragment of known sequence.

* To whom correspondence should be addressed

Present addresses: +Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139 and §Cornell University College of Medicine, New York, NY 10021, USA

MATERIALS AND METHODS

DNA isolation, sequencing and analysis

Plasmid DNA was isolated by the method of Birmboim and Doly (4) and further purified using Qiagen columns (Qiagen, Inc). This DNA was prepared for DNA sequencing (24) as follows: 3 μg of DNA was brought to 0.2 M NaOH/0.2mM EDTA in a 20 μl reaction volume for 5 minutes at room temperature. The reactions were supplemented with 8 μl of 5 M NH_4OAc and 100 μl of 100% EtOH and centrifuged using a tabletop micro centrifuge for 10 minutes. The supernatants were decanted and the pellets were vacuum dried and resuspended in 8 μl of water. The DNA was determined on both strands using modified T7 polymerase (Sequenase, U.S. Biochemicals) and other reagents supplied by the manufacturer. Synthetic oligonucleotides were synthesized using a P250A DNA synthesizer (Cruachem). Sequences adjacent to a Tn5 insertion were determined using the oligonucleotide 5′-AAACGGGAAAGGTTCCG-3′. Inferred protein sequences were used to search DNA sequence databases using TFASTA (21) running as part of the GCG DNA sequence analysis programs (9) at the Cornell Biovax. BLAST computations (2) were performed at the NCBI using the BLAST network service. The *nuc* DNA sequence was submitted to the GenBank DNA sequence database (accession number U00688).

Endonuclease purification

The endonuclease was partially purified from an overproducing *E.coli* strain (35). A 500 ml broth culture of JC2926(pGW1508) was grown to an O.D.(600) of 0.8 in LB broth plus 100 μg ampicillin per ml, centrifuged at 8,000 r.p.m. in a Sorval GSA rotor, resuspended in 10 ml of 100 mM Tris (pH 8.0), and centrifuged again. Cells were then resuspended in 5 mls of cold 20% sucrose, 30 mM Tris (pH 8.0) (buffer A) and 50 μl of freshly prepared lysozyme (1 mg/ml) and 100 μl of 50 mM EDTA were added. The cell suspension was incubated on ice for 30 min, at which time spheroplasts and intact cells were removed by centrifugation. 2.0 mls of supernatant were loaded onto a 1.0 ml MonoS FPLC column (Pharmacia) that had been pre-equilibrated with buffer A. Bound proteins were washed with 20 mls of buffer A and eluted with a 20 ml linear gradient from 0 M NaCl to 0.6 M NaCl. 1 ml fractions were collected. 15 μl aliquots of each fraction were size fractionated by SDS-PAGE using 8 cm long gels (Hoeffer Mighty Small) and stained with Coomassie brilliant blue. 15 μl aliquots were also size fractionated on similar gels containing duplex calf thymus DNA added to a final concentration of 1 μg/ml. After electrophoresis, gels containing DNA were washed in 2 changes of 2.5% Triton X-100 (1 hour each) to remove SDS, washed for 1 hour in buffer containing 50 mM $NaKPO_4$ and 10 mM $MgCl_2$, and then incubated overnight at room temperature in a fresh batch of the same buffer. Gels were stained with ethidium bromide added to a final concentration of 1 μg/ml and photographed over a UV transilluminator. The amino terminal sequence of this protein was determined by the Cornell Oligo/Peptide/Sequencing/Amino Acid Analysis Facility using an Applied Biosystems 470A Gas Phase Protein Sequencer equipped with a 120A PTH analyzer (14).

