Virology **277,** 100–110 (2000)
doi:10.1006/viro.2000.0597, available online at http://www.idealibrary.com on IDEAL®

Identification and Characterization of a Shrimp White Spot Syndrome Virus (WSSV) Gene That Encodes a Novel Chimeric Polypeptide of Cellular-Type Thymidine Kinase and Thymidylate Kinase

Meng-Feng Tsai,*[,1] Hon-Tsen Yu,*[,1] Huey-Fen Tzeng,* Jiann-Horng Leu,† Chih-Ming Chou,† Chang-Jen Huang,† Chung-Hsiung Wang,‡ Jung-Yaw Lin,§ Guang-Hsiung Kou,*[,2] and Chu-Fang Lo*[,2]

*Department of Zoology, ‡Department of Entomology, and §Institute of Biochemistry, National Taiwan University, Taipei 106, Taiwan, Republic of China; and †Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan, Republic of China

Received May 30, 2000; returned to author for revision June 27, 2000; accepted August 17, 2000

From previously constructed genomic libraries of a Taiwan WSSV isolate, a putative WSSV *tk-tmk* gene was identified. Uniquely, the open reading frame (ORF) of this gene was predicted to encode a novel chimeric protein of 388 amino acids with significant homology to two proteins: thymidine kinase (TK) and thymidylate kinase (TMK). Northern blot analysis with a WSSV *tk-tmk*-specific riboprobe detected a major transcript of 1.6 kb. When healthy adult *Penaeus monodon* shrimp were inoculated with WSSV, the *tk-tmk* gene transcript was first detected by RT-PCR analysis at 4 h postinfection and transcription levels continued to increase over the first 18 h. The gene's major *in vitro* transcription and translation product, equivalent to the predicted size (43 kDa), is a single chimeric protein that includes both the TK and TMK functional motifs. Evidence for phylogenetic analysis and sequence alignment suggested that the gene may have resulted from the fusion of a cellular-type TK gene and a cellular-type TMK gene. Its unique arrangement may also provide a valuable gene marker for WSSV.

Key Words: Penaeus monodon; white spot syndrome virus; Taiwan WSSV isolate; WSSV *tk-tmk* gene; thymidine kinase; thymidylate kinase.

INTRODUCTION

Shrimp white spot syndrome (WSS) is one of the most serious diseases faced by the shrimp farming industry all over the world (Chou *et al.,* 1995; Flegel, 1997; Lo *et al.,* 1999). White spot syndrome virus (WSSV), the causative agent of WSS, is a large double-stranded DNA virus (Wang *et al.,* 1995), which shows little genetic variation among WSSV isolates from around the world (Lo *et al.,* 1999). Previous studies (e.g., Lo *et al.,* 1997) have provided evidence that WSSV is a new virus, which may belong to the tentative new genus Whispovirus (van Hulten *et al.,* 2000). However, genome analysis is needed to conclusively establish WSSV's taxonomic position. As part of our continuing work to identify and define the genetic structure of WSSV, we report here on an open reading frame (ORF) that encodes a protein with significant homology to both cellular-type thymidine kinase (TK) and cellular-type thymidylate kinase (TMK). TK is a key enzyme of the pyrimidine salvage pathway of nucleotide biosynthesis and is found in all organisms except for most of the viruses and a few lower eukaryotes (yeast, fungi) (Grivell and Jackson, 1968; Kit, 1985; Gentry, 1992). TK catalyzes the ATP-dependent phosphorylation of thymidine to form thymidine monophosphate (TMP), and in the next step, which is catalyzed by TMK, the TMP is phosphorylated to thymidine diphosphate (Kou and Campbell, 1982; Kit, 1985). Because TMK is involved in both the *de novo* and salvage pathways for thymidine triphosphate synthesis, TMK activity is essential for cell proliferation. Indeed, TMK is found in all prokaryotes and eukaryotes and also in some large DNA viruses (Robertson and Whalley, 1988; Smith *et al.,* 1989).

In this study, the WSSV *tk-tmk* gene is characterized and we show that, like herpesvirus (Robertson and Whalley, 1988; Karlin *et al.,* 1994), the functional domains of TK and TMK of WSSV are both contained in a single polypeptide. A comparison of the deduced WSSV amino acid sequences with other known thymidine kinase and thymidylate kinase sequences suggests that, unlike herpesvirus, which appears to have originated its own TK, the WSSV *tk-tmk* gene probably resulted from the fusion of two preexisting genes captured from a WSSV host and/or from a coinfecting virus in a WSSV host species. In this respect, WSSV would thus resemble some other DNA viruses, for instance, poxviruses (Boyle *et al.,* 1987; Smith *et al.,* 1989) and probably African Swine Fever Virus (Blasco *et al.,* 1990; Yanez *et al.,* 1993) whose separate TK and TMK protein genes were also originally captured from their hosts.

[1] These authors contributed equally to this work.

[2] To whom correspondence and reprint requests should be addressed at National Taiwan University, Department of Zoology, Taipei 106, Taiwan, ROC. Fax: 886-2-23638179. E-mail: ghkou@ccms.ntu.edu.tw or gracelow@ccms.ntu.edu.tw.

0042-6822/00 $35.00

RESULTS

Location of WSSV *tk-tmk*

The WSSV genomic DNA fragment used in this study (pms94, where "pm" indicates *Penaeus monodon* and "s" indicates a *Sal*I fragment) was taken from one of the WSSV plasmid libraries constructed by Wang *et al.* (1995). These libraries were based on the WSSV Taiwan isolate (Lo *et al.,* 1999), which was originally isolated from a batch of WSSV-infected *P. monodon* collected in Taiwan in 1994 (Wang *et al.,* 1995). Random sequencing and alignment with GenBank/EMBL, SWISSPORT, and PIR databases showed that pms94 contained an ORF with a high homology to TK and TMK. The ORF was located between positions 1624–2790 of the 8713-bp pms94 fragment (Fig. 1a). The WSSV *tk-tmk* ORF consists of 1167 nt with the potential to encode a polypeptide of 388 amino acids with a theoretical size of 43 kDa and a pI of 6.28. The sequence surrounding the putative translation initiation codon (ACCATGG) complies with the Kozak rule (Kozak, 1989, 1997) of optimal context for an efficient translation initiation in eukaryotic cells. As Fig. 1b shows, two other alternative start codons were also identified, one at −30 to −28 and the other at 34 to 36, although neither of these is embedded in a sequence that conforms to the Kozak rule. The predicted promoter spans from −66 to −17 where +1 is the putative translation initiation. The putative TATA box is located at −55 to −50 and the sequence TTTTTAT, which is identical to one of the early transcription termination signals (T_5NT) of the vaccinia virus (Yuen and Moss, 1987), appears at nt 1484–1490. In addition, a polyadenylation signal (AATAAA) is located at the second nucleotide of the codon for the last amino acid residue (nt 1163–1168). The WSSV *tk-tmk* gene sequence was submitted to GenBank (accession no. AF272669).

Major transcript of putative WSSV *tk-tmk* in WSSV-infected shrimp

For the transcriptional analysis, subadult *P. monodon* were artificially infected with WSSV by injection (Tsai *et al.,* 1999). A Northern blot of total RNA extracted from these WSSV-infected *P. monodon* at 0, 6, 18, and 36 h postinfection (p.i.) was hybridized with a WSSV *tk-tmk*-specific riboprobe generated by *in vitro* transcription (Sambrook *et al.,* 1989). No WSSV *tk-tmk*-specific transcripts were found at 0 and 6 h p.i. At 18 and 36 p.i., there was one major transcript of approximately 1.6 kb and a minor transcript of approximately 2.4 kb (Fig. 2).

