

## Human PIR1 of the Protein-tyrosine Phosphatase Superfamily Has RNA 5'-Triphosphatase and Diphosphatase Activities\*

(Received for publication, February 12, 1999, and in revised form, March 15, 1999)

Tarangini Deshpande<sup>‡</sup>, Toshimitsu Takagi<sup>§¶</sup>, Luning Hao<sup>||</sup>, Stephen Buratowski<sup>§\*\*</sup>, and Harry Charbonneau<sup>||‡‡</sup>

From the Departments of <sup>‡</sup>Biological Sciences and <sup>||</sup>Biochemistry, Purdue University, West Lafayette, Indiana 47907 and the <sup>§</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

**A human cDNA was isolated encoding a protein with significant sequence similarity (41% identity) to the BVP RNA 5'-phosphatase from the *Autographa californica* nuclear polyhedrosis virus. This protein is a member of the protein-tyrosine phosphatase (PTP) superfamily and is identical to PIR1, shown by Yuan *et al.* (Yuan, Y., Da-Ming, L., and Sun, H. (1998) *J. Biol. Chem.* 272, 20347–20353) to be a nuclear protein that can associate with RNA or ribonucleoprotein complexes. We demonstrate that PIR1 removes two phosphates from the 5'-triphosphate end of RNA, but not from mononucleotide triphosphates. The specific activity of PIR1 with RNA is several orders of magnitude greater than that with the best protein substrates examined, suggesting that RNA is its physiological substrate. A 120-amino acid segment C-terminal to the PTP domain is not required for RNA phosphatase activity. We propose that PIR1 and its closest homologs, which include the metazoan mRNA capping enzymes, constitute a subgroup of the PTP family that use RNA as a substrate.**

The protein-tyrosine phosphatase (PTP)<sup>1</sup> superfamily includes a large number of enzymes that dephosphorylate diverse substrates including proteins, nucleic acids, and lipids (1–6). Members of the PTP superfamily are thought to use a common catalytic mechanism involving the formation and subsequent hydrolysis of a phosphocysteine intermediate (1–6). The essential Cys and Arg residues are located within an active site motif (HCX<sub>5</sub>R) that characterizes all phosphatases of this superfamily.

The *Autographa californica* nuclear polyhedrosis virus expresses a 19-kDa phosphatase of the PTP superfamily desig-

nated herein as BVP (also known as BVH1 and BVPTP) (7–9). BVP was originally characterized as a dual specificity protein phosphatase (7–9), but subsequent studies have demonstrated that its RNA phosphatase activity is several orders of magnitude greater than its activity with protein substrates (3, 4). BVP shares significant sequence similarity with the RNA triphosphatase domain of the metazoan mRNA capping enzymes (2–4, 10). The bifunctional capping enzymes of metazoa contain an N-terminal RNA 5'-triphosphatase domain and a C-terminal GTP::RNA guanylyltransferase domain. The RNA triphosphatase domain removes the  $\gamma$ -phosphate from the 5' end of nascent mRNA to leave a diphosphate terminus and the guanylyltransferase domain catalyzes the subsequent transfer of a guanylyl group from GTP to produce the unmethylated 5' cap structure, G(5')ppp(5')N (11, 12). BVP and the triphosphatase domains of the capping enzymes contain the signature HCX<sub>5</sub>R active site motif common to all PTPs and are thought to employ a catalytic mechanism similar to that used by PTPs to dephosphorylate proteins (2–4, 10, 13, 14). BVP differs from the metazoan capping enzymes in that it lacks a guanylyltransferase domain and releases both  $\gamma$ - and  $\beta$ -phosphates from mRNA to yield a monophosphate at the 5' end (3). BVP is unlikely to be involved in the capping of viral messages because LEF-4, a subunit of the *A. californica* nuclear polyhedrosis virus RNA polymerase, is a bifunctional capping enzyme with RNA triphosphatase and guanylyltransferase activities (15–17). The function of BVP is unknown, but it is not essential for viral replication (18).

Recently, Yuan *et al.* (19) identified and cloned human PIR1, a nuclear phosphatase that interacts with RNA or ribonucleoproteins. To identify a human homolog of BVP, we have independently cloned PIR1 and have expressed the recombinant glutathione *S*-transferase (GST) fusion protein in *Escherichia coli* for further enzymatic characterization. We demonstrate that at comparable substrate concentrations, the RNA triphosphatase and diphosphatase activities of GST-PIR1 exceed its protein phosphatase activity by 2 or more orders of magnitude. The strong preference for RNA substrates and its nuclear localization suggest a potential role for PIR1 in RNA processing. We also show that the C-terminal, noncatalytic domain of PIR1 is not essential for its RNA phosphatase activity.

### EXPERIMENTAL PROCEDURES

**Plasmids and Site-directed Mutagenesis**—A cDNA containing the PIR1 open reading frame was amplified by polymerase chain reaction using oligonucleotide primers with 5'-restriction linkers to facilitate cloning. This fragment was subcloned into the *Bam*HI/*Hind*III sites of pET21a-GST (20). A modification of a polymerase chain reaction-based site-directed mutagenesis method (21) was used to change PIR1 Cys<sup>152</sup> → Ser and introduce a silent mutation within the codon for Arg<sup>158</sup> to yield an *Apa*LI site that was used for screening. The C-terminal truncation mutant was generated by amplifying the open reading frame

\* This work was supported in part by National Institutes of Health Grant CA59935 and a Junior Faculty Research Award from the American Cancer Society (to H. C.). This is Paper 15935 from the Purdue University Agricultural Experimentation Station. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence encoding the PIR1 phosphatase characterized in this paper has been submitted to the GenBank™/EMBL data bank with the accession number AF023917.