Determination of substrate specificity

A 226 bp PCR amplification product containing DNA from the *Agrobacterium* Ti plasmid (13), was synthesized by using a Hybaid thermocycler according to a published procedure (15). The oligonucleotides −5′-GCAGGATCCGGAACGCCTCG-ACCTGCCTGAGA-3′ and 5′-GCATCTAGAGGCCGAA-GCCCATGAGTTGAGAATA-3′ were used to prime DNA synthesis, and a plasmid containing the *occQ-occR* intergenic region was used as substrate. This PCR product was treated with *Bam*HI and *Xba*I and ligated to *Bam*HI, *Xba*I digested pTZ19R DNA, creating plasmid pLW102 (31). A DNA fragment shown in Figure 8A was isolated by cleavage with *Bam*HI and *Hinc*II, end labelled by treatment with modified T7 DNA polymerase (Sequenase) and α[32-P]dATP, dGTP, dCTP, and dTTP, size fractionated by electrophoresis, and eluted from the gel by overnight incubation in a buffer containing 10 mM Tris HCl (pH 7.6), and 50 mM NaCl. This DNA was incubated with approximately 10 ng of endonuclease for 10 min at 30°C in a buffer containing 10 mM Tris-hydrochloride (pH 7.6) and 50 mM $MgCl_2$. The samples were then boiled and size fractionated on a denaturing 5% polyacrylamide gel. The same DNA fragment was also subjected to chemical cleavage at G or A residues (10).

RESULTS

DNA sequence of *nuc*

Previous genetic data indicated that the pKM101 *nuc* gene was localized between *traG* and *fip*, that it lay to the left of the *Sma*I-1 site, and that it was transcribed from left to right as drawn in Figure 1. A derivative of pKM101 containing a Tn5 insertion in *nuc* was also previously isolated (35). In order to determine the sequence of *nuc* DNA adjacent to the Tn, pKM101 *nuc1210*::Tn5 was digested with *Sal*I and a DNA fragment containing about half of Tn5 (including its *neo* gene) and about 5 kilobases of pKM101 DNA ending at the *Sal*I-2 site was cloned into pUC118. This plasmid, pRP4013.2 (Figure 1), was subjected to double strand DNA sequencing using an oligonucleotide which hybridizes to the end of Tn5. The resulting *nuc* sequences allowed us to synthesize additional oligonucleotides and ultimately to 'walk' along the entire gene. The template DNA for these reactions was pRP4014.2 (Figure 1) which contains an intact *nuc* gene.

We report here the DNA sequence of a 1 kilobase region adjacent to the *Sma*I-1 site. This region contains an open reading frame (ORF) of 177 codons expressed from left to right, and encoding a protein having a molecular mass of 19,445 daltons (Figure 2). Confirmation that this ORF encodes *nuc* is provided below. The proposed initiation codon of this ORF is preceded by a possible ribosome binding site (25). We do not see any

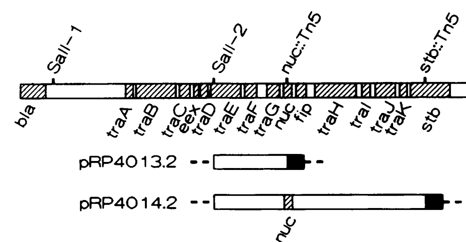

Figure 1. Physical and genetic map of the *tra* region of pKM101. Plasmid pRP4013.2 was constructed by cutting pKM101 *nuc1210*::Tn5 with *Sal*I, and cloning the indicated fragment into pUC118, while plasmid pRP4014.2 was constructed by the same strategy but using pKM101 *stb655*::Tn5. Black boxes represent Tn5 sequences (including *neo*) while dashed lines indicate pUC118 sequences. Descriptions of other pKM101 genes can be found in reference 33.

sequences that resemble any known *E.coli* promoter, suggesting that *nuc* might be expressed as part of an operon. The *nuc* ORF was preceded directly by another ORF reading in the same direction. The stop codon of upstream ORF partially overlaps the start codon of *nuc*, suggesting that their expression may be translationally coupled, and providing further support for a possible operon structure. On the basis of its map position and its homology to other *tra* genes (see below), we believe that the upstream ORF encodes *traG* (Figure 1).

The hydropathy profile of the predicted *nuc* product is shown in Figure 3. Its amino terminal 20 residues are highly hydrophobic, reminiscent of signal sequences that direct protein export. Following this region, the protein is fairly hydrophilic and has an extremely hydrophilic carboxyl terminus. The hydrophilic character of the mature protein is in accord with earlier observations that the endonuclease is highly soluble in aqueous buffers. The pKM101 endonuclease contains 2 histidine residues and numerous acidic residues, which could play a role in catalysis if this enzyme uses a mechanism similar to that of DNase I.