Temporal analysis of WSSV *tk-tmk* transcription by RT-PCR

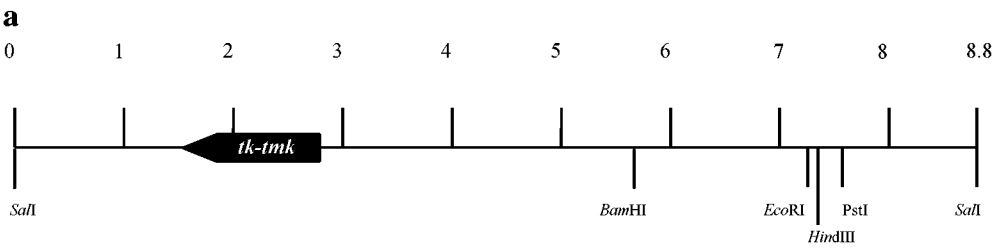
RT-PCR analysis was used to detect the *tk-tmk*-specific transcript in DNase-treated total RNAs extracted from subadult *P. monodon* specimens both before (at 0 h) and after (at 2, 4, 6, 8, 18, 24, 36, and 60 h) artificial infection with WSSV. The *tk-tmk* transcript was first detected at 4 h p.i. and continued to be found through to 60 h p.i. (Fig. 3a). Unlike the *rr1* and *rr2* genes (Figs. 3b and 3c), the intensity of the *tk-tmk* product band increased over time, which suggests that the amount of *tk-tmk* transcript increased as infection advanced. The *tk-tmk* transcript was expressed maximally at 18 and 24 h p.i. The amount of *tk-tmk* transcript was much higher than that of *rr1* and *rr2*. When the DNase-treated total RNAs were treated with RNase and then subjected to RT-PCR with WSSV-specific primer, no RT-PCR products were yielded (data not shown). This quality control confirmed that no viral genomic DNA was left in the prepared RNA for WSSV *tk-tmk* transcription analysis.

In vitro transcription and translation

For *in vitro* transcription and translation, the *tk-tmk* gene (nt −30 to 1158) from pms94 was inserted into pET20b(+) (Novagen, Madison, WI) to yield the expression vector pETTK. The coupled transcription/translation reactions used the TNT Quick Coupled Transcription/Translation System (Promega, Madison, WI) with the rabbit reticulocyte lysate solution for eukaryotic *in vitro* translation. Two potential translation initiation codons were found in this WSSV *tk-tmk* expression vector: ATG at nt −28 to −30 (M −10) and ATG at nt 1 to 3 (M1). Using the TNT Quick system, a major translation product from expression vector pETTK was 43 kDa (Fig. 4), which is the estimated size of the WSSV *tk-tmk* gene product.

Amino acid sequence alignment of WSSV TK-TMK

A multiple alignment (data not shown) of published TK and TMK proteins with the deduced amino acid sequence of the presumed coding region of the WSSV *tk-tmk* gene (Fig. 1) suggested that WSSV TK can be divided into two parts (Fig. 5a). The region at residues 1–182 shows homology to TKs from selected species (similarity for other viruses, vertebrates, and bacteria ranges from 28 to ~58%; Table 1), and the region at residues 191–388 likewise shows homology to TMK (similarity, 34 to ~56%; Table 2). Based on data from the PROSITE database, within these two regions, six domains (I–VI) have high homology to TK and TMK consensus sequences with assigned functions (Fig. 5b). Domains I–IV correspond to the cellular-type TK consensus motifs and domains V and VI correspond to TMK consensus motifs in both order and spacing. The -GPMFAGKS- sequence of domain I conforms to the highly conserved sequence -[AG]-X(4)-G-K-[ST]- for the flexible phosphate-binding loop of ATP/GTP binding motif A. The sequence (-KHSLETRY-) of domain II fits the consensus sequence -[RK]-X(2,3)-[DE]-X(2,3)-Y- for the tyrosine kinase phosphorylation site. Domain III (-VILIDE-) corresponds to nucleotide-binding motif B



b

```
GAT TAC AAA TTT TGC TCT CAA GAG AGG ACC TGC AAC TGG CCA GAA CTT TGT ATC TGT TAA ACT GGA ACA TGA TGG ATC TAA AGC AGA CCT   -91
GTA CAA CGT CAC GTG CTT CTC CAA GCA GCG TGG AGT ATA AAA GGC GGG ACG CAT CAG CAA ATG CAA CTC ATT CTT TCT CAT CAT CTA ACC    -1
ATG GCT GGT CGT GTA GAG CTC GTC ACT GGA CCC ATG TTT GCG GGC AAG TCT ACC TAC CTG AAA AAC ATA TAC CAA CAA GAA AAT GGA GGC    90
 M  A  G  R  V  E  L  V  T  G  P  M  F  A  G  K  S  T  Y  L  K  N  I  Y  Q  Q  E  N  G  G
AAT AAA CAT TGC CTG TTT GTC AAA CAC TCC CTA GAA ACT AGG TAC GGT TGT GGA ACT GGA ACA ATA GTC ACT CAT GCC GGA GAA GTG ATT   180
 N  K  H  C  L  F  V  K  H  S  L  E  T  R  Y  G  C  G  T  G  T  I  V  T  H  A  G  E  V  I
GAA GGT TGT ACT ACA GTT TCT TCT ATC AAG GAA CTA ATC AGT GTG TTA CCA GAA GTT GTG GAT GTG ATT CTC ATT GAC GAA GGG CAA TTC   270
 E  G  C  T  T  V  S  S  I  K  E  L  I  S  V  L  P  E  V  V  D  V  I  L  I  D  E  G  Q  F
TTC ACG GAT TTG GTG CTA GTC AAT AGA CTG GCT GAC AAG GGG AAA AGG ATT GTG ATT GCA GCA CTT GAT GGA ACT TCT GAC CAG CAA ATG   360
 F  T  D  L  V  L  V  N  R  L  A  D  K  G  K  R  I  V  I  A  A  L  D  G  T  S  D  Q  Q  M
TTC AGT CCT ATT CAT AAG CTA TTG CCT TAT ACA AAT TCC ATT GTT AAG CTA GCA TCT AAA TGT ATG ATT TGT AAA ATT GAT ACC AAA GAA   450
 F  S  P  I  H  K  L  L  P  Y  T  N  S  I  V  K  L  A  S  K  C  M  I  C  K  I  D  T  K  E
GCT CCT TTT ACT GTA AGG TTT GGT AAT GAC AAT GAT AAT AAT GTT ATA TGT GTA GGA GGA GCT GAA ATG TAC GCT GCT GCC TGC CGG GAC   540
 A  P  F  T  V  R  F  G  N  D  N  D  N  N  V  I  C  V  G  G  A  E  M  Y  A  A  A  C  R  D
TGT TAC AAA AAA ATT AAC AAG AAA AAG AAC AAG GGG AAA CTT GTT GTA CTT GAA GGA GGT GAC AGG TGC GGT AAG AGT ACC CAA GCC AAA   630
 C  Y  K  K  I  N  K  K  K  N  K  G  K  L  V  V  L  E  G  G  D  R  C  G  K  S  T  Q  A  K
CTC TTG TTG ACC AAT AAA AAC TCG CCT CTT TAT GGA GGA GAA TAC ATG TGC TTT CCC GAC AGG AGC AGC CAT ACG GGT AAA CTC ATC AAT   720
 L  L  L  T  N  K  N  S  P  L  Y  G  G  E  Y  M  C  F  P  D  R  S  S  H  T  G  K  L  I  N
GAT TAT TTA ACT AAG AAA ATT GAA CTA GAT GAT CAT GCA GCT CAC TTG TTA TTT TCT GCA AAT AGA TGG GAA GTT TGT AGT AAA ATT AAG   810
 D  Y  L  T  K  K  I  E  L  D  D  H  A  A  H  L  L  F  S  A  N  R  W  E  V  C  S  K  I  K
CAG TTG TTA GAC GAT GGA ATC CAT GTT GTG ATG GAT AGA TAT TAC TAC TCG GGG ATT GTT TTC TCT TTA GCT AGA GGA GTG GAT ACC GTT   900
 Q  L  L  D  D  G  I  H  V  V  M  D  R  Y  Y  Y  S  G  I  V  F  S  L  A  R  G  V  D  T  V
GAG TGG TGC TCT GCT AGC GAT GAG GGA CTT CCT CAG CCC GAT CTT GTA TTG TTG ATG CTT TTA GAT GTT GAA AAG TGT TCA AAT AGG GAT   990
 E  W  C  S  A  S  D  E  G  L  P  Q  P  D  L  V  L  L  M  L  L  D  V  E  K  C  S  N  R  D
ACT TTT GGT GTC GAA AGA TTT GAG ACA AAT TCC ATT CAA GAA CGT GCT AGA GCC CTA TTC CTA GAC CTC GCA AAT AAG GAC GAA AAG AAT  1080
 T  F  G  V  E  R  F  E  T  N  S  I  Q  E  R  A  R  A  L  F  L  D  L  A  N  K  D  E  K  N
GTA TGG ATT AAG GTA GAC GCT CGC GGC ACC ATT GAG GAG GTG CAA ACT AAA ATT ATA AAT ATT GTA TAT AAT ATT GTT GAA GAA TAA AGA  1170
 V  W  I  K  V  D  A  R  G  T  I  E  E  V  Q  T  K  I  I  N  I  V  Y  N  I  V  E  E  *
ATT GTA AAC TGT TTC CAT GAT GTG TTT CTT TCC TAG TGA TTT TCA CAT GTT ACC TAG AAA GAC TTT GCC CGA CAC TGA AAA TGA TTA TTT  1260
TGT CTT GGA CGA GTC TCT TCT GGA GAA GGT GTA CTA TGA TAA CAA CAA TGA ACT GAT TGT AAG AGT TGG TGG GAT TTA TAT GCA GAT ATG  1350
CAA GTC AAA ATA CAT CTT CCA TCA CGA TGA TCC AGA GAG GTT CTT TTA TAG TGT GTT GGA GGA TTA TCA CCC CAT CAA AGA GAT TGT TGA  1440
ACG ACT AGC AGA AGA GGA TGG GGT ATT TTT AGG ACC GTG GGA GTT TTT ATC GCG CAA ACA AGT GAA CCT CCA ACA CGG GTG CTA CAA AGC  1530
TCT TTT GTC ATT GCC AGA GGA CAA ATA TTG TAA CCT ATT ATT ACC CCA GCA AAT GAA AAC CAA CCT GGA AAA AAT GGA AGA AAT ACA GCG  1620
```

