Functional characterization and structural modelling of late gene expression factor 4 from *Bombyx mori* nucleopolyhedrovirus

Seema SEHRAWAT*, Narayanaswamy SRINIVASAN† and Karumathil P. GOPINATHAN*[1]

*Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India, and †Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Late gene expression factor 4 (LEF4), a multifunctional protein encoded by the *Bombyx mori* nucleopolyhedrovirus has been bacterially expressed and characterized. Sequence analyses and three-dimensional modelling of *B. mori* LEF4 showed that the protein is related to mRNA-capping enzymes, which are organized as two modular domains. Most of the acidic side chains in LEF4 were solvent-exposed and spread all along the fold. A region dominated by negatively charged groups, which protrudes from the larger domain was ideally suited for interactions with proteins having positively charged patches at the surface. The purified LEF4 protein exhibited different enzyme activities associated with mRNA-capping enzymes, i.e. GTP-binding, RNA triphosphatase and guanylate transferase activities. In addition, LEF4 also showed NTP-hydrolysing activity. The kinetic analysis of ATP hydrolysis revealed a sigmoidal response with two deduced binding sites for ATP, whereas the guanylate transferase activity showed a typical hyperbolic response to varying concentrations of GTP with a K_m of $330 \pm 20\ \mu$M. Analysis of the modelled three-dimensional structure of LEF4 suggested the presence of crucial residues in sequence motifs important for the integrity of the fold. Mutation of one such conserved and buried tyrosine residue to cysteine in the motif IIIa, located close to the interlobe region of the model, resulted in a 44 % loss of guanylate transferase activity of LEF4 but had no effect on the ATPase activity.

Key words: ATPase, capping enzyme, guanylate transferase, mRNA capping, RNA triphosphatase.

INTRODUCTION

Bombyx mori nucleopolyhedrovirus (BmNPV) infects the mulberry silkworm *B. mori* and is of major economic concern to sericulture. This virus has also been exploited as a baculovirus expression vector and is next only to the most widely exploited *Autographa californica* multinucleocapsid nucleopolyhedrovirus (AcMNPV) in popularity [1–6]. It is a rod-shaped enveloped virion which contains a circular double-stranded DNA genome of 128.41 kb, encoding 143 predicted open reading frames (ORFs) [7].

Baculovirus gene expression is tightly regulated during viral infection and is classified into three distinct phases: early, late and very late. The early genes are transcribed by host RNA polymerase II [8,9]. The composition of the late transcriptional complex is not fully understood. Transient expression systems combined with complementation analysis have allowed identification of a class of genes, designated as late gene expression factors (*lef*s) in AcMNPV, which influence the late and very late gene expression [10–12]. Of these, *lef*s-4, -5, -6, -7, -8, -9, -10 and -12 and *p47* directly affect late gene transcription whereas *lef*s-1, -2 and -3, and *ie*1, *dnahel* (originally named *p143*) and *dnapol* are essential for viral DNA replication [10,13,14] and thereby influence the expression. Some of the encoded proteins such as LEF2 show both effects [13,15] and yet others (e.g. p35) indirectly influence the late gene expression through delay of apoptosis. All these 19 *lef*s [12] also have their counterparts in BmNPV [7]. However, certain baculoviruses lack a few *lef*s (e.g. *lef*7, *lef*10, *lef*12, *p35* and *ie*2 in *Plutella xylostella* granulovirus [16] and *ie*2, *pe*38, *lef*7 and *p35* in *Spodoptera exigua* and *Lymantria dispar* multinucleocapsid nucleopolyhedroviruses [17]).

Transition between the early and late phases of gene expression in baculoviruses is mediated through a switch from host RNA polymerase II to a novel α-amanitin-resistant RNA polymerase activity [18] that initiates transcription from the late and very late viral promoters but not the early promoters [19]. This minimal polymerase from AcMNPV, constituting four equimolar subunits, LEF4, -8, -9 and P47 [20], exhibited both catalytic and specific promoter-recognition activities. Of these, LEF8 and LEF9 harbour the conserved RNA polymerase motifs. LEF9 shares some sequence similarity (5–7 amino acid residues) with a highly conserved motif, NADFDGD, of the largest subunit of DNA-dependent RNA polymerases [21], and the C-terminal region of LEF8 contains a 13-residue motif ($GXKX_4HGQ$/NKG) that is highly conserved among the RNA polymerases [22]. The other two subunits, LEF4 and P47, do not harbour any recognizable polymerase sequence motifs. However, LEF4 from AcMNPV shows guanylate transferase activity (mRNA-capping function) *in vitro* [23–25]. Here we report the characterization of *lef*4 from BmNPV. The bacterially expressed glutathione S-transferase (GST)-tagged protein has been purified and its biochemical properties analysed. LEF4 showed remarkable similarity to the mRNA-capping enzymes and a three-dimensional model has been generated on the basis of known structure of an mRNA-capping enzyme. In agreement with the prediction, purified LEF4 exhibited all three associated enzymic activities, i.e. RNA triphosphatase, GTP-binding and guanylate transferase, characteristic of the capping enzymes. In addition, the protein also showed NTP-hydrolysing activity.

Abbreviations used: AcMNPV, *Autographa californica* multinucleocapsid nucleopolyhedrovirus; BmLEF4, *B. mori* LEF4; BmNPV, *Bombyx mori* nucleopolyhedrovirus; DTT, dithiothreitol; GST, glutathione S-transferase; IPTG, isopropyl β-D-thiogalactoside; *lef*, LEF, gene and protein for late gene expression factor respectively; ORF, open reading frame; PBCV-1, *Paramecium bursaria Chlorella* virus 1; TCA, trichloracetic acid.

[1] To whom correspondence should be addressed (e-mail kpg@mcbl.iisc.ernet.in).

MATERIALS AND METHODS

Cells and viruses

BmNPV-BGL, a local isolate of BmNPV [26] used in the present study, is nearly identical in restriction-digestion profiles to the prototype T3 strain (GenBank accession no. L33180) except for a few polymorphic restriction sites. The virus stocks were maintained in *B. mori*-derived BmN cells and the viral titres were determined according to standard protocols [27]. Genomic DNA was isolated from the free virions released from the polyhedral bodies isolated from the virus-infected *B. mori* larvae or from purified virus preparations grown in BmN cells.

Plasmid constructs and DNA sequences

The *lef*4 ORF from BmNPV-BGL was PCR amplified using primers P1 (5′-TCGAATTCATGGACCACGGCAATTT-3′, encompassing +1 ATG) and P2 (5′-CGAATTCAAGCTTGT-CGTCAACTGTCC-3′, covering the C-terminal region of LEF4, including the termination codon) and *Taq* DNA polymerase. The PCR product was digested with *Eco*RI and cloned into plasmid vector pBSKS$^+$ to generate the clone pBmLEF4. The *lef*4 insert DNA was completely sequenced (GenBank accession no. AY033655) and mobilized into plasmid vector pGEX4x-1 to facilitate purification of the expressed protein with the GST tag (plasmid pGST-*lef*4).

Expression and purification of GST-LEF4

The plasmid pGST-*lef*4 was transformed into *Escherichia coli* strain BL21 and induced in the presence of 0.4 mM isopropyl β-D-thiogalactoside (IPTG) for 4 h at 37 °C. The bacterial cell pellet was sonicated and the expressed protein, mostly present in the insoluble fraction, was suspended in 5 M guanidinium chloride. After 1 h incubation on ice, the solubilized proteins were refolded by dialysis against decreasing concentrations of guanidinium chloride, followed by dialysis against PBS for 5 h. The dialysate was incubated with glutathione–Sepharose CL 4B (Amersham Biosciences) for 2 h at 4 °C and the matrix was washed with 50 mM Tris buffer (pH 7.5), containing 100 μM PMSF, 1 mM EDTA, 300 mM dithiothreitol (DTT) and 100 mM NaCl. The bound proteins were eluted using 10 mM glutathione in the same buffer. For all biochemical analysis, the GST–Sepharose-affinity-purified protein was used.

LEF4–GMP complex formation

The assay mixture for monitoring GMP–LEF4 complex formation contained, in a final volume of 20 μl, 50 mM Tris/HCl (pH 8.0), 20 mM $MgCl_2$, 5 mM DTT, [α-^{32}P]GTP (2–5 μCi) and the purified GST-LEF4 (780 pg/μl, equivalent to 10 nmol of the protein). The samples were incubated for 15 min at 30 °C and the reaction was stopped by the addition of 1 % SDS and 10 % trichloracetic acid (TCA). The precipitated protein was pelleted by centrifugation, dissolved in gel-loading buffer containing the dye, boiled and subjected to SDS/PAGE (10 % acrylamide/0.1 % SDS). The gels were dried and exposed to X-ray film for autoradiography.