The endonuclease sequence was compared to the translated GenEMBL DNA sequence database (release 76) using TFASTA (21) and using BLAST (2). A partial ORF upstream of the *sog* gene of ColIb-P9 (26) was highly homologous to the carboxyl terminus of the endonuclease, showing 65% identity in a 34 amino acid overlap (Figure 4). The partially sequenced ORF upstream of *nuc* was also used to search the DNA sequence database. This protein fragment was similar to the VirB11 protein of *Agrobacterium tumefaciens* Ti plasmids, showing a 33% identity over 189 amino acids (Figure 3B and data not shown).

```
GTCACGGTCGATCACGTTGTAGAAGCCGTTTATATGATGTACGGC     45
V  T  V  D  H  V  V  E  A  V  Y  M  M  Y  G
'traG
GATGCAGGAAAGATCGGCCCGGTCAGCGCCACTGATGCCCTGCGA     90
D  A  G  K  I  G  P  V  S  A  T  D  A  L  R

GCCTGTATGCGTCTGACACCGGGCCGTATCATCATGACTGAGCTT    135
A  C  M  R  L  T  P  G  R  I  I  M  T  E  L

AGGGATGATGCTGCGTGGGATTATCTTAAAGCACTTAATACCGGC    180
R  D  D  A  A  W  D  Y  L  K  A  L  N  T  G

CATCCAGGCGGTGTTATGTCAACGCACGCTAACTCTGCGCGCGAT    225
H  P  G  G  V  M  S  T  H  A  N  S  A  R  D

GCCTTTAACCGTATTGGGCTGCTTATCAAGGCGACCCCTATCGGC    270
A  F  N  R  I  G  L  L  I  K  A  T  P  I  G

CGTATGCTCGATATGAGCGATATTATGCGAATGCTCTACTCCACC    315
R  M  L  D  M  S  D  I  M  R  M  L  Y  S  T

ATTGACGTTGTGGTGCATATGGAAAAGCGGAAAATCAAAGAAATT    360
I  D  V  V  V  H  M  E  K  R  K  I  K  E  I

TATTTTGACCCTGAATATAAAATGCAGTGTGTGAACGGGAGCCTG    405
Y  F  D  P  E  Y  K  M  Q  C  V  N  G  S  L

TAATGAAAAAATTAGCAACCTGGCTTCTGGCCGCAGCATTTACGA    450
*  *
 M  K  K  L  A  T  W  L  L  A  A  A  F  T  T
 nuc
CAGCCGCCCTGCCCGCCTTTGCGGTGGAACCATCCGTTCAGGTTG    495
 A  A  L  P  A  F  A  V  E  P  S  V  Q  V  G

GCTACTCGCCTGAAGGAAGCGCCCGCGTTCTCGTCCTGAGCGCCA    540
 Y  S  P  E  G  S  A  R  V  L  V  L  S  A  I

TTGACTCCGCAAAAACCTCGATACGGATGATGGCTTATTCTTTTA    585
 D  S  A  K  T  S  I  R  M  M  A  Y  S  F  T

CCGCCCCGGATATTATGAAGGCACTGGTAGCAGCCAAAAAACGGG    630
 A  P  D  I  M  K  A  L  V  A  A  K  K  R  G

GAGTTGATGTGAAAATCGTAATTGATGAAAGGGGCAATACGGGGC    675
 V  D  V  K  I  V  I  D  E  R  G  N  T  G  R

GCGCCAGCATTGCGGCCATGAACTACATAGCGAACAGCGGCATCC    720
 A  S  I  A  A  M  N  Y  I  A  N  S  G  I  P
                                ↓nuc1210::Tn5
CTTTGCGTACTGACAGTAATTTCCCTATCCAGCATGACAAGGTGA    765
 L  R  T  D  S  N  F  P  I  Q  H  D  K  V  I

TCATCGTGGATAATGTGACCGTTGAAACTGGCAGCTTTAATTTCA    810
 I  V  D  N  V  T  V  E  T  G  S  F  N  F  T

CCAAAGCGGCCGAAACGAAAAACTCGGAGAATGCGGTGGTGATCT    855
 K  A  A  E  T  K  N  S  E  N  A  V  V  I  W

GGAACATGCCTAAGCTCGCTGAATCATTCCTCGAACACTGGCAAG    900
 N  M  P  K  L  A  E  S  F  L  E  H  W  Q  D

ACCGCTGGAATCAGGGAAGAGACTACCGTTCCAGCTACTGAATAC    945
 R  W  N  Q  G  R  D  Y  R  S  S  Y  *

TGACCGCCAAAAOCTTTTTTTGTTCAAAGAAACTGATCTCAGTAT    990

CGCGCCCGGG                                      1000
```