<sup>¶</sup> Senior Postdoctoral Fellow of the American Cancer Society, Massachusetts Division, Inc.

\*\* Recipient of an American Cancer Society Junior Faculty Research Award and a Pew Scholar in the Biomedical Sciences Award.

<sup>‡‡</sup> To whom correspondence should be addressed: Purdue University, 1153 Biochemistry, West Lafayette, IN 47907. Tel.: 765-494-4754; Fax: 765-496-6395; E-mail: charb@biochem.purdue.edu.

<sup>1</sup> The abbreviations used are: PTP, protein-tyrosine phosphatase; GST, glutathione *S*-transferase; PEI, polyethyleneimine; BVP, baculoviral phosphatase; PIR1, phosphatase that interacts with RNA and/or ribonucleoproteins; HCE, human capping enzyme; MCE, mouse capping enzyme.

with an antisense primer containing a stop codon after amino acid residue 206. Plasmids were sequenced to confirm their authenticity.

**Expression and Purification of Recombinant Proteins**—*E. coli* BL21(DE3) cells transformed with the GST-PIR1 constructs were grown in LB containing 100 µg/ml ampicillin and 2% (w/v) glucose at 37 °C until the OD<sub>600</sub> was 0.7. The cells were induced with 200 µM isopropyl thio-β-D-galactopyranoside in fresh media, grown for 5–7 h at room temperature, and harvested. GST-PIR1 was affinity purified from bacterial extracts using glutathione-Sepharose as described previously (20). GST-BVP was expressed and purified as described previously (3).

**Protein Phosphatase Assays**—Protein substrates were phosphorylated on tyrosyl residues using recombinant GST-*lyn* kinase and on seryl/threonyl residues using the catalytic subunit of cAMP-dependent kinase as described (3). Phosphatase reactions with *p*-nitrophenyl phosphate and protein substrates were carried out in 50 mM Tris, pH 7.9, 10 mM KCl, 10 mM dithiothreitol for 30 min at 30 °C as described previously (20). Protein was determined by the method of Bradford (22) using bovine serum albumin as a standard. The Coomassie staining intensity of samples on a SDS-polyacrylamide gel suggested that the amount of full-length GST-PIR1 but not GST-PIR1 (1–206) may have been overestimated by as much as 5-fold. However, this apparent error in protein determination has not yet been confirmed by independent methods.

**RNA 5'-Phosphatase Assays**—Triribonucleotide RNA substrates were prepared with [ $\gamma$ -<sup>32</sup>P]ATP (pppApCpC; bold letter denotes labeled phosphate) and [ $\alpha$ -<sup>32</sup>P]ATP (pppApCpC) and used in RNA triphosphatase assays as described previously (3). RNA diphosphatase assays were carried out with diphosphate-terminated triribonucleotides (ppApCpC), prepared with [ $\alpha$ -<sup>32</sup>P]CTP (3). Reaction products were analyzed by thin-layer chromatography (TLC) on polyethyleneimine (PEI)-cellulose plates (3). Phosphate (P<sub>i</sub>) released from the termini of RNA and monophosphate-terminated RNA (pApCpC) was detected by using a Fuji BasX PhosphorImager and autoradiography. Radioactive spots were cut from PEI-cellulose plates and quantitated by liquid scintillation counting. Assays were linear with respect to time and substrate concentration.

For the analysis of the 5' end, RNA was further incubated for 30 min at 37 °C with 2 ng of RNase T2 (purified from *Aspergillus oryzae* and supplied by Dr. S. Norioka, Institute For Protein Research, Osaka University, Japan) in 20 mM ammonium acetate, pH 5.8. After digestion, the products were analyzed by TLC on PEI-cellulose plates with adenosine 3'-monophosphate (Ap) and adenosine 3',5'-diphosphate (pAp) (Sigma) as authentic markers.

## RESULTS

An expressed sequence tag (GenBank™ accession number H60626) encodes a polypeptide exhibiting 47% amino acid sequence identity to an 87-residue segment of the BVP phosphatase (3). Using a probe derived from the H60626 expressed sequence tag, a 1.6-kilobase cDNA was isolated from a human placenta cDNA library (CLONTECH). This cDNA encoded PIR1 (GenBank™ AF023917) that was cloned by Yuan *et al.* (19) while this work was in progress. In accord with Yuan *et al.* (19), Northern analysis detected a single 1.7-kilobase PIR1 transcript in eight different human tissues (data not shown). A sequence-tagged site (GenBank™ accession number AA037694) identical to the PIR1 sequence has been mapped to the short arm of human chromosome 2.

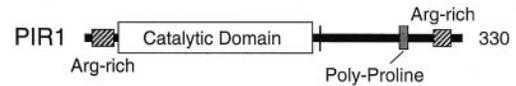
PIR1 is a 38.9-kDa phosphatase that is localized to the nucleus, is characterized by its ability to bind RNA *in vitro* and to interact with the 9G8 and Srp30C splicing factors in the yeast two-hybrid system (19). As shown in the sequence alignment of Fig. 1A, an N-terminal segment of PIR1 exhibits a high degree of sequence similarity to corresponding segments from the BVP RNA phosphatase and the RNA triphosphatase domain of the metazoan mRNA capping enzymes. This conserved region of PIR1 encompasses the hallmark PTP active site motif (HCX<sub>5</sub>R) (1). PIR1 contains a C-terminal segment (residues 207–330) with a polyproline-rich region (residues 276–285) (Fig. 1B) that might serve as a binding site for SH3/WW domain-containing proteins (23).