RNA triphosphatase assay

As a substrate for RNA triphosphatase activity, RNA 5′-end-labelled with γ-^{32}P was synthesized *in vitro* by transcription from Bluescript plasmid pBSKS$^+$ (linearized by digestion with *Sac*I) using T3 RNA polymerase in the presence of [γ-^{32}P]ATP. The 5′-end-labelled RNA (137 nt) was purified after extraction with phenol/chloroform by passing through a column of Sephadex G-50. The labelled RNA was precipitated with 2.5 vol. of ethanol and dissolved in a minimal volume of water. RNA triphosphatase assay was based on the release of [^{32}P]P_i from the 5′-end-labelled RNA. The reaction mixture contained, in a final volume of 20 μl, 50 mM Tris/HCl (pH 7.9), 5 mM DTT, 2.0 mM $MgCl_2$, 5′-end-labelled [γ-^{32}P]RNA (40000 c.p.m.) and affinity-purified LEF4 (as indicated). Incubations were carried out for 15 min at 30 °C. Samples (5 and 10 μl) were spotted on to glass-fibre filters and the remaining labelled RNA was estimated using liquid scintillation spectroscopy.

Capping (guanylate transferase) activity

RNA-capping (guanylate transferase) activity was monitored by transfer of ^{32}P label from [α-^{32}P]GTP to the unlabelled substrate RNA. The unlabelled substrate RNA was generated *in vitro* from the linearized plasmid pBSKS$^+$ (by treatment with restriction enzyme *Sac*I and transcribed using T3 RNA polymerase in presence of all four NTPs). The *in vitro*-synthesized, unlabelled RNA (137 nt) was purified as in the case of labelled RNA (see the previous section). The RNA-capping reaction mixture contained, in a final volume of 20 μl, 50 mM Tris/HCl (pH 8.0), 1–2 mM $MgCl_2$, 5 mM DTT, *in vitro*-synthesized unlabelled RNA, [α-^{32}P]GTP (2 μCi) and affinity-purified GST-LEF4 (as indicated). Incubations were carried out at 37 °C for 30 min. The reaction mixture was first extracted with phenol/chloroform to remove the protein-bound GMP (LEF4–GMP) formed during the reaction. The RNA in the aqueous phase was then precipitated with 10 % TCA and the radioactivity (label transferred to RNA) was monitored in a liquid scintillation counter. The RNA samples were also analysed by electrophoresis on 7 M urea/8 % polyacrylamide gels to confirm the incorporation of radiolabel into the RNA.

NTPase reaction

The reaction mixture contained, in a final volume of 20 μl, 50 mM Tris/HCl (pH 7.5), 5 mM DTT, 1 mM $MnCl_2$, [γ-^{32}P]NTP or [α-^{32}P]NTP (0.02 μCi) and affinity-purified GST-LEF4 (7.5–10 nmol). Incubation was carried out at 30 °C for 30 min and 2 μl samples were spotted on to poly(ethyleneimine) cellulose TLC plates. The solvent system used for TLC was 1.2 M LiCl/100 mM EDTA. The conversion of NTP to NDP and P_i was monitored and quantified in a PhosphorImager or by the direct determination of radioactivity present in NDP, and expressed as the percentage of substrate converted.

Generation of LEF4 antiserum

To raise antiserum against LEF4, the bacterially expressed gel-purified full-length LEF4 carrying the GST tag was utilized. The fusion protein expressed in *E. coli* BL21 cells was subjected to electrophoresis (10 % acrylamide/0.1 % SDS) and the 78 kDa protein band corresponding to GST-LEF4 was eluted from the gel. Following elution from polyacrylamide gel, approx. 700 μg of the fusion protein was injected subcutaneously into a rabbit, after mixing with complete Freund's adjuvant for the initial injection and incomplete Freund's adjuvant for the subsequent booster injections (2–3 weeks between injections). Serum from the immunized rabbit was tested for the presence of LEF4 antibodies by Western blotting. The antibodies were purified by passing through a Protein A–Sepharose column and the antibody

titres were determined by ELISA using the bacterially expressed, purified LEF4 protein as standard.

Western blot analysis

The proteins were subjected to SDS/PAGE (0.1 % SDS/10 % polyacrylamide) and were electrophoretically transferred from the gels to a PVDF-plus transfer membrane for 2 h. The membranes were probed using a 1:4000 dilution of the anti-LEF4 polyclonal antiserum for 4 h at room temperature, followed by incubation with goat anti-rabbit IgG conjugated to horseradish peroxidase, and detected by peroxidase activity using diaminobenzidine reaction (0.04 % diaminobenzidine and 0.3 % H_2O_2).

Molecular modelling of LEF4

The sequence of *B. mori* LEF4 (BmLEF4) was searched for in the non-redundant database using BLAST and similar programs. Analysis was also carried out to match the sequence with profiles of protein domain families. Three-dimensional fold recognition was attempted using GenTHREADER [28], the inverse folding method [29] and 3D-PSSM [30]. These analyses suggested a remarkable similarity of the BmLEF4 sequence to those of mRNA-capping enzymes. The structures of the conserved regions of mRNA-capping enzymes which showed the best similarity to LEF4 were used to define a framework of the model. The overall sequence identity of LEF4 to the capping enzyme of *Paramecium bursaria Chlorella* virus 1 (PBCV-1; PDB code for the crystal structure, 1CKO) was 17.7 %. However, when the occurrence of similar amino acid residues in the mismatch positions was scored, the sequence similarity increased to 55.9 %. A given conserved region of the LEF4 was modelled by extrapolating the sequence in the local region to the equivalent region of 1CKO. The suite of programs encoded in COMPOSER [31–35] and incorporated in SYBYL (Tripos, St. Louis, MO, U.S.A.) was used to generate the three-dimensional model of LEF4. The variable regions were modelled after identifying a suitable segment by searching in the databank of known protein structures and using a template-matching approach to rank the candidate loops [34]. Loops with no short contacts with the rest of the protein were fitted using a ring-closure procedure. Side chains were modelled either by extrapolating from the equivalent positions in the basic structure or by using rules derived from the analysis of known protein structures [32].

The COMPOSER generated model was energy minimized in SYBYL using the AMBER force field [36]. During the initial cycles of energy minimization the backbone was kept rigid and side chains alone were moved. Subsequently, all atoms in the structure were allowed to move. Energy minimization was performed until all short contacts and inconsistencies in geometry were rectified. In the initial stages of minimization the electrostatic term was not included because the main objective was to relieve steric clashes and rectify bad geometry. The modelled structure was analysed using the software SETOR [37].

Since energy minimization was performed, the final modelled structure did not have any short contacts between non-bonded atoms. Stereochemical validation of the model using the program PROCHECK [38] showed that more than 98 % of the Ramachandran angles of the non-glycyl residues were situated in the fully allowed region and the rest of the conformational points were situated in the generously allowed region. The PROCHECK outputs also suggested that there were no unusually long or short bond lengths and bond angles. All the peptide units were almost perfectly planar and the distribution of side-chain conformations was similar to those seen in the experimentally determined protein structures.

Site-directed mutagenesis

To generate site-directed mutants of BmLEF4, megaprime-based PCR mutagenesis was performed. Three oligonucleotide primers, two of which were gene-specific primers, P1 (encompassing +1 ATG) and P2 (covering the C-terminal region of LEF4 including the termination codon), and a third primer, P3 (5′-GTTCAC-GCCGCACTCGCACTGTGTTCGATT-3′) containing the desired mutation (Tyr to Cys in domain IIIa), were used. Primers P1 and P3 were utilized for first PCR and the product purified on agarose gel served as the 'megaprimer' for the second PCR along with primer P2. The amplified product was digested with *Eco*RI and subcloned into pBSKS$^+$. The point mutation was confirmed by automated DNA sequencing.

RESULTS

Cloning and sequence analysis of Bm*lef*4

BmNPV homologue of the *lef*4 ORF was PCR-amplified from the genomic DNA of BmNPV-BGL, a local isolate of BmNPV, and cloned into plasmid vector pBSKS$^+$. The 1.4 kb *lef*4 region encoding the full-length ORF was sequenced (GenBank accession no. AY033655), and was nearly identical to the corresponding region of the BmNPV T3 strain except for a single amino acid change at position 198 from Met to Ile. BmLEF4 has a predicted molecular mass of 54 kDa and an acidic pI of 5.15. The sequence alignment of BmLEF4 with those from a number of other baculoviruses revealed that there was 97 % sequence similarity between the AcMNPV and BmNPV counterparts. BmLEF4 contained 465 amino acids, with one serine residue extra at amino acid position 193 compared with its counterpart from AcMNPV. The similarity to the other baculoviruses ranged from 50 to 27 %.