Figure 2. DNA sequence of *nuc* and part of *traG*. A possible *nuc* ribosome binding site is overlined, while the amino terminal 13 residues of the mature protein are underlined. The arrow at residue 750 indicates the point of insertion of Tn*5* in the *nuc1210* mutant.

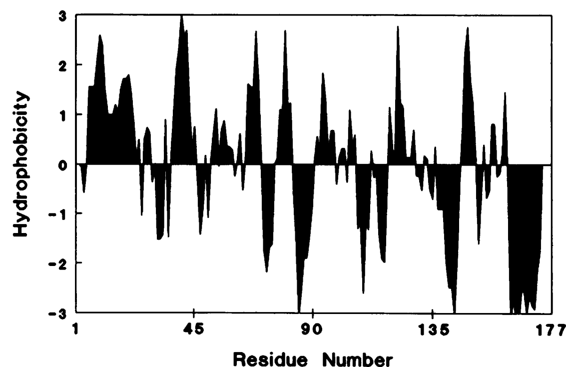

Figure 3. Hydropathy profile of the pKM101 pro-endonuclease, including its signal sequence. The Kyte and Doolittle algorithm (16) was used to plot the hydropathy of an average of seven amino acid residues as a function of amino acid residue number.

A.

```
                   10        20        30        40
Nuc        MKKLATWLLAAAFTTAALPAFAVEPSVQVGYSPEGSARVL

                   50        60        70        80
Nuc        VLSAIDSAKTSIRMMAYSFTAPDIMKALVAAKKRGVDVKI

                   90       100       110       120
Nuc        VIDERGNTGRASIAAMNYIANSGIPLRTDSNFPIQHDKVI

                  130       140       150       160
Nuc        IVDNVTVETGSFNFTKAAETKNSENAVVIWNMPKLAESFL
                                ::::.::. :::..:. .:
colI ORF                       ....NSENVVVLKNMPDVAAQYL
                                           10

                       170
Nuc        EHWQDRWNQGRDYRSSY*
           :::::::::.: :.:.
ColI ORF EHWQDRWNKGTDWRP*
              20        30
```

B.

```
              100       110        120       130
'TraG      VTVDHVVEAVYMMYGDAGK-IGPVSATDALRACMRLTPGR
           ... :   . : ..:...:  .:.:::.. :.:..:. :.:
VirB11     LVIPH-DNHVRLLYSKNGAGLGAVSAEHLLQASLRMRPDR

               140       150       160       170
'TraG      IIMTELRDDAAWDYLKALNTGHPGGVMSTHANSARDAFNR
           :...:.::::::.::... .::::.. ..:...  ..:..
VirB11     ILLGEMRDDAAWAYLSEVVSGHPGSISTIHGANI-QGFKK

               180       190       200       210
'TraG      IGLLIKATPIGRMLDMSDIMRMLYSTIDVVVHMEKRKIKE
           .  :.:..  :   :. .... :: ..:::....  :  ..
VirB11     LFSLVKSSVQGASLEDRTLIDMLSTAIDVIIPF--RAYED

               220       230
'TraG      IYFDPEYKMQCVNGSL*
           .:
VirB11     VYEVGEIWLAADARRRGETIGDLLNQ*
```

Figure 4. Homology between sequenced pKM101 genes and other genes in the GenBank database. **A.** Similarity between pKM101 endonuclease and the probable endonuclease of plasmid ColIb-P9. **B.** Similarity between the probable pKM101 TraG protein and the VirB11 protein of *Agrobacterium tumefaciens*. Colons indicate amino acid identity between these two proteins, while periods denote conservative substitutions.