FIG. 1. WSSV *tk-tmk*. (a) The 8715-nucleotide WSSV *Sal*I fragment (pms94) showing the location (nt 1624 to 2790) of the putative WSSV *tk-tmk* gene and (b) the nucleotide sequence of the putative WSSV *tk-tmk* gene and the deduced WSSV TK protein sequence. The predicted TATA box, T_5AT sequence, and polyadenylation signal (AATAAA) are underlined. The two alternative start codons are boxed.

and is located about 70 amino acids downstream from the phosphate-binding loop. Domain IV (residues 169–182) sequence -GGAEMYAAACRDCY- matches the thymidine kinase cellular-type signature consensus sequence -[GA]-X(1,2)-[DE]-X-Y-X-[STAP]-X-C-[NKR]-X-[CH]-[LIVMFYWH]-. Some of the basic amino acids such as R and K that could play a role in phosphate binding and transfer are also found in domain IV. In the TMK region, domain V (-VGGDRCGKS-) corresponds to the ATP/GTP binding motif A, and domain VI (-VMDRYYYSGIVFS-) conforms to the consensus sequence for the thymidylate kinase signature, -[LIV]-[LIVMG-

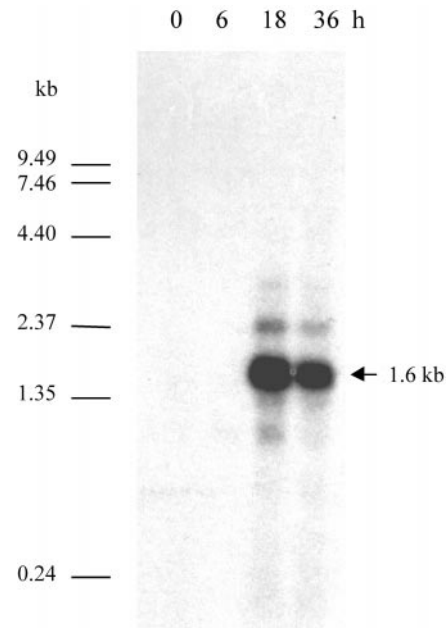


FIG. 2. Northern blot hybridization of total RNA isolated from WSSV-infected *P. monodon* with *tk-tmk*-specific riboprobes. The size standards are indicated using RNA marker (Promega). Lane headings show h p.i.

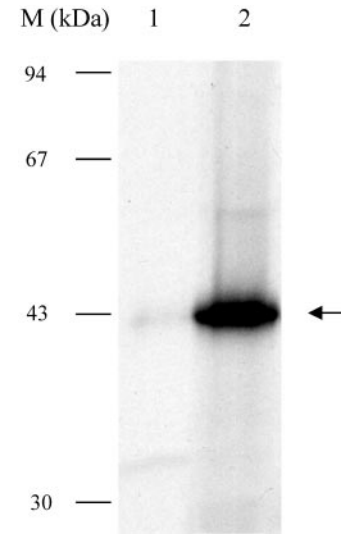


FIG. 4. *In vitro* transcription and translation of WSSV *tk-tmk* gene from pET20b(+) (lane 1, as control) and pETTK (lane 2).

STC]-[DT]-[RH]-[FYHCS]-X(2)-S-[GSTNP]-X-[AV]-[FY]-[STANQ]-. Sequence -MDRYYYS-, which is important for substrate binding, is also located in domain VI.

The pairwise identity and similarity (BLOSUM 35) of WSSV TK-TMK to TKs and TMKs of other viruses, vertebrates, yeast, and bacteria are shown in Tables 1 and 2. Conservation of the amino acid sequences in domains I to VI is high, and is particularly striking for domains I, III, V, and VI, which contain the nucleotide-binding motifs (Tables 1 and 2; Fig. 5). The highest similarities for domains I, III, V, and VI are 100, 83, 75, and 84%, respectively, whereas the intervening regions A to G show a much greater sequence variability (Tables 1 and 2). The highest relatedness of the overall amino acid sequence of the WSSV TK and TMK regions is to Canarypox virus and human, respectively. The WSSV TMK region also has a high degree of relatedness to Variola virus and Vaccinia virus. The E region, which forms the junction between the TK and TMK regions has very low similarity to the corresponding (i.e., C-terminus) region of the cellular TKs and even to viral TKs of similar length (VAR, CPV, etc.), but it shows a relatively high degree of similarity to the corresponding (N-terminus) region of the selected TMKs. When multiple alignment was used to divide the E region into the E1 (residues 183–190) and E2 (residues 190–198) subregions, relative to the entire E region, E2 was found to have higher similarity to the N-terminus region of TMKs. The E2 region also has a surprisingly high degree of relatedness to *E. coli* and Vaccinia virus. Like the E region, the E1 region still has very low homology to TKs and TMKs, but on the other hand, E1 is a K N-rich region, and in this it resembles certain groups of poxvirus TKs.

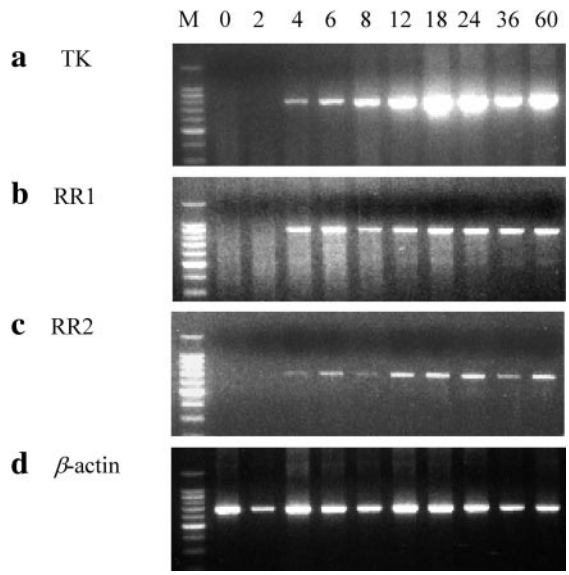


FIG. 3. RT-PCR (a) with *tk-tmk*-specific primers, (b) with *rr1*-specific primers, (c) with *rr2*-specific primers, and (d) with β-actin degenerate primers (internal control). M represents a DNA marker (100-bp DNA Ladder; Promega). Lane headings show h p.i.

Phylogenetic analyses

Only neighbor-joining (NJ) trees (Saitou and Nei, 1987) are shown here (i.e., parsimony trees are not shown) because both types of trees yielded similar results; moreover, the NJ trees revealed finer structures for the phylogenetic relationships.