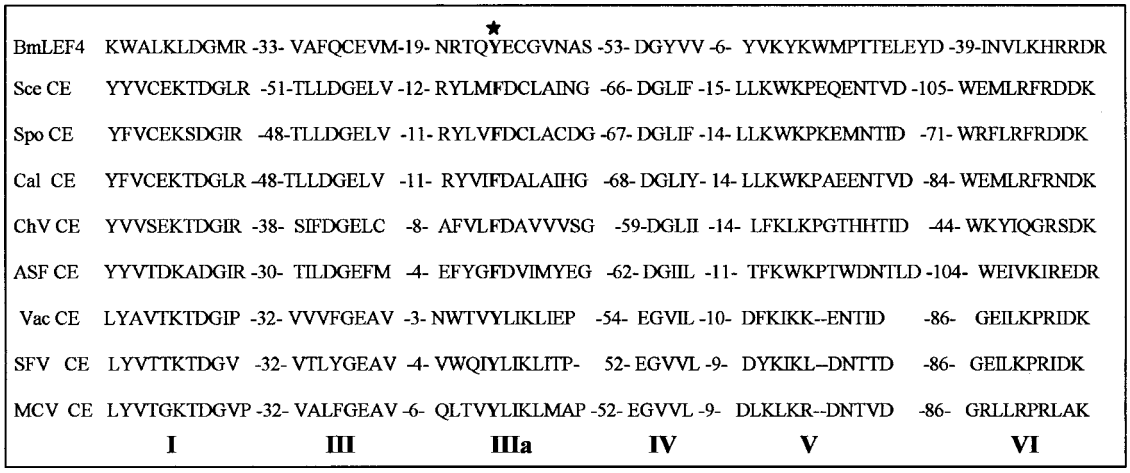
Bacterial expression and purification of BmLEF4

LEF4 was expressed as a GST fusion protein. The bacterially expressed LEF4 was present in the insoluble fraction as inclusion bodies under a variety of induction conditions (various growth temperatures, IPTG concentrations and times for incubation). However, the protein could be solubilized by treatment with 5 M guanidinium chloride and refolded by dialysis against decreasing concentrations of guanidinium chloride (4, 3, 2 and 1 M), and finally against buffer. The protein was purified by adsorption to and elution from the affinity matrix GST–Sepharose. The bacterially expressed GST-LEF4 showed an apparent molecular mass of about 78 kDa, in agreement with the predicted size of LEF4 ORF, including the 26 kDa GST tag, although small amounts of partially degraded protein were visible in some preparations.

Comparative modelling of the three-dimensional structure of LEF4

The amino acid sequences of LEF4, analysed by extensive database searching and matching of signature motifs using multiple sequence alignments, showed significant similarity to the mRNA-capping enzymes from *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Chlorella* virus, African swine fever virus, vaccinia virus, Shope fibroma virus and *Molluscum contagiosum* virus (Figure 1a). The extent of similarity, though generally low, was evident, as suggested by the

(a)



(b)

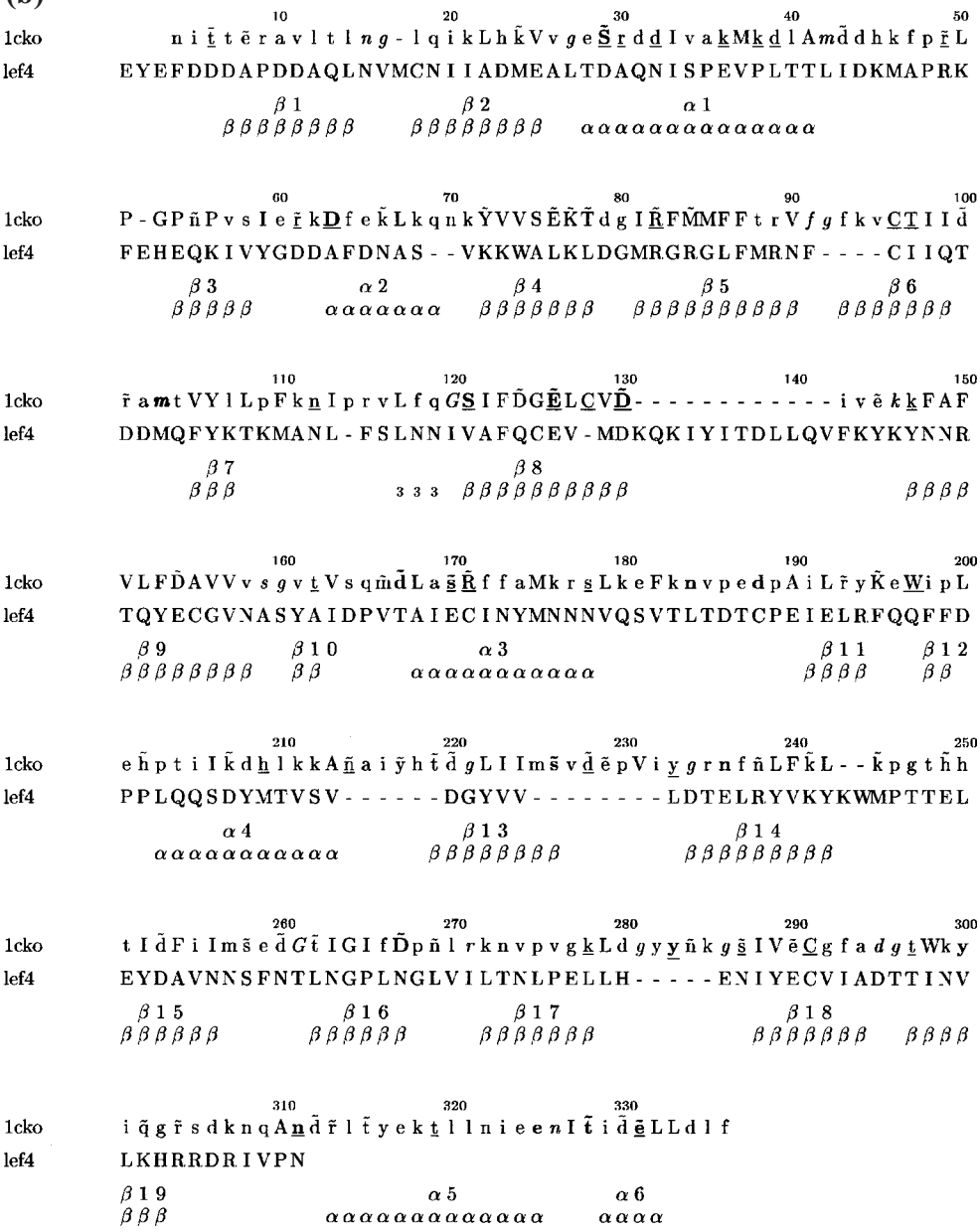
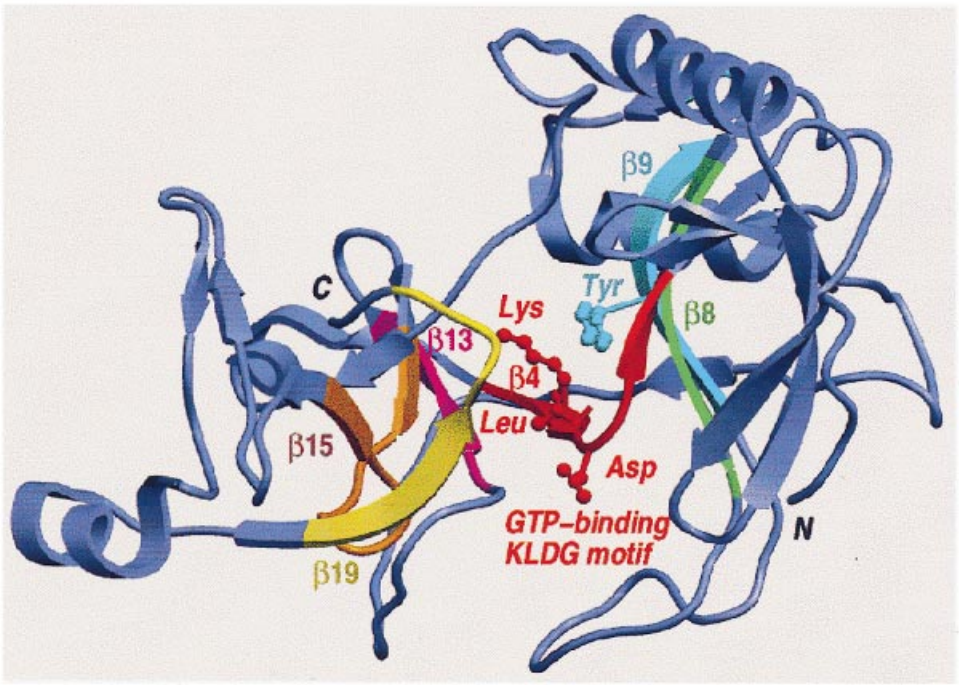


Figure 1 For legend see facing page.