**Protein Phosphatase Activity of PIR1**—GST-PIR1 dephosphorylates phosphotyrosyl-containing protein substrates and at least one substrate phosphorylated on serine/threonine res-

## A

PIR1	H <b>E</b> P <b>E</b> F <b>K</b> D <b>Y</b> L <b>P</b> V <b>C</b> <b>Q</b> R <b>M</b> P <b>C</b> T <b>R</b> F <b>T</b> A <b>R</b> <b>G</b> V <b>P</b> L <b>I</b> Q <b>K</b> S <b>E</b> --K <b>K</b> L <b>A</b> P <b>E</b> E--C <b>F</b> S <b>P</b> L <b>D</b> E <b>F</b> N <b>K</b> I <b>R</b> E <b>Q</b> N <b>E</b> E
T23G7	R <b>P</b> D <b>P</b> <b>N</b> L <b>I</b> D <b>N</b> V <b>E</b> R <b>D</b> D <b>G</b> T <b>R</b> F <b>V</b> P <b>P</b> K <b>P</b> L <b>D</b> S <b>S</b> F <b>D</b> G <b>K</b> N <b>M</b> P <b>V</b> L <b>L</b> <b>Q</b> G <b>V</b> K <b>T</b> L <b>I</b> S <b>L</b> A <b>Q</b> A <b>N</b> K <b>O</b>
BVP	M <b>E</b> P <b>A</b> F <b>H</b> I <b>N</b> L <b>Y</b> L <b>Q</b> C <b>V</b> <b>V</b> <b>K</b> D <b>S</b> N <b>L</b> I <b>C</b> R <b>K</b> F <b>R</b> L <b>P</b> E <b>L</b> F--A <b>Y</b> T <b>S</b> E <b>D</b> --V <b>W</b> T <b>A</b> Q <b>I</b> L <b>K</b> O <b>N</b> P <b>S</b>
BMVP	M <b>E</b> P <b>A</b> F <b>H</b> I <b>N</b> L <b>Y</b> L <b>Q</b> C <b>V</b> <b>V</b> <b>K</b> D <b>S</b> N <b>L</b> I <b>C</b> R <b>K</b> F <b>R</b> L <b>P</b> E <b>L</b> F--A <b>Y</b> T <b>S</b> E <b>D</b> --V <b>W</b> T <b>T</b> Q <b>I</b> V <b>K</b> O <b>N</b> P <b>S</b>
MCE	K <b>E</b> P <b>P</b> <b>M</b> L <b>N</b> C <b>P</b> R <b>R</b> <b>C</b> P <b>V</b> A <b>G</b> --R <b>E</b> L <b>P</b> <b>K</b> R <b>I</b> L <b>G</b> P <b>R</b> D <b>D</b> --S <b>Q</b> A <b>E</b> N <b>E</b> --R <b>H</b> P <b>S</b> M <b>L</b> S <b>N</b> L <b>K</b> S <b>L</b> K <b>V</b> K
HCE	K <b>E</b> P <b>P</b> <b>M</b> L <b>N</b> C <b>P</b> R <b>R</b> <b>C</b> P <b>V</b> A <b>G</b> --R <b>E</b> L <b>P</b> <b>K</b> R <b>I</b> L <b>G</b> P <b>R</b> D <b>D</b> --S <b>Q</b> A <b>E</b> N <b>E</b> --R <b>H</b> P <b>S</b> M <b>L</b> S <b>N</b> L <b>K</b> S <b>L</b> K <b>V</b> K
CEL-1	G <b>L</b> P <b>E</b> <b>M</b> L <b>H</b> C <b>P</b> K <b>T</b> <b>L</b> E <b>N</b> N--L <b>F</b> R <b>P</b> P <b>K</b> <b>P</b> L <b>C</b> K <b>M</b> D <b>D</b> --N <b>Q</b> A <b>E</b> R <b>Y</b> Q <b>H</b> P <b>A</b> E <b>V</b> F <b>S</b> H <b>P</b> L <b>H</b> G <b>K</b>
F54C8	V <b>M</b> R <b>K</b> <b>D</b> <b>S</b> K <b>F</b> Q <b>P</b> V <b>E</b> <b>N</b> V <b>P</b> R <b>T</b> R <b>I</b> V <b>E</b> R <b>P</b> <b>I</b> N <b>S</b> L <b>S</b> --T <b>K</b> H <b>K</b> Q <b>E</b> --R <b>E</b> T <b>T</b> N <b>D</b> L <b>F</b> R <b>Q</b> L <b>S</b> E <b>R</b> G <b>Q</b> Y
PIR1	I <b>G</b> L <b>V</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
T23G7	I <b>G</b> L <b>V</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
BVP	I <b>G</b> A <b>L</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
BMVP	I <b>G</b> A <b>L</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
MCE	I <b>G</b> L <b>V</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
HCE	I <b>G</b> L <b>V</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
CEL-1	I <b>G</b> L <b>V</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
F54C8	I <b>G</b> L <b>V</b> <b>D</b> L <b>I</b> <b>S</b> E <b>T</b> <b>P</b> R <b>T</b> <b>V</b> <b>D</b> K <b>D</b> L <b>T</b> G <b>M</b> C <b>V</b> <b>L</b> <b>H</b> E <b>V</b> <b>N</b> C <b>F</b> <b>R</b> G <b>F</b> <b>I</b> R <b>D</b> D <b>C</b> V <b>S</b> E <b>F</b> Q <b>V</b> L <b>Q</b> <b>D</b> <b>T</b> K <b>D</b> C <b>D</b> E
PIR1	D <b>K</b> L <b>L</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> L <b>L</b> C <b>R</b> <b>I</b> <b>L</b> <b>D</b> <b>V</b> E <b>G</b> V <b>R</b> P <b>D</b> D <b>A</b> E <b>L</b> <b>N</b> R <b>C</b> R <b>S</b> H <b>C</b> L <b>E</b> R <b>O</b> N <b>I</b> <b>L</b> E <b>I</b> R
T23G7	D <b>K</b> L <b>L</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> L <b>L</b> C <b>R</b> <b>I</b> <b>L</b> <b>D</b> <b>V</b> E <b>G</b> V <b>R</b> P <b>D</b> D <b>A</b> E <b>L</b> <b>N</b> R <b>C</b> R <b>S</b> H <b>C</b> L <b>E</b> R <b>O</b> N <b>I</b> <b>L</b> E <b>I</b> R
BVP	G <b>M</b> L <b>V</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> M <b>V</b> C <b>R</b> <b>I</b> <b>L</b> <b>M</b> H <b>T</b> L <b>I</b> G <b>A</b> P <b>E</b> A <b>D</b> E <b>R</b> E <b>K</b> A <b>R</b> <b>S</b> H <b>K</b> L <b>E</b> R <b>O</b> N <b>I</b> <b>L</b> <b>Q</b> <b>D</b> L
BMVP	G <b>M</b> L <b>V</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> M <b>V</b> C <b>R</b> <b>I</b> <b>L</b> <b>M</b> H <b>T</b> L <b>I</b> G <b>A</b> P <b>E</b> A <b>D</b> E <b>R</b> E <b>K</b> A <b>R</b> <b>S</b> H <b>K</b> L <b>E</b> R <b>O</b> N <b>I</b> <b>L</b> <b>Q</b> <b>D</b> L
MCE	P <b>E</b> L <b>L</b> <b>V</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> L <b>L</b> C <b>A</b> F <b>L</b> V <b>E</b> K <b>M</b> D <b>W</b> S <b>E</b> A <b>A</b> V <b>A</b> P <b>A</b> Q <b>A</b> <b>R</b> <b>S</b> P <b>G</b> T <b>Y</b> K <b>D</b> <b>P</b> <b>K</b> <b>L</b>
HCE	P <b>E</b> L <b>L</b> <b>V</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> L <b>L</b> C <b>A</b> F <b>L</b> V <b>E</b> K <b>M</b> D <b>W</b> S <b>E</b> A <b>A</b> V <b>A</b> P <b>A</b> Q <b>A</b> <b>R</b> <b>S</b> P <b>G</b> T <b>Y</b> K <b>D</b> <b>P</b> <b>K</b> <b>L</b>
CEL-1	D <b>R</b> V <b>V</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> L <b>L</b> A <b>A</b> <b>L</b> <b>L</b> F <b>V</b> E <b>Y</b> G <b>L</b> D <b>A</b> <b>A</b> G <b>E</b> F <b>A</b> N <b>R</b> <b>K</b> O <b>G</b> T <b>Y</b> K <b>D</b> <b>P</b> <b>K</b> <b>L</b>
F54C8	D <b>A</b> L <b>L</b> <b>G</b> V <b>H</b> C <b>T</b> <b>G</b> <b>N</b> R <b>C</b> <b>G</b> L <b>L</b> C <b>R</b> <b>E</b> L <b>L</b> D <b>R</b> L <b>G</b> <b>S</b> H <b>E</b> R <b>A</b> D <b>A</b> E <b>Q</b> A <b>R</b> <b>S</b> <b>T</b> E <b>K</b> <b>G</b> A <b>L</b> <b>W</b> <b>M</b> <b>A</b>

