Original Paper

Journal of Biomedical Science

J Biomed Sci 2001;8:170–175

Received: September 25, 2000
Accepted: November 14, 2000

Construction of a Tagging System for Subcellular Localization of Proteins Encoded by Open Reading Frames

Chen-Hua Chuang Sheng-Chieh Hsu Chia-Ling Hsu Tsui-Chi Hsu
Wan-Jr Syu

Institute of Microbiology and Immunology, National Yang Ming University, Taipei, Taiwan, ROC

Key Words
Monoclonal antibody · Epitope · Tag · Immunofluorescence

Abstract
We have previously characterized a monoclonal antibody (SC1D7) that is directed to maltose-binding protein (MBP) of *Escherichia coli* and other closely related enteric bacteria. SC1D7 does not cross-react with proteins in eucaryotes and appears to be a highly specific tool in immunochemical analyses. To better map the epitope, we took advantage of an available plasmid, pMAL-c2, that encodes the *E. coli* MBP-coding sequence and constructed plasmids to express MBP fragments. A construct containing the N-terminal portion of MBP does not react with SC1D7, whereas a second construct expressing glutathione S-transferase fused with the C-terminal half of MBP does react with SC1D7. To precisely define the epitope, random peptides displayed on M13 were used to react with SC1D7. Sequences of reactive peptides were aligned, and a consensus sequence of XDXRIPX was deduced. This sequence matches MBP with an amino acid stretch of KDPRIAA. To consolidate the mapping result, a sequence encoding this epitope was inserted into an expression vector and the resulting recombinant protein did react with SC1D7. Thereafter, this epitope was incorporated into a eucaryotic expression plasmid containing a previously defined hepatitis delta virus epitope for protein tagging. This two-epitope-tagging vector is useful in various molecular analyses. We demonstrate its usage for localization of a bacterial virulence factor in host cells. This vector should be applicable for high-throughput characterization of new open reading frames found in genome sequencing.

Introduction

Fast characterization of open reading frames (ORFs) is increasingly demanded in functional genomics. The currently available vectors that carry a tag at the N- or C-terminus of the expressed proteins have greatly facilitated the detection of targeted proteins. These tags, exemplified by the Myc-tag [5, 6, 13], His-tag [2, 21], FLAG-tag [9, 14], HA-tag [3], EBNA-1 tag [1] and T7-tag [4, 15, 18], can be detected by one or more monoclonal antibodies (MAb). These epitope-tagged proteins can be quickly analyzed before ORF-specific antibodies become available. One example of these analyses is the high-throughput subcellular localization of target proteins by immunostaining [10]. However, proteins labeled with tags at the N-terminus may encounter degradation, and the result of localization by staining the tag may be complicated by observing the products without the C-terminus. Thus, a system with

KARGER
Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7770/01/0082–0170$17.50/0
Accessible online at:
www.karger.com/journals/jbs

W.-J. Syu
Institute of Microbiology and Immunology
National Yang Ming University, Shih-Pai
Taipei 112, Taiwan (ROC)
Tel. +886 2 2826 7112, Fax +886 2 2821 2880, E-Mail wjsyu@ym.edu.tw

a C-terminal tag may assure the detection of the full-length product. However, another concern may arise from the C-terminal tagging system. It is possible to observe products translated from internal initiation codons of the given ORFs. Therefore, simultaneous labeling of ORFs with tags at both ends and collecting data derived from the two reporting tags may provide a better analysis.

We previously characterized a mouse MAb (HP6A1) that recognizes the human hepatitis delta antigen (HDAg) in a strain-specific manner [7]. The epitope defined by this MAb is located at the N-terminal residues 4–10 of HDAg (strain 25). HDAg is a protein exogenous to mammalian cells, and HP6A1 reacts with no cellular proteins as revealed by immunoprecipitation, Western blotting and immunocytostaining. Therefore, the epitope defined by HP6A1 is a good expression tag, and so an additional epitope that could be simultaneously used was sought.

MAb SC1D7 recognizes maltose-binding proteins of enteric bacteria. It does not cross-react with bacterial proteins of distantly related species [8]. Neither does it cross-react with proteins extracted from eucaryotic cells. Therefore, we mapped the binding site of SC1D7 and inserted the coding sequence into plasmids that have a previously engineered HP6A1-defined tag. As a result, ORFs of interest cloned into these plasmids could be flanked by two different epitopes. Utilization of this system was demonstrated on subcellular localization of proteins of interest.