(c)



(d)

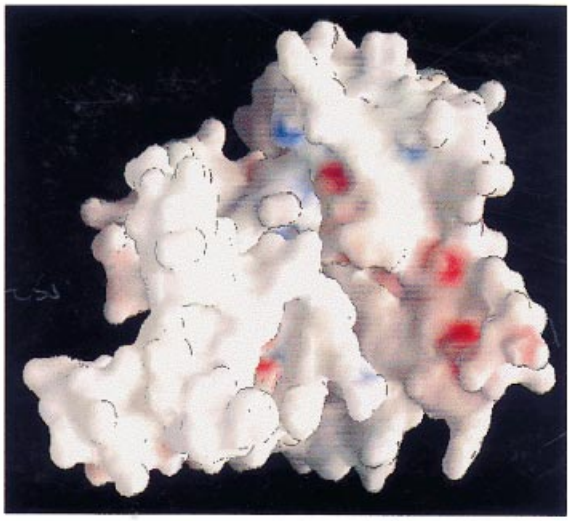


Figure 1 A molecular model of BmLEF4

(**a**) Phylogeny of mRNA-capping enzymes. The amino acid sequences were aligned for capping enzymes (CEs) encoded by *S. cerevisiae* (Sce), *S. pombe* (Spo), *C. albicans* (Cal), *Chlorella* virus (ChV), African swine fever virus (ASF), vaccinia virus (Vac), Shope fibroma virus (SFV) and *M. contagiosum* virus (MCV). The mutation of the Tyr to Cys residue generated in BmLEF4 in the conserved domain IIIa is marked with an asterisk. (**b**) The amino acid sequence of BmLEF4 was aligned to the capping enzyme from PBCV-1 (PDB code, 1CKO). The structural environment of the residues in 1CKO was generated by JOY [48]. Key to JOY alignments: solvent inaccessible (upper case), solvent accessible (lower case), positive (italic), hydrogen bond to other side chain ($\sim$ above the letter), hydrogen bond to main-chain amide (bold), hydrogen bond to main chain carbonyl (underlined), α-helix (α) and 3_{10}-helix (3) and β-sheet (β). Various α-helices and β-strands are marked as $\alpha1$–$\alpha6$ and $\beta1$–$\beta19$ respectively. (**c**) The conserved capping enzyme motifs in the three-dimensional model of BmLEF4 are shown in orange, yellow, magenta, red, green and blue (figure generated using SETOR [37]). Residues involved in GTP binding are marked. (**d**) Electrostatic charge distribution on the surface of LEF4 was generated using GRASP [49]. Acidic regions are in red and basic regions are in blue.

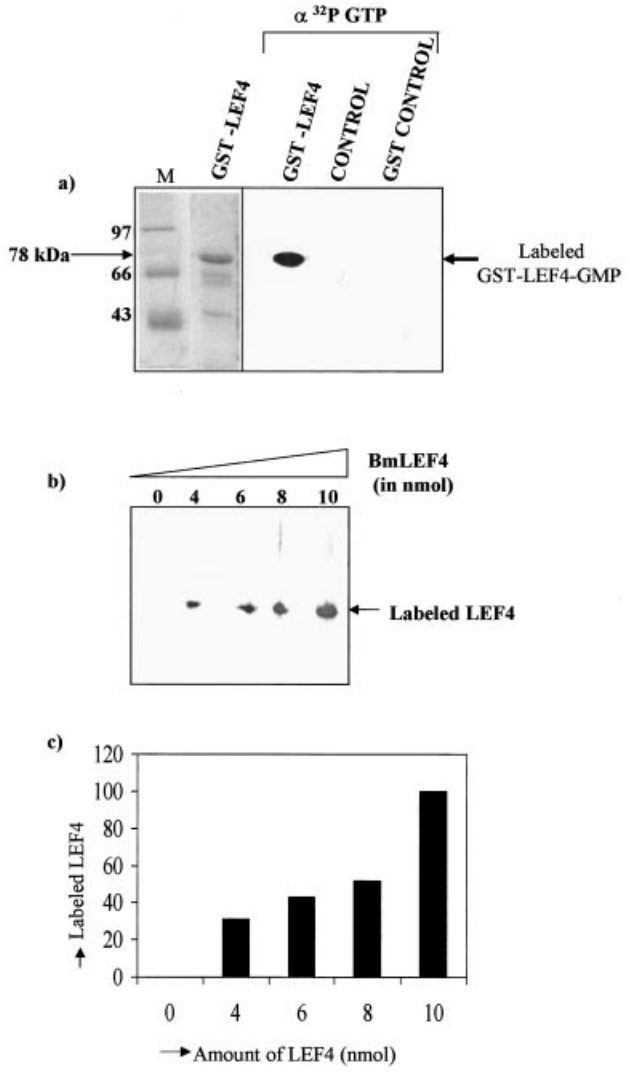


Figure 2 BmLEF4–GMP complex formation

(**a**) Bacterial expression of BmLEF4. (**b**) and (**c**) BmLEF4–GMP complex formation and its quantification. (**a**) *Bmlef4* ORF was PCR-amplified from the BmNPV genomic DNA, cloned into plasmid vector and overexpressed as a GST fusion protein. The bacterially expressed, affinity-purified GST-LEF4 was assayed for guanylate transferase activity by monitoring the transfer of radioactivity from [α-^{32}P]GTP to the LEF4 protein by electrophoresis of the reaction mixture on 0.1% SDS/10% PAGE (right-hand panel marked GST-LEF4). Appropriate controls in which the reactions were terminated at time 0 (lane marked control), or assays containing only GST (lane marked GST control), were included. Other lanes: M, molecular size markers; left-hand lane marked GST-LEF4, bacterially expressed affinity-purified protein stained with Coomassie Blue. (**b**, **c**) BmLEF4–GMP complex formation monitored as a function of increasing concentrations of the protein (**b**) and its quantification by PhosphorImager scanning (**c**) are presented.

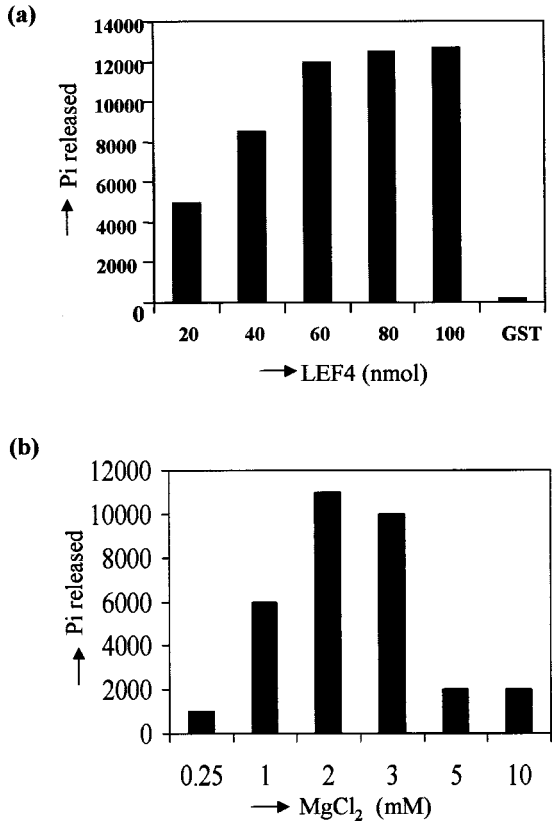


Figure 3 RNA triphosphatase activity of BmLEF4

RNA triphosphatase activity as a function of the BmLEF4 protein (**a**) and concentration of Mg^{2+} (**b**) are presented. For details of the assay, see text. The amount of unhydrolysed labelled RNA remaining was estimated by TCA precipitation and the results are presented as the amount of P_i released (c.p.m. of RNA remaining subtracted from c.p.m. of input RNA).

matches of many key residue positions in the signature motifs. The presence of a few sequence motifs of significant length with a number of conserved or conservatively subsituted residues (Figure 1a) enabled the identification of a good match between LEF4 and mRNA-capping enzymes. The overall sequence identity was 17.7%, but when the occurrence of similar amino acid residues in the mismatch positions were scored the sequence similarity increased to 55.9%.

The presence of a conserved KXDG motif in domain I, Y in domain IIIa and DG in domain IV are important in capping activity [39]. The secondary structure alignment (Figure 1b) between the LEF4 ORF and the mRNA-capping enzyme from *Chlorella* virus PBCV-1 was arrived at by considering the features in the crystal structure of the latter [40]. Features such as solvent accessibility, secondary structure and hydrogen-bonding pattern at various residue positions have been considered in assessing the conformational compatibility of the sequence of LEF4 with the three-dimensional structure of mRNA-capping enzyme from PBCV-1 (PDB code 1CKO). The probability of a residue in LEF4 adopting the structural environment of equivalent residues in the known mRNA-capping enzyme structure has been considered in deciding on the alignment. The region of LEF4 that aligns with 1CKO is amino acid positions 182–465.