## B



**FIG. 1. Multiple sequence alignment of PIR1 homologs.** A, the following amino acid sequences were aligned using the Clustal X computer program: PIR1 (GenBank™ accession number AF023917, amino acid residues 30–199), T23G7 (Z68319, 52–225), BVP (L22858, 1–166), BMVP (L33180, 1–166), MCE (AF025653, 5–173), HCE (AB009022, 5–173), CEL-1 (AF003925, 2–171), and F54C8 (Z22178, 7–179). *Solid black boxes* identify residues that are identical within all sequences. *Shaded boxes* enclose residues that are either identical or conserved in six of the eight sequences aligned. T23G7 and F54C8 designate protein sequences predicted for the T23G7.5 and F54C8.4 genes of *C. elegans*, respectively. *BMVP* designates a BVP-like protein encoded by the *B. mori* nuclear polyhedrosis virus. B, diagram illustrating the structural features of PIR1. The *open box* designates the conserved catalytic domain, whereas the *cross-hatched* and *shaded boxes* indicate the position of the arginine-rich (19) and polyproline-rich regions, respectively. The *vertical line* indicates the position of residue 206, which is the C-terminal residue of the catalytic fragment that was expressed as a GST fusion protein.

idues (Table I). The activity of GST-PIR1 toward protein substrates is comparable to that of GST-BVP and the RNA triphosphatase domain of the *Caenorhabditis elegans* capping enzyme, CEL-1 (1–236). Of the four phosphotyrosyl substrates tested, the activity of GST-PIR1 was the highest toward the acidic substrate, poly(Glu<sub>4</sub>:Tyr<sub>1</sub>).