For the phylogenetic analysis of the TK region (residues 1–182) of the WSSV TK-TMK protein, 37 TK amino

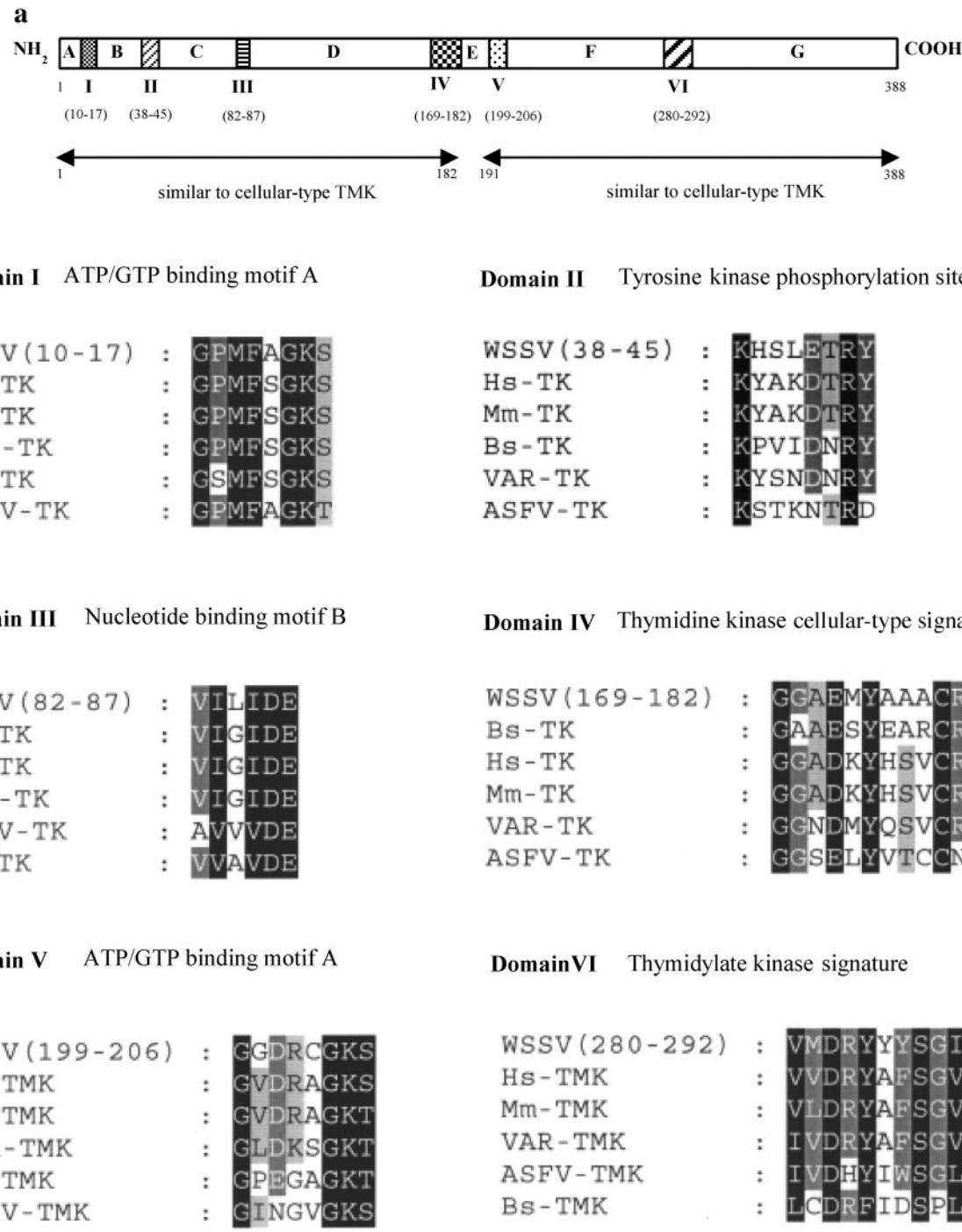


FIG. 5. (a) Schematic representation of the WSSV TK-TMK protein and (b) amino acid sequences for six domains of the WSSV TK-TMK protein, aligned with corresponding regions of selected TK and TMK proteins. The labels I to VI indicate six highly conserved regions with assigned functions and A to G indicate the intervening regions. Amino acid residue numbers for these regions are shown in parentheses.

acid sequences were included in the NJ tree (21 viruses: T4, CIV, CamPV, VAR, MPV, CoV, EV, CPV, RaPV, ShPV, SwPV, YMTV, MV, RFV, CaPV, FWPV, ASFV, WSSV, AmEPV, CfEPV, and CbEPV; eight bacteria: Sgc, Ec, Enc, Tm, Rsp, Dr, Bs, and Mp; four vertebrates: Hs, Cg, Mm, and Gg; two plants: At and Os; one slime mold, Dd; and one nematode, Ce; see Fig. 6). Strikingly, the viral TK genes are more similar to those of their hosts than to one another. For instance, the vertebrate poxviruses form a cluster with the vertebrates and yet the insect poxviruses form a cluster of their own instead of with the vertebrate poxviruses; likewise, T4 is clustered with *E coli* and two other bacteria (Sgc and Enc); otherwise, this gene tree largely reflects the known phylogenetic relationships of the organisms. Thus, the bacterial genes form two clusters, one with five bacteria (Tm, Rsp, Dr, Bs, and Mp) and

TABLE 1

Pairwise Comparison of the Amino Acid Sequence (Identity/Similarity; BLOSUM 35) of the TK Region of WSSV TK-TMK Protein (Residues 1–182, Consisting of Consensus Regions I–IV and Intervening Regions A–D), the Junction Region E (Residues 183–198), and Its Subregions E1 (Residues 183–190) and E2 (Residues 190–198) with Other TK Proteins.

		Consensus region				Intervening region				Junction region		
Source	TK (1–182)	I	II	III	IV	A	B	C	D	E	E1	E2
CaPV	36/58	75/100	50/75	83/83	64/71	20/40	14/47	21/37	36/59	6/6	0/12	0/0
EV	34/55	87/100	50/62	83/83	57/71	22/55	4/38	13/41	36/57	0/6	12/25	0/20
MV	34/55	100/100	25/37	83/83	50/71	16/41	14/42	22/47	34/58	0/12	0/0	0/0
SwPV	33/57	87/100	50/62	50/83	57/64	20/40	4/42	22/44	32/60	0/0	0/12	0/0
MPV	33/56	87/100	50/62	83/83	57/71	20/50	4/38	13/41	35/58	0/6	12/25	0/20
CoV	33/55	87/100	50/62	83/83	57/71	22/55	4/38	13/41	35/57	0/6	12/25	0/20
VAR	33/55	87/100	50/62	83/83	57/71	20/50	4/38	13/41	35/57	0/6	0/0	0/11
CPV	32/55	87/100	50/62	66/83	57/71	22/55	4/38	11/38	35/57	0/6	12/25	0/20
ShPV	32/54	87/100	37/50	66/83	57/64	20/40	9/38	11/47	32/54	12/18	0/0	22/33
FWPV	29/53	75/100	50/62	83/83	50/64	20/60	9/47	18/39	25/53	0/0	12/12	0/0
CfEPV	28/52	62/100	37/62	66/83	50/71	22/55	9/47	10/33	30/57	0/12	12/25	0/22
CbEPV	28/52	62/100	37/62	66/83	50/71	22/55	9/47	10/33	30/57	0/12	12/25	0/22
AmEPV	27/52	75/100	25/62	66/83	50/64	22/44	9/52	12/30	27/55	0/0	12/12	0/0
Gg	27/43	87/100	50/75	83/83	57/71	4/24	14/42	8/27	36/57	2/11	5/8	2/8
Dr	27/42	87/100	37/75	83/83	35/35	20/40	0/33	17/32	32/49	0/6	0/0	7/14
Bs	26/45	75/87	37/62	50/83	50/50	21/42	0/33	15/42	28/48	12/31	22/22	8/16
Mm	25/43	87/100	50/75	83/83	57/71	4/24	9/42	8/36	36/57	0/6	4/4	0/6
Hs	25/42	87/100	50/75	83/83	57/71	4/24	9/42	11/36	36/56	0/6	4/4	0/6
ASFV	25/42	87/100	37/50	33/83	42/64	0/5	19/33	12/32	26/45	5/15	15/30	0/20
At	23/40	87/100	62/75	83/83	50/64	5/13	14/33	12/29	27/55	10/13	3/6	6/16
T4	18/41	62/62	37/75	33/83	35/57	0/33	14/38	12/25	19/40	6/6	25/62	11/11
Ec	17/39	62/75	12/50	50/83	50/57	0/33	14/33	17/30	15/40	4/14	0/5	5/15
CIV	8/28	62/75	25/25	0/16	14/21	0/22	0/19	5/23	5/24	5/11	22/22	10/20