As reflected by the extent of sequence identity (see Figure 1b), the mRNA-capping enzyme and LEF4 were only distantly related. Nonetheless, the similarity was clearly implicated by the presence of motifs conserved within the family of mRNA-capping enzymes and LEF4. Most of the α-helical and β-strand regions were conserved between 1CKO and LEF4, a feature characteristic of similarly folded proteins. The length of the secondary structures and packing between them could vary because the relationship between the two proteins was distant but the overall fold was expected to be same.

The overall fold of LEF4 (Figure 1c) deduced on the basis of 1CKO is comprised of two domains. The GTP-binding motif was located in the inter-domain region and was conserved in

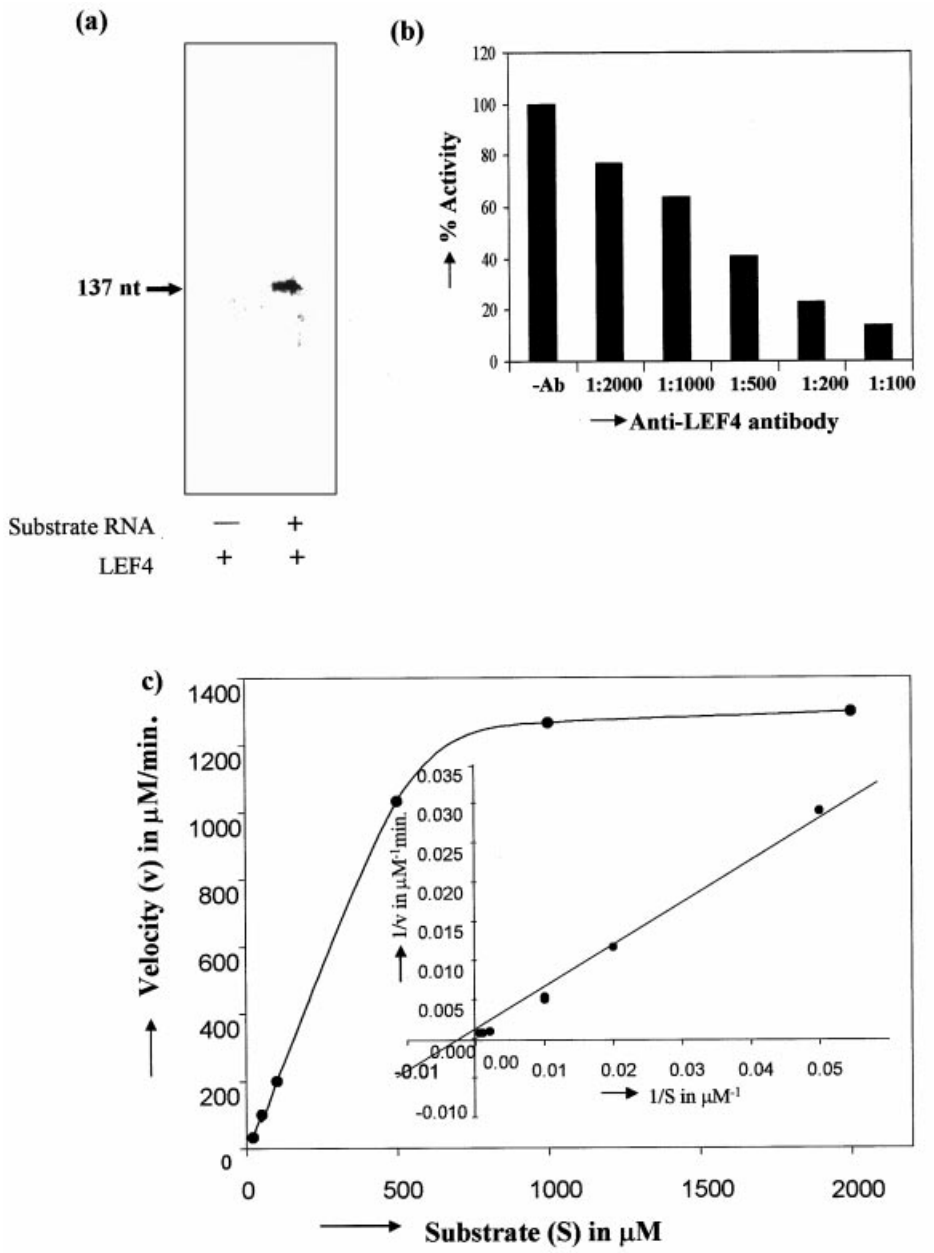


Figure 4 The RNA 5′-end-capping activity of BmLEF4

(**a**) Transfer of label from GTP to RNA. (**b**) Effect of LEF4 antibodies on capping activity. (**c**) Effect of varying concentration of GTP on RNA-capping activity of LEF4. For details of assay, see text. In (**a**), the precipitated RNA sample at the end of the reaction was subjected to electrophoresis on 7 M urea/8% polyacrlamide gel. (**b**) For testing the effect of LEF4 antibodies on RNA-capping activity, the protein was incubated with different dilutions of LEF4-specific polyclonal antibody as indicated and the RNA-capping activity was quantified by the radiolabel transfer from GTP to the RNA. (**c**) The effect of varying GTP concentration on the capping activity was analysed as above. The Lineweaver–Burk plot from a typical experiment is presented as an inset.

terms of residues and location, as in mRNA-capping enzymes. The electrostatic charge distribution of the model is shown in Figure 1(d). The GTP-binding region was slightly basic in nature. A display of all the acidic side chains in LEF4 (results not shown) indicated that all of them were solvent-exposed and spread all along the fold. However, many of the acidic residue charges were compensated by the presence of basic residues in the vicinity (indicated by white patches in Figure 1d). A region which protrudes from the larger domain was dominated by negatively charged groups. This region is ideally suited for interaction with other proteins with predominantly positively charged patches at the surface.

The GTP-dependent capping enzymes and ATP-dependent ligases make up a superfamily of covalent nucleotidyltransferases [41,42]. The relative spatial orientation of six sequence motifs in BmLEF4 (shown in Figure 1a) is perhaps important for NTP binding. The structural regions of these motifs are highlighted in the overall fold of BmLEF4 (Figures 1b and 1c). These sequence motifs cover much of the secondary structural elements in the inter-domain region.