**RNA Phosphatase Activity of PIR1**—To test for RNA 5'-phosphatase activity, GST-PIR1 was incubated with an RNA trinucleotide labeled at the  $\gamma$ -position of its 5'-triphosphate end (pppApCpC). Analysis of the reaction products by TLC showed that the labeled  $\gamma$ -phosphate was released from this RNA trinucleotide substrate (Fig. 2A, lanes 9 and 10), indicating that PIR1 possesses RNA triphosphatase activity. As shown in Fig. 2A, the triphosphatase activity of PIR1 is comparable to that of BVP (3, 4).

In the GST-PIR1 C152S mutant, the essential active site Cys has been replaced by Ser. At a concentration exceeding the wild type protein by 10-fold (Fig. 2A, lanes 11–13), GST-PIR1 C152S had no detectable RNA triphosphatase activity, confirming that this activity can be ascribed to PIR1 and not contaminating *E. coli* proteins. Both the RNA and protein phosphatase activities of PIR1 were inhibited by sodium vanadate, an inhibitor of tyrosine-specific and dual specificity protein phosphatases. The absolute requirement for Cys<sup>152</sup> and its sensitivity to vanadate, suggest that PIR1 hydrolyzes RNA substrates via the mechanism utilized by the tyrosine and dual specificity protein phosphatases (24, 25).

RNA triphosphatases from yeast and viral capping enzymes

TABLE I  
Substrate specificity of PIR1

Substrate	Concentration $\mu\text{M}$	Specific activity $\text{nmol min}^{-1} \text{mg}^{-1}$		
		Cel-1(1-236) <sup>a</sup>	GST-BVP <sup>a</sup>	GST-PIR1
pppACC	3	1000	200	260
ppACC	5	ND <sup>b</sup>	600	3000
<i>p</i> -Nitrophenyl phosphate	50,000	35	200	66
pY-MBP <sup>c</sup>	5	0.006	0.13	0.073
pY-casein	2.5	0.024	0.18	0.008
pY-RCML <sup>d</sup>	5	0.012	0.22	0.011
pY-poly(Glu <sub>4</sub> :Tyr <sub>1</sub> )	5	— <sup>e</sup>	— <sup>e</sup>	0.61
pS/T-MBP <sup>c</sup>	5	ND	ND	ND
pS-casein	5	ND	0.003	ND
pS/T-histone	5	0.001	0.006	ND
pS/T-RCML <sup>d</sup>	5	ND	0.001	— <sup>e</sup>

<sup>a</sup> Values taken from Takagi *et al.* (3).

<sup>b</sup> ND, no detectable phosphate was released (the limit of detection for protein substrate assays was ~0.005–0.1 pmol of phosphate released).

<sup>c</sup> MBP, myelin basic protein.

<sup>d</sup> RCML, reduced, carboxamidomethylated and maleylated lysozyme.

<sup>e</sup> This combination was not tested.

do not contain the HCX<sub>2</sub>R hallmark motif of the PTPs, require Mg<sup>2+</sup> for their activity and possess nucleotide phosphohydrolyase activity (15, 16, 26–30). In contrast, the RNA phosphatase activity of PIR1 does not require Mg<sup>2+</sup> and instead is inhibited about 50% by 100  $\mu\text{M}$  MgCl<sub>2</sub> (data not shown). PIR1 did not exhibit significant ATPase activity (Fig. 2C, lanes 5–7), suggesting that its phosphatase activity is specific to polynucleotides.

Using RNA radiolabeled at the  $\alpha$ -phosphate (pppApCpC), both BVP and PIR1 produced a product that migrates almost as fast as inorganic phosphate on PEI-cellulose plates (Fig. 3A, lanes 3–5 and 6–8). To determine if the observed product contained monophosphate-terminated trimeric RNA (not shown) or free phosphate (lane 1), the reaction products were further digested with RNase T2 prior to analysis by TLC. As shown in Fig. 3B, lanes 3–5 and 6–8, both BVP-treated and PIR1-treated RNA released pAp. We conclude that PIR1, like BVP, leaves a 5'-monophosphate end on RNA.

PIR1 (Fig. 3C, lanes 6–8) and BVP (Fig. 3C, lanes 3–5) converted diphosphate-terminated RNA (ppApCpC) to a species that migrated similarly to 5'-hydroxyl-terminated RNA (ApCpC) produced by treatment with calf intestinal phosphatase (Fig. 3C, lane 1). To distinguish 5'-monophosphate-ended trimer from 5'-hydroxyl-ended trimer, the reaction products were digested with RNase T2. This treatment revealed that pAp was released from BVP-treated and PIR1-treated RNA (Fig. 3D, lanes 3–5 and 6–8), demonstrating that PIR1 also possesses potent RNA diphosphatase activity. Its activity toward diphosphate-terminated RNA was more than an order of magnitude higher than its activity toward triphosphate-terminated RNA, whereas that of BVP is only 3-fold higher (Table I). The presence of a mixture of diphosphate- and monophosphate-terminated RNA in the products from triphosphate-terminated substrate obtained at low concentrations of BVP and PIR1 (Fig. 3, A and B, lanes 3–5 and 6–8), suggests that these two enzymes carry out the triphosphatase and diphosphatase reactions sequentially. The RNA diphosphatase and triphosphatase activities of both BVP and PIR1 are also inhibited to the same extent by magnesium and vanadate (data not shown). Both BVP and PIR1 can be distinguished from the mRNA capping enzyme CEL-1 by their ability to remove the  $\beta$ -phosphate from the diphosphate termini of RNA substrates.