Note. Refer to Fig. 5 for the location of each region.

the other with three (Enc, Ec, and Sgc); the vertebrates form a tight cluster (Hs, Cg, Mm, and Gg); and the two plants form a cluster (At and Os). Among the vertebrate poxviruses, the bird poxviruses form a robust cluster (bootstrap support over 90%), as do the mammal poxviruses (except in sheep and pigs). Two rabbit viruses (MV

TABLE 2

Pairwise Comparison of the Amino Acid Sequence (Identity/Similarity; BLOSUM 35) of the TMK Region of WSSV TK-TMK Protein (Residues 191–388, Consisting of Consensus Regions V–VI and Intervening Regions F–G), the Junction Region E (Residues 183–198), and Its Subregions E1 (Residues 183–190) and E2 (Residues 191–198) with Other TMK Proteins.

		Consensus region		Intervening region		Junction region		
Source	TMK (191–388)	V	VI	F	G	E	E1	E2
Hs	38/56	75/75	53/84	42/63	30/46	31/43	7/14	41/58
Mj	38/51	50/62	53/69	31/50	23/45	12/37	8/16	20/60
Ce	37/54	75/75	53/76	41/54	31/50	25/50	6/12	28/50
Sp	36/56	75/75	61/84	41/58	26/50	37/62	6/20	38/69
Sc	36/56	62/75	69/84	43/60	26/47	25/56	15/23	36/81
Mm	35/52	62/75	53/84	47/63	23/40	31/43	7/14	41/58
At	33/51	75/75	61/84	43/60	22/43	6/6	0/0	11/11
VAR	30/54	50/75	38/84	42/63	18/41	25/43	8/16	40/70
VV	30/53	50/75	38/84	42/63	17/40	25/43	8/16	40/70
Tm	24/45	62/62	30/69	27/43	21/41	6/31	12/25	11/55
Ec	18/37	37/62	23/38	15/31	21/38	18/43	18/27	33/77
ASFV	17/34	50/62	30/76	30/46	7/21	18/43	9/18	33/77
Bs	13/39	37/62	23/53	17/35	12/38	12/37	9/18	22/66

Note. Refer to Fig. 5 for the location of each region.

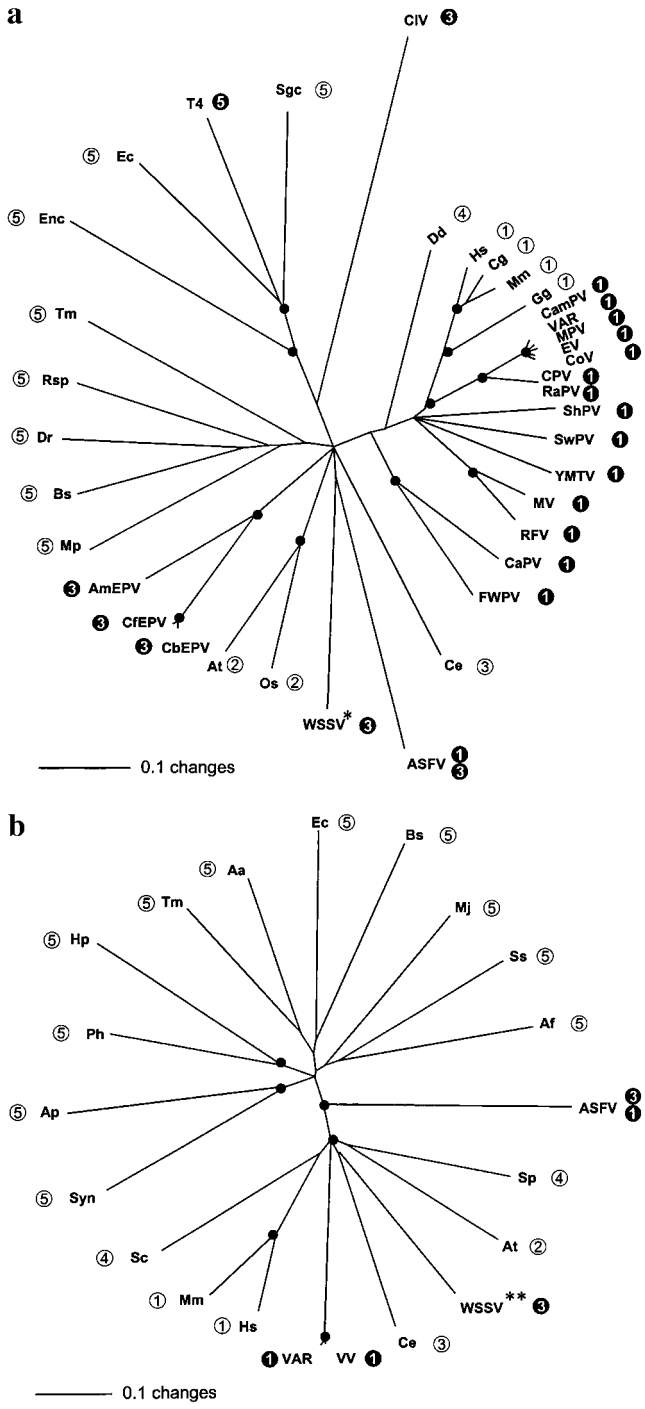


FIG. 6. Unrooted neighbor-joining phylogenetic tree of the TK (a) and TMK (b) genes from different organisms, based on amino acid sequences. Each node is supported by over 50% of the bootstrap replicates and dark dots identify the nodes that are supported by over 90% of the bootstrap replication. *, comparison based only on the TK region of WSSV, i.e., amino acid residues 1–182 of the WSSV TK-TMK polypeptide. **, comparison based only on the TMK region of WSSV, i.e., amino acid residues 191–388 of the WSSV TK-TMK polypeptide. Viruses are indicated by the black circles; the numbers correspond to the type of organism (white circles) or the type of host(s) (black circles). 1, vertebrates; 2, plants; 3, invertebrates; 4, fungi; 5, bacteria and archaea.

and RFV) also form a robust cluster supported by over 90% bootstrap replicates. WSSV forms a cluster with ASFV.

In the phylogenetic tree of the TMK region (residues 191–388) of WSSV TK-TMK protein and other TMKs, two major clusters are well supported by 90% of bootstrap replications. One comprises six eukaryotes (two yeasts, one plant, one nematode, and two mammals) and four viruses (two poxviruses, WSSV, and ASFV). The other includes six bacteria and five archaea (Archaebacteria). Within the eukaryote–virus cluster, five major clades were identified: (1) human and mouse (also joined by baker's yeast); (2) another yeast (fission yeast) with a plant; (3) two closely related poxviruses; (4) WSSV and Ce; and (5) ASFV, which is a clade of itself. Finally, although their relationships are unresolved, the first four clades then join to form one cluster that is distinct from ASFV. Three unresolved clades are identified within the bacteria–archaea cluster. The first includes three archaea (Mj, Ss, and Af), the second includes four bacteria (Bs, Ec, Aa, and Tm), and the third includes two subclades, each containing one bacterium and one archaea (Hp and Ph; and Ap and Syn).

DISCUSSION

In addition to the ATG for M1, WSSV *tk-tmk* has two other potential in-frame ATGs located at nt −28 to −30 (M-10) and nt 34 to 36 (M12), respectively. Kozak (1995) found that when a second ATG follows close upon the first, the first ATG is the most likely translation start site. Context, however, is also important: in eukaryotic cells, ATGs that have a G at position +4, an A at −3, and do not have a T at position +5 are favored (Kozak, 1997). Because the ATG for M1 matches these criteria, whereas the ATGs for M-10 and M12 do not, the latter two are therefore not likely to serve as an efficient translation start site. Empirically, in the WSSV *tk-tmk* expression vector pETTK for *in vitro* transcription and translation assay, either the ATG at nt −28 to −30 (M−10) or the ATG at nt 1 to 3 (M1) (i.e., the first and second ATGs, respectively) may serve as the translation initiation codon. From Fig. 4, it is hard to resolve the M-10 and M1 *in vitro* translation products. However, the asymmetric shape of the major band suggested that it may have merged from a thin (i.e., less intense) band at the top (higher molecular weight) and a thick band at the bottom. If so, this would suggest that the ATGs at both M-10 and M1 served as translation initiation codons in this assay system, and that the translation started from M1 was more efficient than that from M-10.