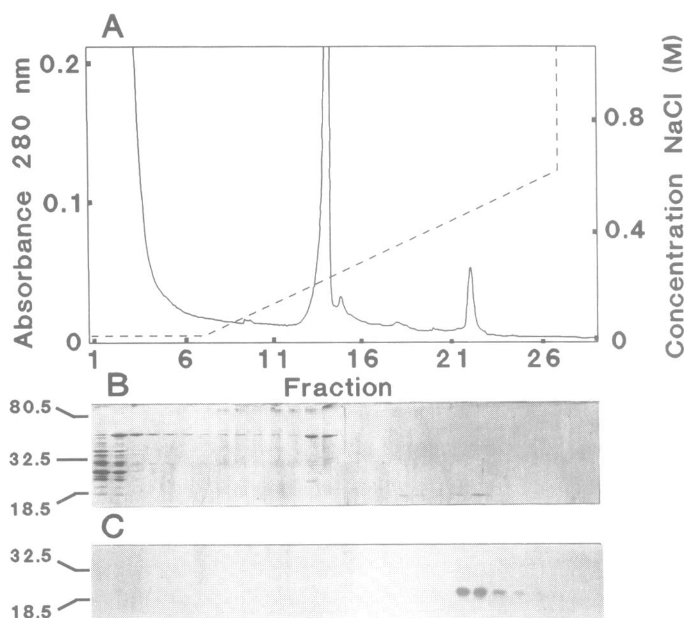

Figure 5. Purification of the endonuclease **A:** Cation exchange chromatography. The solid line indicates OD(280) of the eluate (scale on left axis), while the dashed line indicates the NaCl concentration used to elute proteins (scale on right axis). **B:** Coomassie blue stained samples of each fraction after size fractionation by SDS-PAGE (18). C: Nuclease activity gel assays of each fraction using polyacrylamide gels containing calf thymus DNA.

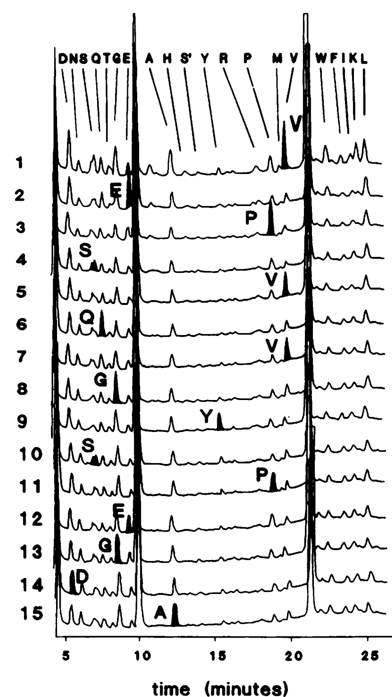

Figure 6. Amino terminal peptide sequence of the endonuclease. Letters across the top indicate positions at which each amino acid elutes. Numbers on the left side of each trace refer to Edman degradation cycle. Amino acids released from the endonuclease amino terminus during each cycle are filled in. All other peaks are due to contaminating amino acids.

Purification of the endonuclease

The endonuclease was partially purified from an *E.coli* strain containing a plasmid in which the *nuc* gene was fused to a *tet* promoter (35). About 75% of cellular endonuclease activity was released from concentrated cells by osmotic shock (20), and this fraction was purified free of detectable contaminating proteins by cation exchange chromatography as described in the Materials and Methods. A single peak of nuclease activity eluted at approximately 0.5 M NaCl (Figure 5A).