A comparison of the protein and RNA phosphatase activities of PIR1 (Table I) reveals that it dephosphorylates triphosphate-

and diphosphate-terminated RNA with a specific activity that is 2 and 3 orders of magnitude greater than that obtained with a comparable concentration of poly(Glu<sub>4</sub>:Tyr<sub>1</sub>), the best phosphoprotein substrate tested. With artificial protein substrates, the specific activity of GST-PIR1 is about 5 orders of magnitude lower than that of PTP 1B (31), a typical tyrosine-specific phosphatase and at least 10-fold lower than that of the dual specificity protein phosphatase Cdc14 (20). In contrast, PIR1 dephosphorylates the 5'-end of RNA at rates comparable to CEL-1, which acts as a mRNA capping enzyme *in vivo* (2, 3). From these *in vitro* measurements, we conclude that PIR1 is a poor protein phosphatase and a more efficient catalyst with RNA substrates.

**Role of the C-terminal Noncatalytic Domain**—A truncated form of GST-PIR1 containing only the catalytic domain (residues 1–206) was expressed in bacteria and purified in a manner similar to that of the full-length protein. When expressed on a molar basis, the specific activity of GST-PIR1 (1–206) toward *p*-nitrophenyl phosphate and the best protein substrate, poly(Glu<sub>4</sub>:Tyr<sub>1</sub>), was approximately 4- and 3-fold higher than that of the full-length enzyme, respectively (data not shown). The specific activity of GST-PIR1 (1–206) toward both triphosphate- and diphosphate-terminated RNA was 2-fold higher than that of the full-length enzyme (Fig. 2B, lanes 8–10, and data not shown). Like the full-length enzyme, the PIR1 catalytic domain did not dephosphorylate mononucleotides such as ATP (Fig. 2C, lanes 8–10).

Thus, the C terminus of PIR1 is not required for phosphatase activity and its removal does not significantly alter the strong preference for RNA or the level of RNA triphosphatase activity. The effects of C-terminal truncation on phosphatase activity measured with the small RNA and artificial protein substrates employed in these studies may not reflect the properties of PIR1 with physiologic substrates. Moreover, these findings do not eliminate a role for the noncatalytic domain in modulating activity by serving as a binding site for regulatory proteins or other effectors and/or by providing sites for phosphorylation or other post-translational modifications.

## DISCUSSION

We demonstrate herein that PIR1 has both RNA tri- and diphosphatase activities like the viral BVP phosphatase (3, 4). The relatively poor protein phosphatase activity of PIR1 suggests that it is unlikely to be involved in controlling the level of protein phosphorylation. Instead, PIR1 has the potential to play a direct role in RNA metabolism by using RNA as its substrate.

**The RNA Phosphatase Subgroup of PTPs**—At least 10 proteins exhibiting a high degree of sequence similarity to PIR1 (31–46% identity) have been identified. These include the metazoan capping enzymes from *C. elegans* (CEL-1) (2, 10), mice (MCE) (13, 32), and humans (HCE) (13, 33, 34); BVP and its homologs from four other nuclear polyhedrosis viruses<sup>2</sup>; and two predicted proteins (T23G7.5 and F54C8.4) of unknown function from *C. elegans* (Fig. 1A). We propose that these 11 proteins constitute a unique subgroup within the PTP superfamily because they exhibit much greater sequence similarity to one another than to other PTPs and all those that have been studied dephosphorylate the 5'-end of RNA much more efficiently than proteins.

Based on their common function in mRNA processing, simi-

<sup>2</sup> BVP homologs are encoded by the *Bombyx mori*, *Rachiplusia ou*, *Anagrapha falciferia*, and *Orgyia pseudo tsugata* nuclear polyhedrosis viruses. The corresponding nucleotide sequences have the GenBank™ accession numbers L33180, AF068270, U64896, and U75930, respectively.