Two consensus sequences could be involved in the transcriptional termination of the WSSV *tk-tmk* gene: AATAAA, which for most genes signals the addition of a poly A tail at a position 12 to 13 nucleotides downstream; and TTTTTAT, which conforms to the consensus se-

quence (TTTTTNT). T_5NT signals transcriptional termination of vaccinia virus early genes and results in polyadenylation downstream of the T_5NT signal. Since the result of 3′ RACE only indicated that a poly A tail was added to the WSSV *tk-tmk* transcript at the position downstream of the AATAAA signal, the WSSV *tk-tmk* gene may not use the vaccinia early gene termination signal at all. If, in fact, the T_5NT signal is used, the absence of this species of transcript (i.e., a species with a larger 3′UTR) in the 3′ RACE suggests that it can be expressed only at low levels and can probably be expressed only during the early transcription phase.

The Northern blot analysis (Fig. 2) provides evidence that the putative WSSV *tk-tmk* gene is expressed during viral infection. However, if transcription of the putative WSSV *tk-tmk* gene begins several nucleotides downstream of the predicted TATA box location, and terminates downstream of the poly A signal, then even allowing for the addition of a 3′ poly A tail of 200 nucleotides, the major transcript (1.6 kb; Fig. 2) is still larger than the predicted size of ~1.4 kb. This discrepancy could be accounted for if the early transcription termination signal TTTTTAT was used and if the poly A tail was added downstream of this T_5AT sequence; unfortunately, the 3′ RACE results indicate that, in fact, the transcription ends several bases downstream of the poly A signal. An alternative explanation is that the promoter usage is different from the predicted one. A promoter about 200 nucleotides upstream of the TATA box at −55 to −50 (Fig. 1) would increase the length of the 5′-noncoding leader sequences of the transcript such that it would be consistent with the results of Northern blot analysis (Fig. 2). A primer extension experiment would be required to test this possibility.

Temporal analysis of the WSSV *tk-tmk* transcript by RT-PCR showed that the *tk-tmk* expression pattern is different from that of WSSV *rr1* and *rr2* (Fig. 3). This suggests that different regulation mechanisms may be involved in *rr* and *tk-tmk* gene expression. This is further supported by the fact that, except for a TATA box in *rr* and *tk-tmk,* no other shared consensus sequence was observed in the predicted promoter regions or the transcriptional initiation site for *rr* and *tk-tmk.* Figure 3 also shows that the *tk-tmk* transcripts accumulate in the course of infection, whereas the *rr1* and *rr2* transcripts remain at a low level of expression throughout the observed infection period. This may be because the *tk-tmk* transcript is much more stable than that of either *rr1* or *rr2.* Alternatively, transcription of WSSV *tk-tmk* may be driven by different promoters at the early and late transcription phases. If so, the late promoter is evidently stronger than the early promoter.

The two gene trees show that some viral genes are more closely related to their hosts than to their own taxonomic sister groups. The T4 TK gene clusters with that of its host *E. coli.* The viruses that infect vertebrates form a clade with their hosts, so that mammalian viral TK genes (CamPV, VAR, MPV, EV, CoV, CPV, RaPV, ShPV, SwPV, YMTV, MV, and RFV) are closely related to mammalian *tk* genes, whereas the insect poxvirus TK genes (AmEPV, CfEPV, cbEPV) form their own group and are phylogenetically distant from the poxvirus that infect vertebrates. These results imply intimate interactions between the viral and host genetic components. This inference is also supported by the fact that different viral genes (such as MV and RFV, which both infect rabbits) tend to cluster together if they infect the same host. In the TK tree, WSSV and ASFV form a cluster. Although ASFV infects pigs, it is also capable of infecting arthropods (Kleiboeker *et al.,* 1999), and it may be that these two viruses form a clade, albeit deep-branched, because sometime in the evolutionary past they may both have infected hosts (arthropods) that had a close phylogenetic relationship. The current situation might thus have arisen from viral and host genetic interactions in the remote evolutionary past.

To date, WSSV is the only organism known to have both cellular-type TK and TMK functional domains on a single coding gene. We propose that the *tk* and *tmk* regions of the WSSV *tk-tmk* gene might originally have been captured either directly or indirectly from a WSSV host. Studies suggest that viruses can recruit both *tk* (Boyle *et al.,* 1987; Blasco *et al.,* 1990) and *tmk* genes (Smith *et al.,* 1989; Yanez *et al.,* 1993) by capturing cellular genes (presumably in the form of cDNA); alternatively, the *tk* and *tmk* genes could have been captured from a coinfecting virus such as the poxvirus, which is already known to encompass a number of genes with closely related cellular counterparts (Esposito and Knight, 1984; Boyle *et al.,* 1987; Koonin and Senkevich, 1993). As shown in Tables 1 and 2, TK and TMK proteins of certain poxviruses share high homology with those of WSSV. In either case, incorporation of the cellular genes, which might have been captured simultaneously or at different times, would have been achieved via recombination.

MATERIALS AND METHODS

Virus, plasmid clone, and sequence analysis

From WSSV plasmid libraries constructed by Wang *et al.* (1995) and based on the WSSV Taiwan isolate (Wang *et al.,* 1995; Lo *et al.,* 1999), a plasmid clone containing the pms94 fragment was first sequenced by using universal M13 forward and reverse primers. The internal sequences of the cloned fragments were obtained by sequence walking via custom-synthesized primers. All of the sequences were confirmed by sequencing both strands completely. Automatic sequencing was done commercially (Mission Biotech, Taiwan). Sequence data were compiled and analyzed using three computer programs: GeneWorks (IntelliGenetics, Campbell, CA), UWGCG (release 9.0; Genetics Computer Group, Madi-

son, WI), and Neural Network for promotor prediction (NNPP) (Reese, 1994; Reese and Eeckman, 1995; Reese *et al.,* 1996). The DNA and the deduced amino acid sequences were compared with the latest GenBank/EMBL, SWISSPORT, and PIR databases using FASTA and BLAST.

WSSV *tk-tmk* transcription analysis

Preparation of RNA for analysis. Healthy (that is, two-step WSSV diagnostic PCR negative) (Lo *et al.,* 1996a) subadult *P. monodon* (15–20 g) were infected with WSSV using the method described previously by Tsai *et al.* (1999). Briefly, the gills, stomach, and exoskeleton from *P. monodon* with a pathologically confirmed WSSV infection were homogenized (0.1 g/ml in 0.9% NaCl) and then centrifuged at 1000 *g* for 10 min. The supernatant was diluted to 1:100 with 0.9% NaCl and filtered through a 0.45-μm filter. For the experimental infection, 50 μl of the filtrate was injected into each shrimp between the tergal plates at the dorsal lateral region of the fourth abdominal segment forward into the third abdominal segment. At various times over the course of the next 60 h, two or three specimens were selected at random and their pereiopods were excised. The collected pereiopods were immediately frozen and stored in liquid nitrogen until used as described below.

For the isolation of total RNA, the frozen pereiopods (500 mg) from WSSV-infected *P. monodon* were homogenized in 5 ml TRIzol-LS reagent (Life Technologies, Gaithersburg, MD) and then subjected to ethanol precipitation according to the manufacturer's recommendations. The total RNA was stored in 75% ethanol at −20°C.