The amino terminus sequence of this protein was determined by automated Edman degradation. Importantly, the resulting data were interpreted without any knowledge of the predicted amino acid sequence. 19 amino acid residues were analyzed, and of these residues 1–13 precisely match residues 23–35 of the *nuc* open reading frame (underlined amino acid residues in Figure 2), and the 15th residue is identical to residue 37 (Figure 6). This finding provides a definitive confirmation that the predicted *nuc* gene does in fact encode the enzyme that we had purified. We can also conclude that the primary translation product must be proteolytically processed to form the mature protein, probably by the periplasmic peptidase that removes signal peptides during protein export (22). The mature protein would therefore contain 155 amino acid residues and have a molecular mass of 17,155 kDa, which is slightly larger than the predicted molecular mass as estimated by SDS-PAGE.

Substrate specificity

The pKM101 endonuclease was tested for site-specificity by using it to digest a 222 bp end-labelled duplex DNA (Figure 7). The endonuclease creates detectable cleavage products at approximately half of all residues, although a small number of highly preferred cleavage sites and a larger number of less preferred sites were detected. These sites are shown on Figure 8A, and these sites are aligned in Figure 8B in an effort to identify any common sequence determinants. These cleavage sites do not share any obvious similarity, although the enzyme may show some preference for purines at the first, second, and third bases 5′ to the cleavage site, and at the second base 3′ to the cleavage site. However, a larger sample of cleavage sites will be required to determine whether this pattern is significant.

DISCUSSION

To begin a molecular analysis of an endonuclease encoded by pKM101, we have determined the DNA sequence of the *nuc* gene, purified the enzyme, and analyzed its substrate specificity. This periplasmic protein is processed from a larger precursor during export across the cytoplasmic membrane. Its mature molecular mass is approximately 17 kDa. Earlier studies showed that the native molecular mass was approximately 70 kDa, suggesting that the protein has a tetrameric quaternary structure. The enzyme attacks many but not all phosphodiester bonds in duplex DNA, and its preferred cleavage sites do not share any obvious sequence beyond a possible purine richness.

The carboxyl terminus of the pKM101 endonuclease is homologous to a partially sequenced open reading frame upstream of the *sog* gene of ColIb-P9. It is noteworthy that an EDTA-resistant endonuclease activity was detected in all tested strains containing IncI plasmids (35), and that a gene encoding such an enzyme in ColIb-P9 was previously localized close to *sog* (23). The homology between *nuc* and this ORF suggests strongly that the ColIb-P9 *nuc* gene lies 94 bp upstream of *sog*. Both *nuc* and the COlIb-P9 ORF lie directly downstream of a cluster of genes required to synthesize conjugal pili (23).

Figure 7. Substrate specificity of the pKM101 endonuclease. Lane 1: Products of partial endonuclease digestion of a 3′ end labelled duplex template. Lane 2: Cleavage products of the same template after exposure to chemical reagents that cleave at G or A residues (10).

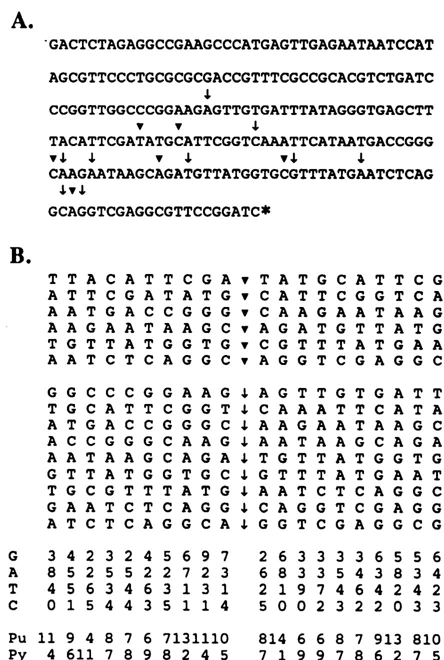

Figure 8. Preferred endonuclease cleavage sites. **A:** The DNA sequence of the duplex substrate used in Figure 7. The asterix denotes the 3′ label. Highly preferred cleavage sites (▼) and less highly preferred cleavage sites (↓) are indicated above the 5′ end of the resulting fragment. **B:** Alignment of highly preferred (▼) and less highly preferred (↓) cleavage sites.