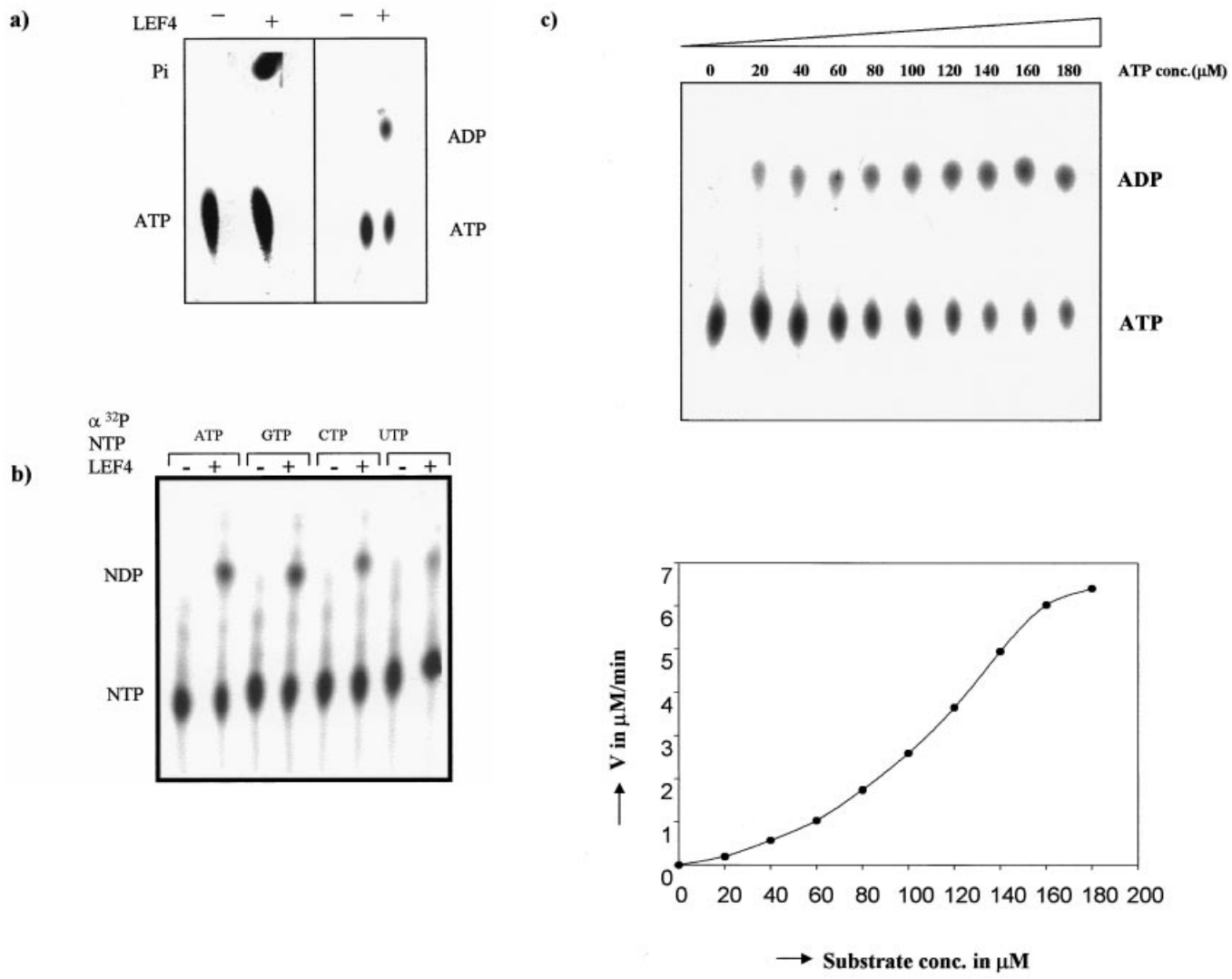


Figure 5 NTPase activity of BmLEF4

(**a**) ATPase activity using (left-hand two lanes) γ-^{32}P- and (right-hand two lanes) α-^{32}P-labelled ATP as substrates. (**b**) NTPase activity using different α-^{32}P-labelled NTPs as substrates. (**c**) Effect of varying concentrations of ATP on ATPase activity (top panel) and the reaction rates quantified in a PhosphorImager and plotted as a function of substrate concentration (lower panel).

Functional characterization of LEF4

LEF4 from AcMNPV, with which the BmNPV counterpart shares substantial homology, has been shown to possess mRNA-capping activity [23–25]. Further, since the molecular modelling of BmLEF4 suggested its similarity to the mRNA-capping enzymes, the capping activity of the purified LEF4 (bacterially expressed and GST-tagged) has been analysed in detail. mRNA capping comprises different biochemical reactions, i.e. RNA triphosphatase, in which the 5′-phosphate-terminated primary transcript is converted to a diphosphate-terminated RNA, GTP-RNA guanylate transferase, in which the GMP is transferred to the diphosphate-terminated RNA in a 5′–5′ linkage, and NTP hydrolysis. All these enzymic activities were tested with purified BmLEF4.

Characterization of guanylate transferase activity of LEF4

The guanylate transferase component of the enzyme catalyses two sequential nucleotidyl transfer reactions involving a covalent enzyme–guanylate intermediate. The nucleotide is then transferred to the 5′ end of the acceptor RNA (diphosphate-terminated) to form an inverted 5′–5′ triphosphate bridge structure. LEF4 from BmNPV catalysed both. The first partial reaction, i.e. the transfer of radiolabel from [α-^{32}P]GTP to the protein leading to the formation of labelled LEF4, is presented in Figure 2(a). The size and location of the purified GST-LEF4 (Figure 2a, lane 2), and the presence of labelled LEF4 following the enzyme reaction (Figure 2a, lane 3) are shown. Appropriate controls where the reaction was terminated at time 0 (Figure 2a, lane 4), or where only GST was used as substrate instead of GST-LEF4 (Figure 2a, lane 5) are also presented. The formation of labelled LEF4–GMP was dependent on the concentration of the LEF4 protein added (Figures 2b and 2c).

Characterization of RNA triphosphatase activity of LEF4

RNA triphosphatase catalyses the first step of mRNA cap formation, i.e. the hydrolysis of the 5′-γ-phosphate of the

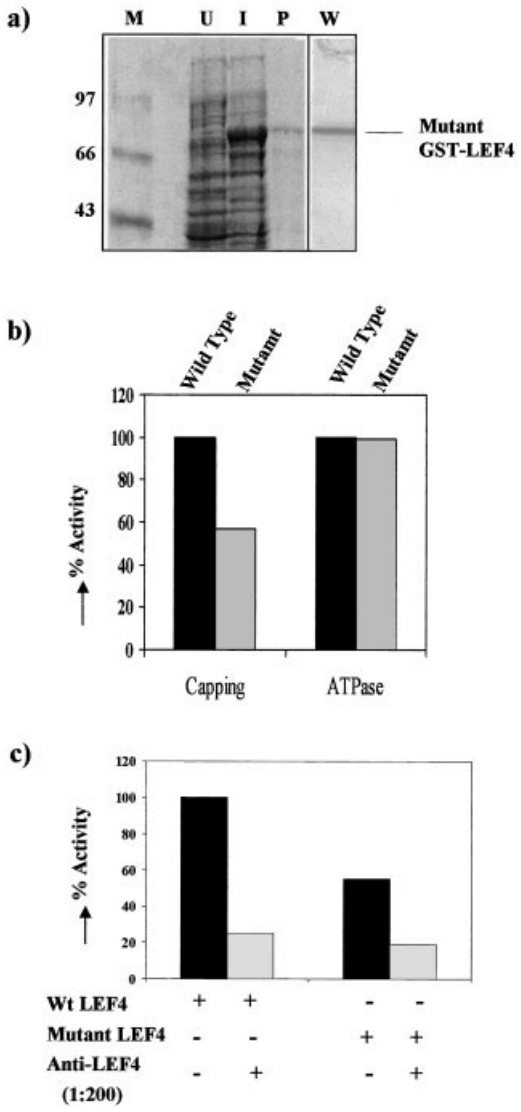


Figure 6 Mutational analysis of BmLEF4

(**a**) Bacterial expression of mutant protein. (**b**) Effect of mutation (Tyr to Cys) on the ATPase and RNA-capping activities. (**c**) Effect of anti-LEF4 antibodies on the capping activity of mutant LEF4. (**a**) The mutation Tyr to Cys in the conserved domain IIIa (indicated in Figure 1a) was generated by megaprime PCR-based mutagenesis (see text for details). The mutant protein expressed as GST fusion in bacteria was solubilized, purified by binding to the affinity matrix GST–Sepharose and confirmed by Western blot using anti-LEF4 antibodies. M, molecular size markers; U, uninduced bacterial extract; I, induced bacterial extract; P, mutant BmLEF4 after affinity purification on GST–Sepharose. All these lanes show staining patterns with Coomassie Blue; W, Western blot of the purified mutant LEF4 protein (same as in lane P). (**b**) RNA-capping activity was quantified from the radiolabel transfer from [α-^{32}P]GTP to RNA, in a liquid scintillation counter (100% activity corresponds to 38000–40000 c.p.m. incorporated). ATPase activity was assayed by monitoring the hydrolysis of [α-^{32}P]ATP on TLC plates.

triphosphate-terminated pre-mRNA to form a diphosphate end, which serves as the acceptor for 5′ capping. This activity (measured as the amount of P_i released) was estimated indirectly from the radioactivity remaining on the labelled RNA after addition of BmLEF4. The amount of P_i released from 5′-γ-phosphate-end-labelled RNA was proportional to the concentration of LEF4 (Figure 3a). The reaction was saturated by 80–100 nmol of LEF4. The kinetic analysis of the RNA triphosphatase activity showed that the reaction rates were linear up to 45 min. The optimal activity was achieved at 2 mM Mg^{2+}, which declined at higher concentrations (5 mM or more) of the cation (Figure 3b). Only 30% activity was seen in the presence of Mn^{2+} (results not shown).

Capping activity of BmLEF4

The final step in capping activity was monitored by the transfer of radioactive label from [α-^{32}P]GTP to the RNA. BmLEF4 also showed the transfer of radioactive label from [α-^{32}P]GTP to the RNA (Figure 4a). Since the isotopic label from [α-^{32}P]GTP was also transferred to the LEF4 protein (which is an intermediate in the capping reaction; see the previous reaction), it was necessary to remove the GMP-bound protein by phenol extraction, prior to monitoring the incorporation of radiolabel into the RNA. The RNA-capping activity was inhibited by the presence of anti-LEF4 antibodies (Figure 4b). More than 80% of inhibition of activity was seen at 1:100 dilution of the antibody. There was no inhibition of activity by non-immune serum, which was included as control. The kinetic parameters of the capping activity were determined (three independent experiments) by varying the concentration of substrate (GTP) and from a double-reciprocal plot; the K_m for GTP was calculated to be 330 ± 20 μM (Figure 4c).