Detection of WSSV tk-tmk transcripts in WSSV-infected shrimp by Northern blot hybridization analysis with tk-tmk gene-specific riboprobes. The WSSV *tk-tmk*-specific [α-^{32}P]rCTP-labeled riboprobes for Northern blot analysis were generated by *in vitro* transcription (Sambrook *et al.,* 1989) using T7 RNA polymerase (Boehringer Mannheim, Indianapolis, IN). The RNA polymerase promoter addition kit Lig'nScribe (Ambion, Inc., Austin, TX) was used in accordance with the manufacturer's instructions to generate templates from WSSV *tk-tmk*-specific PCR product for the *in vitro* transcription. T7 RNA polymerase promoter was added to the PCR fragment by the Lig'nScribe reaction. The WSSV *tk-tmk*-specific PCR product was amplified by the primer set *tk*-R1 and *tk*-F1 (5′-AATACAAGATCGGGCTGAGG-3′ and 5′-CTCATGCCGGAGAAGTGATT-3′, respectively).

For Northern blotting, approximately 10 μg total RNA was separated on 1% formaldehyde-agarose gel, transferred to a Hybond-N$^+$ membrane (Amersham, Pharmacia Biotech, Piscataway, NJ) and then hybridized to a WSSV *tk-tmk*-specific riboprobe using standard protocols (Sambrook *et al.,*1989).

Temporal analysis of WSSV tk-tmk transcription by RT-PCR. Total RNA in 75% ethanol was centrifuged at 14,000 *g* for 30 min at room temperature. The pellet was resuspended in DEPC-water and quantified by spectrophotometry at 260 nm. An aliquot of 10 μg RNA was treated with 200 U of RNase-free DNase I at 37°C for 30 min to remove any residual DNA and then extracted with phenol-chloroform. The DNase-treated total RNA (~10 μg) was denatured by heating at 85°C for 10 min in 10 μl DEPC-water containing 100 pmol oligo dT primer (Roche Diagnostics GmbH Roche Molecular Biochemicals, Mannheim, Germany). The first-strand cDNA was synthesized by the addition of 4 μl Superscript II 5× buffer, 1 μl 100 mM DTT, 1 μl 10 mM dNTPs, 10 U RNasin (Promega), and 100 U Superscript II reverse transcriptase (Life Technologies). DEPC-water was added to make a final volume of 20 μl. The reverse transcription proceeded at 37°C for 1 h, followed by heating at 95°C for 5 min to stop the reaction. One tenth of the products of the cDNA reaction (2 μl; ~1 μg) was subjected to RT-PCR in a 50-μl reaction buffer containing 10 mM Tris–HCl, pH 8.8, 1.5 mM $MgCl_2$, 150 mM KCl, 0.1% Triton X-100, 0.2 mM dNTPs, 100 pmol of each primer (*tk*-F1/*tk*-R1 for *tk*; *rr1*-F1/*rr1*-R1 [5′-GACATGATTATGCGTGTGG-3′/5′-CTCCCGGTAATCAACATAG-3′] for *rr1*; and *rr2*-F1/*rr2*-R1 [5′-ATGCAAGTAGATACGTTTAAACAC-3′/5′-CAATTCGGACGGCTTCAGTGAC-3′] for *rr2,* where *rr1* and *rr2* are the genes that encode, respectively, the large and small subunits of ribonucleotide reductase; they are used here for comparison), and 2 U DyNAzyme II DNA polymerase (Finnzymes, Espoo, Finland). The PCR cycles were as follows: 94°C for 2 min, 40 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, followed by an elongation at 72°C for 30 min. The β-actin transcript amplified by using actin-F1/actin-R1 primer set (5′-GAYGAYATGGAGAAGATCTGG-3′/5′-CCRGGGTACATGGTGGTRCC-3′) served as an internal control for RNA quality and amplification efficiency.

In vitro transcription and translation

A single expression vector with the WSSV *tk-tmk* gene (nt −30 to 1158) was constructed for *in vitro* transcription and translation. The insertion region was amplified from pms94 by PCR with synthetic oligomers. The 5′ primer was GAC GGA TCC GCA TAT GCA ACT CAT TCT TTC T and the 3′ primer, AAT TCT TTA CTC GAG AAC AAT ATT ATA. The underlined bases indicate the restriction sites. The 5′ primer was designed to insert an *Nde*I site immediately 5′ to nt −30 and the 3′ primer was designed to insert an *Xho*I site immediately 3′ to nt 1158. Twenty-five cycles consisting of 45 s at 94°C, 45 s at 50°C, and 80 s at 72°C were run using 50 ng of pms94, 100 pmol of each primer, 0.2 mM of each dNTP, and 1 unit of *Vent*$_R$ (New England BioLabs, Beverly, MA) DNA polymerase. The PCR product was cloned into pGEM-T Easy vector (Pro-

mega) according to the manufacturer's instructions, and the resulting construct was designated pGTETK, which was digested with the restriction endonucleases *Nde*I and *Xho*I. The resulting *Nde*I/*Xho*I fragment containing the WSSV TK-TMK coding region was purified by 1% agarose gel electrophoresis, extracted with the QIAquick gel extraction kit (Qiagen, Chatsworth, CA), and then ligated with *Nde*I- and *Xho*I-digested pET20b(+) (Novagen). The resulting expression construct was designated pETTK. This expression vector was purified by QIAprep spin miniprep kit (Qiagen), and quantified by spectrophotometry at 260 nm.

TNT (Promega) Quick Master Mix (8 μl) was mixed with 0.4 μl [^{35}S]methionine (1000 Ci/mmol, 10 mCi/ml) and pETTK (0.2 μg in 1.6 μl nuclease-free water). The reaction mixture was incubated at 30°C for 90 min, and the translation product (4.5 μl) was analyzed by 12.5% SDS–PAGE. After electrophoresis, the gel was stained, destained, dried, and exposed to Fuji medical film at room temperature for 17 h.

Amino acid sequence comparison

The following selected TK and TMK proteins available in GenBank were used in the alignment and phylogenetic analysis: African swine fever virus (ASFV; KIXFAS/JQ2161), *Aeropyrum pernix* (Ap; BAA81101), *Aquifex aeolicus* (Aqa; O67099), *Amsacta moorei* entomopoxvirus (AmEPV; P28852), *Arabidopsis thaliana* (At; AAF21190/AAC33288), *Archaeoglobus fulgidus* (Arf; O30175), *Bacillus subtilis* (Bs; Q03221/P37537), *Caenorhabditis elegans* (Ce; CAA19501/CAA91347), Camelpox virus (CamPV; AAB24618), Canarypox virus (CaPV; BAA77559), Chilo iridescent virus (CIV; AAB94460), *Choristoneura biennis* entomopoxvirus (CbEPV; Q05879), *Choristoneura fumiferana* entomopoxvirus (CfEPV; Q05880), Coliphage T4 (T4; AAD42663), Cotia virus (CoV; AAF21102), Cowpox virus (CPV; AAF21104), Cricetulus griseus (Cg; P09768), *Deinococcus radiodurans* (Dr; AAF11536), Ectromelia virus (EV; AAF21103), Encephalitozoon cuniculi (Enc; CAA07261), *Escherichia coli* (Ec; AAC74320/BAA35905), Fowlpox virus (FWPV; P10052), *Gallus gallus* (Gg; P04047), *Helicobacter pylori* J99 (Hp; AAD06956), *Homo sapiens* (Hs; P04183/P23919), *Methanococcus jannaschii* (Mj; AAB98278), Monkeypox virus (MPV; KIVZMV), *Mus musculus* (Mm; P04184/P97930), Mycoplasma pneumoniae (Mp; P75070), Myxoma virus (MV; P28851), Oryza sativa (Os; AAC31168), Pyrococcus horikoshii (Ph; O59366), Raccoonpox virus (RaPV; AAA93128), Rabbit fibroma virus (RFV; P07605), Rattus norvegicus (Rn; AAA75560), Rhodothermus sp. (Rsp; AAC98909), *Saccharomyces cerevisiae* (Sc; KIBYT8), *Schizosaccharomyces pombe* (Sp; CAA19357), Sheeppox virus (ShPV; BAA00324), Streptococcus gordonil challis (Sgc; P47848), Sulfolobus solfataricus (Ss; CAB57523), Swinepox virus (SwPV; P23335), Synechocystis Pcc6803 (Syn; Q55593), *Thermotoga maritima* (Tm; AAD35486/AAD36175), Vaccinia virus (VV; P13410), Variola virus (VAR; KIVZVV/P33803), and Yaba monkey tumor virus (YMTV; BAA88780). The pairwise comparisons of amino acid sequences were done using GeneDoc (Nicholas *et al.,* 1997). The multiple sequence alignments were done by the multiple sequence alignment program CLUSTAL_X (Thompson *et al.,* 1997). Phylogenetic analyses based on TK and TMK gene amino acid sequences were performed using neighbor-joining (NJ) and parsimony methods with the PAUP 4.0b1 program (Swofford, 1998), using CLUSTAL_X (Thompson *et al.,* 1997) to produce input files of aligned protein sequences. One thousand bootstrap replicates were generated to test the robustness of the trees.