It is interesting that IncN, IncI, IncH, and IncM plasmids share EDTA-resistant endonucleases, and in light of the homology described above, it is tempting to speculate that all these enzymes may be homologous. While other plasmids did not direct the synthesis of detectable EDTA-resistant enzymes, data from other groups indicates that at least one IncW plasmid (7) and probably at least one IncP plasmid and one IncF1 plasmid encode periplasmic EDTA-sensitive nucleases (11, 36). The IncW, IncP, and IncF1 enzymes are homologous to each other and to a *Staphylococcus aureus* nuclease (8). The roles of each of these enzymes in the biology of their respective plasmids remains to be determined. The pKM101 endonuclease was not required for any tested role, including error-prone DNA repair, conjugation, and entry exclusion. However, the fact that this gene appears to be translationally coupled to an upstream cluster of *tra* genes suggests that this enzyme might play some ancillary role in conjugation or a related process.

The *nuc* initiation codon partially overlaps the termination codon of the upstream *traG* gene. The lack of any intergenic DNA between *traG* and *nuc* was surprising for two reasons. First, a Tra^+, Nuc^+ insertion of Mu*d* Ap *lac* was previously isolated and was thought to lie between these two genes (33). This insertion most likely fell within the carboxyl terminus of *traG*, indicating that this part of the protein is dispensable for conjugation. The second reason that the *traG-nuc* linkage was surprising is that transposon insertions in *traG* are not polar on *nuc* (33). We conclude either that *traG* contains an internal promoter which expresses *nuc*, or that *traG* transcription is not terminated by insertions either of Tn*5* or of Mu*d* Ap *lac*. We have found other examples of pKM101 *tra* genes whose sequences suggest that they are part of an operon, but whose transcription is apparently not terminated by transposon insertions in upstream genes (unpublished data).

The observed homology between TraG and VirB11 is reminiscent of other studies showing that VirB11 is homologous to the TrbB(KlbA) protein of plasmid RK2, which, like the pKM101 TraG protein, is required for conjugal transfer and for sensitivity to male-specific bacteriophage (26, 29). This finding also provides further weight to the hypothesis that the *A. tumefaciens vir* regulon evolved from a conjugal transfer system. TraG, TrbA, and VirB11 are all homologous to a larger family of proteins, including ComG of *Bacillus subtilis* (1) and PtlG of *Bordetella pertussis* (32), each of which appears to act in transporting macromolecules across the bacterial cell envelope. These proteins generally have ATP binding site motifs, and VirB11 was shown to bind and hydrolyze ATP (6). TraG also has an ATP binding site motif (data not shown), and it is therefore tempting to speculate that this protein may provide energy for the conjugal export of DNA.

ACKNOWLEDGEMENTS

We thank Ted Thannhauser for expert assistance in peptide sequencing. We also thank Clay Fuqua, Brian Wilkins, Laura Frost, and Karin Ippen-Ihler for helpful discussions. This work was supported by N. I. H. (Grant GM2893). We also thank the Pew Charitable Trusts and the National Science Foundation Instrumentation and Laboratory Improvement Program (Grant

USE 9152210) for contributing to the development of the undergraduate molecular research facilities at Colgate University.