Characterization of NTPase activity of LEF4

The GTP-dependent capping enzymes and ATP-dependent ligases make up a superfamily of covalent nucleotidyl transferases. BmLEF4 showed ATP-hydrolysing activity. The ATPase activity was confirmed by using [γ-^{32}P]ATP and [α-^{32}P]ATP as substrates, from which [^{32}P]P_i or [^{32}P]ADP were generated, respectively (Figure 5a). The NTPase activity of LEF4 was also examined using different NTPs as substrates. As seen in Figure 5(b), ATP and GTP were hydrolysed to the same extent, followed by CTP and UTP. The conversion of ATP to ADP was proportional to the concentration of LEF4 and was linear up to 40–60 min. The ATPase activity of LEF4 showed a specific requirement for Mn^{2+} with maximal activity at 1 mM, whereas Mg^{2+} showed no activity at this pH. The kinetic analysis of ATPase as a function of varying ATP concentration in a typical experiment is presented in Figure 5(c). The reaction velocity had a sigmoid dependence on the concentration of ATP (Figure 5c, lower panel) and two binding sites for ATP on LEF4 were calculated from the Hill plot.

Mutational analysis of LEF4

A point mutation was generated in the LEF4 ORF, converting a conserved Tyr to Cys in the C-terminal region showing homology to the capping enzyme domain IIIa (marked in Figure 1a). The mutant protein was overexpressed and purified as a GST fusion protein (Figure 6a), similar to the parental LEF4. The bulk of the bacterially expressed mutant LEF4 was also present in the insoluble fraction and therefore a similar method as used for solubilization and purification of the wild-type protein was adopted. The mutation had no effect on the ATPase activity, whereas the capping activity was reduced to 56% (Figure 6b). The capping activity of the mutant LEF4 was also inhibited by LEF4 antibodies to almost the same extent as the wild-type protein (Figure 6c).

The last step associated with the capping function, i.e. methyl group transfer activity, could not be detected with LEF4 when

[methyl-^{14}C]*S*-adenosylmethionine was included as a substrate in the capping reaction.

DISCUSSION

Almost all eukaryotic mRNAs are capped at their 5′ termini. The m7GpppN cap structure of eukaryotic mRNA is formed by the sequential action of three enzymes: RNA triphosphatase, RNA guanylate transferase and RNA (guanine-7)-methyltransferase. The LEF4 protein from AcMNPV carries out the mRNA-capping reaction [23–25]. BmLEF4 showed 97% homology with its AcMNPV counterpart. In addition, the signature motifs of mRNA-capping enzymes were discernible in the molecular model predicted here (Figure 1b). In agreement, BmLEF4 showed all the catalytic functions, i.e. RNA triphosphatase, guanylate transferase and RNA guanylate transferase activities. In fact, the transfer of GTP to the RNA 5′ end was not shown in previous studies on AcMNPV. The protein also showed NTPase activity. The GTP-dependent capping enzymes and ATP-dependent ligases make up a superfamily of covalent nucleotidyltransferases [41,42].

Notwithstanding the remote sequence similarity between BmLEF4 and other capping enzymes, the compatibility of the BmLEF4 sequence to the three-dimensional fold of capping enzymes of known structures could be demonstrated. *Chlorella* virus PBCV-1 guanylate transferase, a 38 kDa protein, is the smallest guanylate transferase reported to date and no triphosphatase activity is associated with this protein [42]. Vaccinia virus mRNA-capping enzyme, on the other hand, is a multifunctional protein with RNA triphosphatase, RNA guanylate transferase and RNA (guanine-7)-methyltransferase activities [43,44]. We could not detect any methyltransferase activity associated with BmLEF4. The mammalian capping enzyme, a single polypeptide having both triphosphatase and guanylate transferase activities, also do not show any methyltransferase activity [42]. In trypanosomes, mRNA capping occurs post-transcriptionally by *trans* splicing, where the capped spliced leader sequence is transferred to the 5′ ends of all mRNAs [45].

Transguanylation is a two-stage reaction involving a covalent enzyme–GMP intermediate [46]. The active site is composed of six protein motifs that are conserved in terms of order and spacing among yeast and DNA virus capping enzymes. Both RNA triphosphatase and RNA guanylate transferase activities were associated with BmLEF4. As seen with other RNA triphosphatases from DNA viruses (poxviruses and baculoviruses) and *S. cerevisiae*, the activity of BmLEF4 was dependent on $MgCl_2$ whereas $MnCl_2$ showed only 33% activity. In contrast, the NTPase activity was exclusively dependent on Mn^{2+}, and Mg^{2+} or Ca^{2+} failed to support the reaction at the pH used (pH 7.5). The cation requirements for these reaction were similar to those seen for the AcMNPV counterpart and even in this case, there was no Mg^{2+}-dependent activity at pH 7.5 but significant activity was seen at pH 9.0 [47]. The ATPase activity of BmLEF4 depended on ATP concentration in a sigmoid manner and two active sites were calculated from the Hill plot. The capping (RNA guanylate transferase) activity, however, showed a typical hyperbolic response and, based on the kinetic patterns, there appeared to be one binding site for GTP in the capping reaction (from two independent determinations). These results are consistent with the prediction of separate domains for the triphosphatase and capping activities [47] and with the results of mutational analysis.

The guanylate transferase activity of AcMNPV LEF4 has been shown to be only at the level of radioactive label transfer from labelled GTP to the LEF4 protein, which is an intermediate in the total reaction. Here we have demonstrated both steps, i.e. transfer of label from [α-^{32}P]GTP to the protein as well as to the RNA substrate. Moreover, antibodies raised against the bacterially expressed protein also showed inhibitory effects on all the biochemical reactions of LEF4. Thus all the steps characteristic of the capping mechanism have been demonstrated for the BmLEF4. The deletion of C-terminal domain of LEF4 resulted in the loss of capping activity (results not shown).

In the three-dimensional model constructed here most of the conserved residues were located in and around the inter-domain interface, which contains the essential functional region. A number of conserved hydrophobic residues, Val-79, Phe-146, Ile-216, Leu-215 and Ile-217, were buried in the structure and were well packed with other hydrophobic residues. Only some of these residues were close to the inter-domain interface. For example, Tyr in motif IIIa of the model is equivalent to Phe-146 of 1CKO and is 9 Å (where 1 Å = 0.1 nm) away from the KXDG functional motif. As this residue is well buried and packed, a disruption of this residue is likely to affect markedly the structure of the protein and, consequently, its function. We have explored this deduction by site-directed mutagenesis of a Tyr residue in motif IIIa located at the middle of the β-strand involved in the β-sheet (designated as β9 in Figures 1b and 1c). Mutation of this conserved Tyr residue to Cys reduced the capping activity to 56% but had little effect on the ATPase function. Interestingly, ten contiguous residues in this β-strand, centred about the Tyr of interest, were buried by the rest of the structure, suggesting that this β-strand is an integral component of the fold adopted by the protein. Reduction in the capping activity without any noticeable change in ATPase function in the mutant indicated that the alteration in conformation and exposure of the residues involved in capping activity was not profound enough to disrupt the ATPase activity. However, in the absence of crystal or NMR structures for such mutants of capping enzymes, it is difficult to predict the exact changes in the structure of the mutant enzyme.

We thank the Department of Biotechnology, Government of India, and the Council of Scientific and Industrial Research for financial support, and our Institute's Divisional Facilities for DNA sequencing and PhosphorImaging. N.S. is supported by an International Senior Fellowship Award in Biomedical Sciences by the Wellcome Trust, U.K.