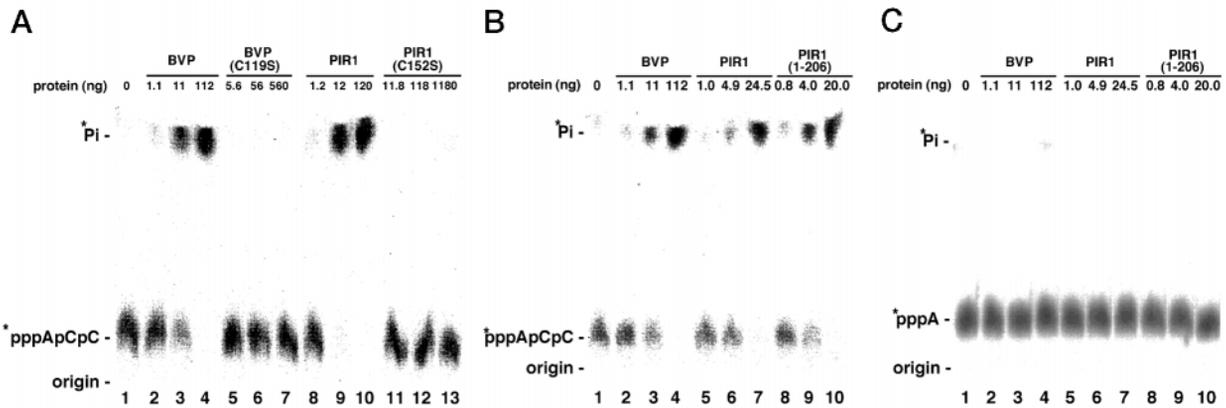


FIG. 2. PIR1 releases  $\gamma$ -phosphate from triphosphate-terminated RNA. GST-PIR1 and RNA substrates were incubated for 10 min at 30 °C in buffer containing 50 mM Tris-HCl, pH 7.9, 5 mM dithiothreitol, 50  $\mu$ g/ml bovine serum albumin. Products were analyzed by TLC on PEI-cellulose plates. The position of the  $^{32}$ P label is designated by the asterisk. The position of [ $^{32}$ P]inorganic phosphate was determined by treating RNA substrates with calf intestine alkaline phosphatase (data not shown). Radioactivity was detected using a Fuji BasX PhosphorImager. A, [ $\gamma$ - $^{32}$ P]ATP-terminated trinucleotide RNA (3  $\mu$ M of termini) was incubated with buffer (lane 1), the indicated amounts of wild-type GST-BVP (lanes 2–4), mutant GST-BVP (C119S) (lanes 5–7), wild-type GST-PIR1 (lanes 8–10), or mutant GST-PIR1 (C152S) (lanes 11–13). B, [ $\gamma$ - $^{32}$ P]ATP-terminated trinucleotide RNA (3  $\mu$ M termini) was incubated with buffer (lane 1), the indicated amounts of GST-BVP (lanes 2–4), GST-PIR1 (lanes 5–7), or GST-PIR1 (1–206) (lanes 8–10). C, nucleotide phosphohydrolysis assay. Reactions in lanes 1–10 were identical to those in B, except [ $\gamma$ - $^{32}$ P]ATP-terminated RNA was replaced with 3  $\mu$ M [ $\gamma$ - $^{32}$ P]ATP.

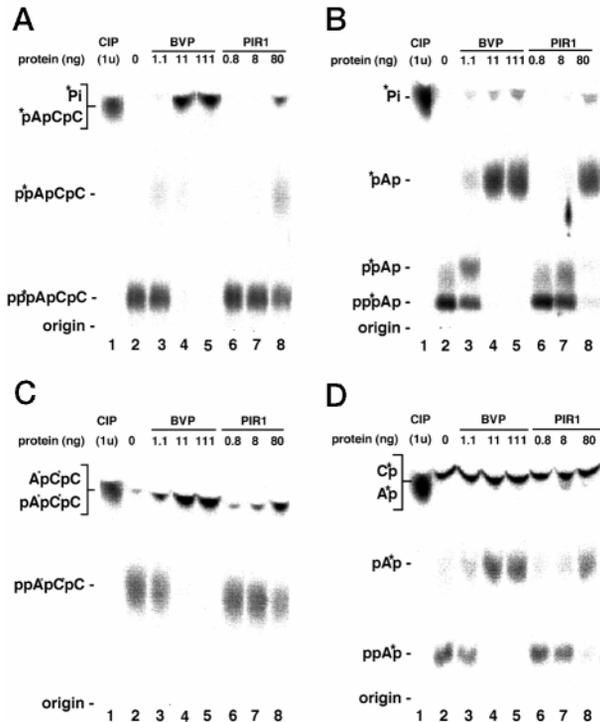


FIG. 3. PIR1 possesses RNA 5'-triphosphatase and diphosphatase activities but leaves a 5'-monophosphate end. RNA triphosphatase assays (panels A and B) and diphosphatase assays (panels C and D) were performed with either [ $\alpha$ - $^{32}$ P]ATP-terminated trinucleotide RNA or 5'-diphosphate-terminated RNA trinucleotide that was labeled with [ $^{32}$ P] at internal phosphates, respectively (label is indicated by asterisk). RNA substrates were incubated for 10 min at 30 °C with the following: 1 unit of calf intestinal phosphatase (CIP) (lane 1), buffer containing 50 mM Tris-HCl, pH 7.9, 5 mM dithiothreitol, 50  $\mu$ g/ml bovine serum albumin (lane 2), the indicated amounts of GST-BVP (lanes 3–5) or GST-PIR1 (lanes 6–8). Products were analyzed by PEI-cellulose TLC directly (panel A and C), or after cleavage with RNase T2 (panel B and D). Positions of the authentic markers, adenosine 3'-monophosphate (Ap) and adenosine 3',5'-diphosphate (pAp) (Sigma), were detected by UV light.

lar domain organization and high degree of sequence similarity within their RNA triphosphatase domains (41% identity between enzymes from worms and man), it is likely that the metazoan capping enzymes (CEL-1, HCE, and MCE) are or-

thologs. On the other hand, PIR1, BVP, the four viral BVP-like proteins and the T23G7 protein from *C. elegans* appear to constitute a distinct group of orthologous proteins that exhibit more similarity to one another (40–46% identity) than to the capping enzymes (31–37% identity). Thus, like PIR1 and BVP, *C. elegans* T23G7 and the BVP-like viral orthologs are expected to have the ability to remove  $\gamma$ - and  $\beta$ -phosphates from RNA and to carry out similar cellular functions. This proposed evolutionary relationship raises the possibility that the nuclear polyhedrosis viruses may have acquired a PIR1-like gene from an insect host through a mechanism involving horizontal gene transfer. The existence of a cellular gene might explain why the BVP gene of *A. californica* nuclear polyhedrosis virus is not essential (18).

**Potential Role for the PIR1 RNA Phosphatase**—It is unlikely that PIR1 functions in mRNA capping because its RNA diphosphatase activity leaves a 5'-monophosphate end on RNA, which is incompatible with the formation of the cap structure found on eukaryotic mRNAs. The possibility exists that a small subset of eukaryotic transcripts might form a mRNA cap structure using a 5'-monophosphate end to which GDP is transferred (35). To date, this unique RNA capping mechanism has only been found in some rhabdoviruses, a class of non-segmented, cytoplasmic RNA viruses (36–39).