ACKNOWLEDGMENTS

We are grateful to Shao-Wen Sun for her help in sequence analysis. This work was supported by the Council of Agriculture Grants 89-BT-2.1-FID-01(13), NSC89-2311-B-002-041, and NSC89-2311-B-002-042. We are indebted to Paul Barlow for his helpful criticism of the manuscript.

REFERENCES

Blasco, R., Lopez-Otin, C., Munoz, M., Bockamp, E. O., Simon-Mateo, C., and Vinuela, E. (1990). Sequence and evolutionary relationships of African swine fever virus thymidine kinase. *Virology* **178**, 301–304.

Boyle, D. B., Coupar, B. E., Gibbs, A. J., Seigman, L. J., and Both, G. W. (1987). Fowlpox virus thymidine kinase: Nucleotide sequence and relationships to other thymidine kinases. *Virology* **156,** 355–365.

Chou, H. Y., Huang, C. Y., Wang, C. H., Chiang, H. C., and Lo, C. F. (1995). Pathogenicity of a baculovirus infection causing white spot syndrome in cultured penaeid shrimp in Taiwan. *Dis. Aquat. Org.* **23,** 165–173.

Esposito, J. J., and Knight, J. C. (1984). Nucleotide sequence of the thymidine kinase gene region of monkeypox and variola viruses. *Virology* **135,** 561–567.

Flegel, T. W. (1997). Special topic review: Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. *World J. Microbiol. Biotechnol.* **13,** 433–442.

Gentry, G. A. (1992). Viral thymidine kinases and their relatives. *Pharmacol. Ther.* **54,** 319–355.

Grivell A. R., and Jackson, J. F. (1968). Thymidine kinase: Evidence for its absence from *Neurospora crassa* and some other microorganisms, and the relevance of this to the specific labelling of deoxyribonucleic acid. *J. Gen. Microbiol.* **54,** 307–317.

Jong, A. Y. S., Kou, C., and Campbell, J. L. (1984). The CDC8 gene of yeast encodes thymidylate kinase. *J. Biol. Chem.* **259,** 11052–11059.

Karlin, S., Mocarski, E. S., and Schachtel, G. A. (1994). Molecular evolution of herpesviruses: Genomic and protein sequence comparisons. *J. Virol.* **68,** 1886–1902.

Kit, S. (1985). Thymidine kinase. *Microbiol. Sci.* **2,** 369–375.

Kleiboeker, S. B., Scoles, G. A., Burrage, T. G., and Sur, J. (1999). African swine fever virus replication in the midgut epithelium is required for infection of *Ornithodoros* ticks. *J. Virol.* **73,** 8587–8598.

Koonin, E. V., and Senkevich, T. G. (1993). Fowlpox virus encodes a protein related to human deoxycytidine kinase: Further evidence for independent acquisition of genes for enzymes of nucleotide metabolism by different viruses. *Virus Genes* **7,** 289–295.

Kozak, M. (1995). Adherence to the first-AUG rule when a second AUG codon follow closely upon the first. *Proc. Natl. Acad. Sci. USA* **92,** 1734.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is

augmented by G in position +4 but not generally affected by the nucleotides in positions +5 and +6. *EMBO J.* **16,** 2482–2492.

Lo, C. F., Ho, C. H., Chen, C. H., Liu, K. F., Chiu, Y. L., Yeh, P. Y., Peng, S. E., Hsu, H. C., Liu, H. C., Chang, C. F., Su, M. S., Wang, C. H., and Kou, G. H. (1997). Detection and tissue tropism of white spot syndrome baculovirus (WSBV) in captured brooders of *Penaeus monodon* with a special emphasis on reproductive organs. *Dis. Aquat. Org.* **30,** 53–72.

Lo, C. F., Hsu, H. C., Tsai, M. F., Ho, C. H., Peng, S. E., Kou, G. H., and Lightner, D. V. (1999). Specific genomic fragment analysis of different geographical clinical samples of shrimp white spot syndrome virus. *Dis. Aquat. Org.* **35,** 175–185.

Lo, C. F., Leu, J. H., Ho, C. H., Chen, C. H., Peng, S. E., Chen, Y. T., Chou, C. M., Yeh, P. Y., Huang, C. J., Chou, H. Y., Wang, C. H., and Kou, G. H. (1996). Detection of baculovirus associated with white spot syndrome (WSBV) in penaeid shrimps using polymerase chain reaction. *Dis. Aquat. Org.* **25,** 133–141.

Nicholas, K. B., Nicholas, H. B., Jr., and Deerfield II, D. W. (1997). GeneDoc: Analysis and visualization of genetic variation. *EMBNEW. NEWS* **4,** 14.

Reese, M. G. (1994). Erkennung von Promotoren in pro- und eukaryontischen DNA-Sequenzen durch k stliche Neuronale Netze. Master Thesis. German Cancer Research Center, Heidelberg, Germany.

Reese, M. G., and EecKman, F. H. (1995). New neural network algorithms for improved eukaryotic promoter site recognition. *In* "The Seventh International Genome Sequencing and Analysis Conference." Hilton Head Island, SC.

Reese, M. G., Harris, N. L., and EecKman, F. H. (1996). Large scale sequencing specific neural networks for promoter and splice site recognition. *In* "Biocomputing: Proceedings of the 1996 Pacific Symposium" (L. Hunter and T. E. Klein, Eds.). World Scientific Publishing, Singapore.

Robertson, G. R., and Whalley, J. M. (1988). Evolution of the herpes thymidine kinase: Identification and comparison of the equine herpesvirus 1 thymidine kinase gene reveals similarity to a cell-encoded thymidylate kinase. *Nucleic Acids Res.* **16,** 11303–11317.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Smith, G. L., de Carlos, A., and Chan, Y. S. (1989). Vaccinia virus encodes a thymidylate kinase gene: Sequence and transcriptional mapping. *Nucleic Acids Res.* **17,** 7581–7590.

Swofford, D. L. (1998). PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, MA.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882.

Tsai, M. F., Kou, G. H., Liu, H. C., Liu, K. F., Chang, C. F., Peng, S. E., Hsu, H. C., Wang, C. H., and Lo, C. F. (1999). Long-term presence of white spot syndrome virus (WSSV) in a cultured shrimp population without disease outbreaks. *Dis. Aquat. Org.* **38,** 107–114.

Tsai, M. F., Lo, C. F., van Hulten, M. C. W., Tzeng, H. F., Chou, C. M., Hung, C. J., Wang, C. H., Lin, J. W., Vlak, J. M., and Kou, G. H. (2000). Transcriptional analysis of the ribonucleotide reductase genes of shrimp white spot syndrome virus. *Virology* **277,** 92–99.

van Hulten, M. C. W., Tsai, M. F., Schipper, C. A., Lo, C. F., Kou, G. H., and Vlak, J. M. (2000). Analysis of a genomic segment of white spot syndrome virus of shrimp containing ribonucleotide reductase genes, and repeat regions. *J. Gen. Virol.* **81,** 307–316.

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1,** 945–951.

Wang, C. H., Lo, C. F., Leu, J. H., Chou, C. M., Yeh, P. Y., Chou, H. Y., Tung, M. C., Chang, C. F., Su, M. S., and Kou, G. H. (1995). Purification and genomic analysis of baculovirus associated with white spot syndrome (WSBV) of *Penaeus monodon*. *Dis. Aquat. Org.* **23,** 239–242.

Yanez, R. J., Rodrigues, J. M., Rodrigues, J. F., Salas, M. L., and Vinuela, E. (1993). African swine fever virus thymidylate kinase gene: Sequence and transcriptional mapping. *J. Gen. Virol.* **74,** 1633–1638.

Yuen, L., and Moss, B. (1987). Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. *Proc. Natl. Acad. Sci. USA* **84,** 6417–6421.