REFERENCES

1 Maeda, S., Kawai, T., Obinata, M., Fujiwara, H., Horiuchi, T., Saeki, Y., Sato, Y. and Furusawa, M. (1985) Production of human α-interferon in silkworm using a baculovirus vector. Nature (London) **315**, 592–594

2 Miyajima, K., Schreurs, J., Otsu, K., Kondo, A., Arai, K. and Maeda, S. (1987) Use of the silkworm, *Bombyx mori*, and an insect baculovirus vector for high-level expression and secretion of biologically active mouse interleukin-3. Gene **58**, 273–281

3 Maeda, S. (1989) Expression of foreign genes in insects using baculovirus vectors. Annu. Rev. Entomol. **34**, 351–372

4 Palhan, V. B., Sumathy, S. and Gopinathan, K. P. (1995) Baculovirus mediated high level expression of luciferase in silkworm cells and larvae. BioTechniques **19**, 97–105

5 Sehgal, D. and Gopinathan, K. P. (1998) Recombinant *Bombyx mori* nucleopolyhedrovirus harboring green fluorescent protein. BioTechniques **25**, 997–1000

6 Acharya, A., Sriram, S., Sehrawat, S., Rahman, M., Sehgal, D. and Gopinathan, K. P. (2002) *Bombyx mori* nucleopolyhedrosis virus: molecular biology and biotechnological applications for large scale synthesis of recombinant proteins. Curr. Sci. **83**, 455–465

7 Gomi, S., Majima, K. and Maeda, S. (1999) Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. J. Gen. Virol. **80**, 1323–1337

8 Glocker, B., Hoopes, Jr, R. R., Hodges, L. and Rohrmann, G. F. (1993) *In vitro* transcription from baculovirus late gene promoters: accurate mRNA initiation by nuclear extracts prepared from infected *Spodoptera frugiperda* cells. J. Virol. **67**, 3771–3776

9 Huh, N. E. and Weaver, R. F. (1990) Identifying the RNA polymerase that synthesize specific transcripts of *Autographa californica* nuclear polyhedrosis virus. J. Gen. Virol. **71**, 195–199

10 Lu, A. and Miller, L. K. (1995) Differential requirements for baculovirus late expression factor genes in two cell lines. J. Virol. **69**, 6265–6272

11 Todd, J. W., Passarelli, A. L. and Miller, L. K. (1995) Eighteen baculovirus genes, including *lef*11, *p35*, *39K* and *p47*, support late gene expression. J. Virol. **69**, 968–974

12 Rapp, J. C., Wilson, J. A. and Miller, L. K. (1998) Nineteen baculovirus open reading frames, including LEF12, support late gene expression. J. Virol. **72**, 10197–10206

13 Sriram, S. and Gopinathan, K. P. (1998) A potential role of late gene expression factor *Lef*2 from *Bombyx mori* nuclear polyhedrosis virus in very late gene transcription and DNA replication. Virology **251**, 108–122

14 Kool, M., Ahrens, C. M., Goldbach, R. W., Rohrmann, G. F. and Vlak, J. M. (1994) Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. Proc. Natl. Acad. Sci. U.S.A. **91**, 11212–11216

15 Merrington, C. L., Kitts, P. A., King, L. A. and Possee, R. D. (1996) An *Autographa californica* nucleopolyhedrovirus lef-2 mutant: consequences for DNA replication and very late gene expression. Virology **217**, 338–348

16 Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita, T., Sano, Y. and Matsumoto, T. (2000) Sequence analysis of the *Plutella xylostella* granulovirus genome. Virology **275**, 358–372

17 Kuzio, J., Pearson, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M. and Rohrmann, G. F. (1999) Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. Virology **253**, 17–34

18 Fuchs, L. Y., Woods, M. S. and Weaver, R. F. (1983) Viral transcription during AcNPV infection: a novel RNA polymerase induced in infected *Spodoptera frugiperda* cells. J. Virol. **48**, 641–646

19 Beniya, H., Funk, C. J., Rohrmann, G. F. and Weaver, R. F. (1996) Purification of a virus-induced RNA polymerase from *Autographa californica* nuclear polyhedrosis virus-infected *Spodoptera frugiperda* cells that accurately initiates late and very late transcription *in vitro*. Virology **216**, 12–19

20 Guarino, L. A., Xu, B., Jin, J. and Dong, W. (1998) A virus encoded RNA polymerase purified from baculovirus-infected cells. J. Virol. **72**, 7985–7991

21 Lu, A. and Miller, L. K. (1994) Identification of three late expression factor genes within the 33.8- to 43.4-map-unit region of *Autographa californica* nuclear polyhedrosis virus. J. Virol. **68**, 6710–6718

22 Passarelli, A. L., Todd, J. W. and Miller, L. K. (1994) A baculovirus gene involved in late gene expression predicts a large polypeptide with conserved motif of RNA polymerases. J. Virol. **68**, 4673–4678

23 Guarino, L. A., Jin, J. and Dong, W. (1998) Guanyltransferase activity of the LEF-4 subunit of baculovirus RNA polymerase. J. Virol. **72**, 10003–10010

24 Gross, C. H. and Shuman, S. (1998) RNA 5′-triphosphatase, nucleoside triphosphatase and guanylyltransferase activities of baculovirus LEF-4 protein. J. Virol. **72**, 10020–10028

25 Jin, J., Dong, W. and Guarino, L. A. (1998) The LEF-4 subunit of baculovirus RNA polymerase has 5′ RNA triphosphatase and ATPase activities. J. Virol. **72**, 10011–10019

26 Palhan, V. B. and Gopinathan, K. P. (1996) Characterization of a local isolate of *Bombyx mori* nuclear polyhedrosis virus. Curr. Sci. **70**, 147–153

27 O'Reilly, D. R., Miller, L. K. and Lukow, V. E. (1992) The Baculovirus Expression Vectors: a Laboratory Guide, Chapman and Hall, London

28 Jones, D. T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. **287**, 797–815

29 Fischer, D., Rice, D., Bowie, J. U. and Eisenberg, D. (1996) Assigning amino acid sequences to 3-dimensional protein folds. FASEB J. **10**, 126–136

30 Kelly, L. A., MacCallum, R. M. and Sternberg, M. J. E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. **299**, 499–520

31 Sutcliffe, M. J., Haneef, I., Carney, D. and Blundell, T. L. (1987) Knowledge based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. Protein Eng. **1**, 377–384

32 Sutcliffe, M. J., Hayes, F. R. and Blundell, T. L. (1987) Knowledge based modelling of homologous proteins, part II: rules for the conformations of substituted sidechains. Protein Eng. **1**, 385–392

33 Srinivasan, N. and Blundell, T. L. (1993) An evaluation of the performance of an automated procedure for comparative modeling of protein tertiary structure. Protein Eng. **6**, 501–512

34 Srinivasan, N., Bax, B., Blundell, T. L. and Parker, P. J. (1996) Structural aspects of the functional modules in human protein kinase-Cα deduced from comparative analyses. Proteins **26**, 217–235

35 Topham, C. M., McLeod, A., Eisenmenger, F., Overington, J. P., Johnson, M. S. and Blundell, T. L. (1993) Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables. J. Mol. Biol. **229**, 194–220

36 Weiner, S. I., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) A new force field for molecular mechanical stimulation of nucleic acids and proteins. J. Am. Chem. Soc. **106**, 765–784

37 Evans, S. V. (1993) SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. J. Mol. Graph. **11**, 134–138

38 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283–291

39 Cong, P. and Shuman, S. (1995) Mutational analysis of mRNA capping enzyme identifies amino acids involved in GTP binding, enzyme-guanylate formation, and GMP transfer to RNA. Mol. Cell. Biol. **15**, 6222–6231

40 Hakansson, K. and Wigley, D. B. (1998) Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. Proc. Natl. Acad. Sci. U.S.A. **95**, 1505–1510

41 Shuman, S. and Schwer, B. (1995) RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. Mol. Microbiol. **17**, 405–410

42 Wang, S. P., Deng, L., Ho, C. K. and Shuman, S. (1997) Phylogeny of mRNA capping enzymes. Proc. Natl. Acad. Sci. U.S.A. **94**, 9573–9578

43 Venkatesan, S., Gershowitz, A. and Moss, B. (1980) Modification of the 5′ end of mRNA. Association of RNA triphosphatase with the RAN guanylyltransferase-RNA (guanine-7-)methyl transferase complex from vaccinia virus. J. Biol. Chem. **253**, 903–908

44 Shuman, S., Surks, M., Furneaux, H. and Hurwitz, J. (1980) Purification and characterization of a GTP-pyrophosphatase exchange activity from vaccinia virions. Association of the GTP-pyrophosphate exchange activity with vaccinia virus mRNA guanylyl transferase-RNA (guanine-7-)methyl transferase complex (capping enzyme). J. Biol. Chem. **255**, 11588–11598

45 Silva, E., Ullu, E., Kobayashi, R. and Tschudi, C. (1998) Trypanosome capping enzymes display a novel two-domain structure. Mol. Cell. Biol. **18**, 4612–4619

46 Shuman, S. and Hurwitz, J. (1981) Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme-guanylate intermediate. Proc. Natl. Acad. Sci. U.S.A. **78**, 187–191

47 Martins, A. and Shuman, S. (2001) Mutational analysis of baculovirus capping enzynme LEF4 delineates an autonomous triphosphatase domain and structural determinants of divalent cation specificity. J. Biol. Chem. **276**, 45522–45529

48 Mizuguchi, K., Deane, C. M., Blundell, T. L., Johnson, M. S. and Overington, J. P. (1998) JOY: protein sequence-structure representation and analysis. Bioinformatics **14**, 617–623

49 Nicholls, A., Sharp, K. A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins **11**, 281–296

Received 22 March 2002/5 August 2002; accepted 8 August 2002
Published as BJ Immediate Publication 8 August 2002, DOI 10.1042/BJ20020456