Yuan *et al.* (19) found that PIR1 interacts with accessory splicing factors 9G8 and SRp30C in the yeast two-hybrid system and suggested that PIR1 might be localized to nuclear sites where RNAs are undergoing splicing. However, it should be noted that these interactions could not be confirmed by independent methods such as co-immunoprecipitation (19). To our knowledge, RNA intermediates containing 5'-tri- and diphosphate ends do not occur in splicing, making it unlikely that PIR1 plays a role in this process.

Two smaller splice variants (HCE1A and HCE1B) of the bifunctional human capping enzyme HCE, were recently isolated and found to possess RNA 5'-triphosphatase activity but no guanylyltransferase activity (33). Together with PIR1, these HCE variants constitute a group of RNA phosphatases that might carry out RNA processing reactions other than those involved in the capping and splicing of mRNAs. For instance, they might be involved in processing of rRNA or tRNA or in the turnover of RNAs. Elucidation of the precise role of PIR1 in RNA metabolism must await further study.

**Acknowledgments**—We thank Dr. S. Norioka for generously providing RNase T2. We also thank Drs. R. Kuhn and S. Rossie for suggestions and critical comments on this manuscript.

## REFERENCES

- Barford, D., Das, A. K., and Egloff, M.-P. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27**, 133–164
- Takagi, T., Moore, C. R., Diehn, F., and Buratowski, S. (1997) *Cell* **89**, 867–873
- Takagi, T., Taylor, G. S., Kusakabe, T., Charbonneau, H., and Buratowski, S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9808–9812
- Gross, C. H., and Shuman, S. (1998) *J. Virol.* **72**, 7057–7063
- Maehama, T., and Dixon, J. E. (1998) *J. Biol. Chem.* **273**, 13375–13378
- Myers, M. P., Pass, I., Batty, I. H., Vanderkaay, J., Stolarov, J. P., Hemmings, B. A., Wigler, M. H., Downes, C. P., and Tonks, N. K. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13513–13518
- Sheng, Z., and Charbonneau, H. (1993) *J. Biol. Chem.* **268**, 4728–4733
- Kim, D., and Weaver, R. F. (1993) *Virology* **195**, 587–595
- Hakes, D. J., Martell, K. J., Zhao, W.-G., Massung, R. F., Esposito, J. J., and Dixon, J. E. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4017–4021
- Wang, S. P., Deng, L., Ho, C. K., and Shuman, S. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9573–9578
- Shuman, S. (1995) *Prog. Nucleic Acids Res. Mol. Biol.* **50**, 101–129
- Mizumoto, K., and Kaziro, Y. (1987) *Prog. Nucleic Acids Res.* **34**, 1–28
- Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., and Shatkin, A. J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12758–12760
- Wen, Y. X., Yue, Z. Y., and Shatkin, A. J. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12226–12231
- Gross, C. H., and Shuman, S. (1998) *J. Virol.* **72**, 10020–10028
- Jin, J., Dong, W., and Guarino, L. A. (1998) *J. Virol.* **72**, 10011–10019
- Guarino, L. A., Jin, J., and Dong, W. (1998) *J. Virol.* **72**, 10003–10010
- Li, Y., and Miller, L. K. (1995) *J. Virol.* **69**, 4533–4537
- Yuan, Y., Da-Ming, L., and Sun, H. (1998) *J. Biol. Chem.* **273**, 20347–20353
- Taylor, G. S., Liu, Y., Baskerville, C., and Charbonneau, H. (1997) *J. Biol. Chem.* **272**, 24054–24063
- Fisher, C. L., and Pei, G. K. (1997) *BioTechniques* **23**, 570–574
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994) *Science* **266**, 1241–1247
- Guan, K. L., and Dixon, J. E. (1991) *J. Biol. Chem.* **266**, 17026–17030
- Denu, J. M., and Dixon, J. E. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5910–5914
- Shuman, S., Surks, M., Furneaux, H., and Hurwitz, J. (1980) *J. Biol. Chem.* **255**, 11588–11598
- Itoh, N., Mizumoto, K., and Kaziro, Y. (1984) *J. Biol. Chem.* **259**, 13930–13936
- Ho, C. K., Srisakanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) *J. Biol. Chem.* **273**, 9577–9585
- Ho, C. K., Pei, Y., and Shuman, S. (1998) *J. Biol. Chem.* **273**, 34151–34156
- Ho, C. K., Schwer, B., and Shuman, S. (1998) *Mol. Cell. Biol.* **18**, 5189–5198
- Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1988) *J. Biol. Chem.* **263**, 6731–6737
- McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Amgen EST Program, Shuman, S., and Bentley, D. L. (1997) *Genes Dev.* **11**, 3306–3318
- Yamada-Okabe, T., Doi, R., Shimmi, O., Arisawa, M., and Yamada-Okabe, H. (1998) *Nucleic Acids Res.* **26**, 1700–1706
- Tsukamoto, T., Shibagaki, Y., Murakoshi, T., Suzuki, M., Nakamura, A., Gotoh, H., and Mizumoto, K. (1998) *Biochem. Biophys. Res. Commun.* **243**, 101–108
- Shatkin, A. J. (1976) *Cell* **9**, 645–653
- Abraham, G., Rhodes, D. P., and Banerjee, A. K. (1975) *Cell* **5**, 51–58
- Gupta, K. C., and Roy, P. (1980) *J. Virol.* **33**, 292–303
- Shuman, S. (1997) *Virology* **227**, 1–6
- Bisallion, M., and Lemay, G. (1997) *Virology* **236**, 1–7