REFERENCES

1. Albano, M., Breitling, R., and Dubnau, D. B. (1989). *J. Bacteriol.* **171**,5386–5404.
2. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). *J. Mol. Biol.* **215**,403–410.
3. Belogurov, A. A., Delver, E. P., and Rodzevich, O. V. (1993). *J. Bacteriol.* **175**,4843–4850.
4. Birnboim, H. C. and Doly, J. (1979). *Nucl. Acids Res.* **7**,725–727.
5. Bukhari, A. I., Shapiro, J. A., and Adhya, S. L. (1977). DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
6. Christie, P. J., Ward, J. E., Gordon, M. P., and Nester, E. W. (1989). *Proc. Natl. Acad. Sci. USA* **86**,9677–9681.
7. Close, S. M., and Kado, C. I. (1992). *Mol. Microbiol.* **6**,521–527.
8. Cotton, F. A., Hazen Jr., E. E. and Legg, M. J. (1979). *Proc. Natl. Acad. Sci. USA* **76**,2551–2555.
9. Devereux, J., Haeberli, P., and Smithies, O. (1984). Nucl. Acids Res. **12**:387–395.
10. Eckert, R. L. (1987). in: Current Protocols in Molecular Biology, ed. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (John Wiley & Sons). pp 7.5.1–7.5.9.
11. Gerlitz, M., Hrabak, O., and Schwab, H. (1990). *J. Bacteriol.* **172**,6194–6203.
12. Galas, D. J. and Schmitz, A. (1978). *Nucl. Acids Res.* **5**,3157–3170
13. Habeeb, L., Wang, L., and Winans, S.C. (1991). *Molec. Plant-Microbe Interact.* **4**:379–385.
14. Hunkapiller, M. W., Granlund-Moyer, K., and Whiteley, N. W. (1986). in: Methods of Protein Microcharacterization, ed. Shively, J. E. (Humana Press) pp. 223–247.
15. Innis, M. A., and Gelfand, D. H. (1990). in: PCR Protocols, ed. Innis, M. A., Gelfand, D. H., Sninsky, J. J., White, T. J. (Academic Press) pp. 3–12.
16. Kyte, J. and Doolittle, R. F. (1982). *J. Mol. Biol.* **157**,105–132.
17. Lackey, D., Walker, G. C., Keng, T., and Linn, S. (1977). *J. Bacteriol.* **131**,583–588.
18. Laemmli, U. K. (1970). *Nature* **227**,680–685.
19. McCann, J., Springar, N. E., Kobori, J. and Ames, B. N. (1975). *Proc. Natl. Acad. Sci. U. S. A.* **79**,979–983.
20. Neu, H. C., and Heppel, L. A. (1964). *J. Biol. Chem.* **239**,3893–3899.
21. Pearson, W. R. and Lipman, D. J. (1988). *Proc. Natl. Acad. Sci. USA.* **85**,2444–2448.
22. Perlman, D., and Halvorson, H. O. (1983). *J. Mol. Biol.* **167**,391–409.
23. Rees, C. E. R., Bradley, D. E., and Wilkins, B. M. (1987). *Plasmid* **18**,223–236.
24. Sanger, F., Nicklen, S. and Coulson, A. R. (1977). *Proc. Natl. Acad. Sci. USA* **74**,5463–5467.
25. Shine, J. and Dalgarno, L. (1974). *Proc. Natl. Acad. Sci. USA* **71**,1342–1346.
26. Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992). *J. Biol. Chem.* **267**,13062–13072.
27. Suck, D., Lahm, A. and Oefner, C. (1988). *Nature* **332**,464–468.
28. Suck, D. and Oefner, C. (1986). *Nature* **321**,620–625.
29. Thomson V.J., Jovanovic, O.S., Pohlman, R.F., Chang, C.H., Figurski, D.H. (1993). J. Bacteriol. **175**,2423–2435.
30. Walker, G. C. (1985). *Annu. Rev. Biochem.* **54**,425–457.
31. Wang, L., Helmann, J. A., and Winans, S. C. (1993). *Cell* **69**,659–667.
32. Weiss, A. A., Johnson, F. D. and Burns, D. L. (1993). *Proc. Natl. Acad. Sci. USA* **90**,2970–2974.
33. Winans, S. C. and Walker, G. C. (1985). *J. Bacteriol.* **161**,402–410.
34. Winans, S. C. and Walker, G. C. (1985). *J. Bacteriol.* **161**,425–427.
35. Winans, S. C. and Walker, G. C. (1982). *J. Bacteriol.* **154**,1117–1125.
36. Yoshioka, Y., Fujita, Y. and Ohtsubo, E. (1990). *J. Mol. Biol.* **214**,39–53.