Materials and Methods

Plasmid Construction

Plasmid pMAL-c2 (New England Biolabs, Beverly, Mass.) contains an ORF coding for 462 amino acids. For easy discussion, this full-length ORF is named MBP(F). To generate fragments of MBP(F), pMAL-c2 was digested with *Bgl*II and *Bam*HI. After removing a 0.8-kb *Bgl*II-*Bam*HI fragment, the remaining portion of the plasmid was self-ligated to generate pMAL-N. The resulting ORF, designated as MBP(N), was reduced to 180 residues containing the N-terminal 120-residue fragment of MBP fused with 58 residues derived from the vector. In parallel, pMAL-c2 was digested with *Bgl*II and *Eco*RI, and a 0.8-kb *Bgl*II/*Eco*RI fragment was isolated. This fragment was ligated with *Bam*HI/*Eco*RI-cut pGEX-1 (Pharmacia, Piscataway, NJ) to result in plasmid pGEX-MAL(C) that encodes a C-terminal 280-residue fragment of MBP fused to the C-terminus of glutathione S-transferase; this 53-kd recombinant protein was designated as GST-MBP(C).

Pin-M was constructed by inserting an adapter (generated by the annealing of primers 5′TGAAGGATTCGCGGATTCCGACTTAGGTAC3′ and 5′CTAAGTCGGAATCCGCGAATCCTTCA3′) into the *Puv*II- and *Kpn*I-digested plasmid PinPoint-Xa-1 (Promega, Madison, Wisc.). Plasmids pCMV-DD2 and pCMV-DD6 were modified from pCMV-DS [7] with a similar strategy. All resulting plasmids were examined for correct restriction enzyme digestion patterns, and the ligated junctions were confirmed by DNA sequencing.

Epitope Mapping with M13 Phage Peptide Display

MAb preparations have been described previously [8]. Antibodies were affinity-purified with a protein G-Sepharose column (Pierce, Rockford, Ill.) according to the manufacturer's instructions. The purity of the antibodies was confirmed by sodium dodecyl sulfate-containing polyacrylamide gel electrophoresis (SDS-PAGE). M13 phages displaying random heptapeptides at the N-terminus of its minor coat protein pIII were obtained from New England Biolabs. Antibody coating and phage panning were performed as previously described [7]. The antibody-bound phages were determined for the 5′-end nucleotides of gene III by manual sequencing.

Western Blotting

Proteins were separated on SDS-PAGE, transferred to nitrocellulose membranes, reacted with immunochemical reagents as described previously [19]. In brief, rabbit antibodies were diluted with 5% skimmed milk in Tris-buffered saline and incubated with the nitrocellulose membranes. After incubation and washing, the blots were reacted with horseradish peroxidase (HRP)-labeled goat anti-rabbit IgG (heavy- and light-chain-specific; Sigma, St. Louis, Mo.), and finally developed with Western blot chemiluminescence reagent (NEN Life Science, Boston, Mass.).

When primary antibodies were MAb, the immunochemical reactions were carried out similarly except that the secondary antibodies used were HRP-conjugated goat anti-mouse IgG (Sigma).

Cell Transfection and Immunofluorescence Staining

HeLa cells in DMEM-10% fetal calf serum were maintained in 145-mm tissue culture dishes. Transient expression of antigens was carried out by calcium chloride transfection as previously described [11]. Cells used for immunofluorescence staining were similarly prepared except that the cells were cultured on cover slides. Cells were rinsed with phosphate-buffered saline (PBS), fixed, and permeated with methanol/acetone (1:1) 48 h after transfection. After blocking nonspecific binding with 1% bovine serum albumin (BSA) in PBS, the cells were reacted with primary antibodies diluted 1 to 200 in the same buffer. The secondary antibodies used were rhodamine-labeled goat anti-rabbit IgG or FITC-labeled goat anti-mouse IgG (Sigma). These fluorochrome-conjugated secondary antibodies were species specific and did not cause cross-reaction. Both conjugates were diluted 1 to 500 in PBS containing 1% BSA.

Results and Discussion

To map where SC1D7 binds to MBP, we generated two recombinant proteins containing different MBP portions (fig. 1a). MBP(N) contains the N-terminal 120 amino acids of MBP, whereas GST-MBP(C) contains GST and residues 121–391 of MBP. These two recombinant proteins appeared to be overexpressed in bacteria since in the Coomassie blue-stained protein profiles (fig. 1b), bands with the expected electrophoretical mobility could

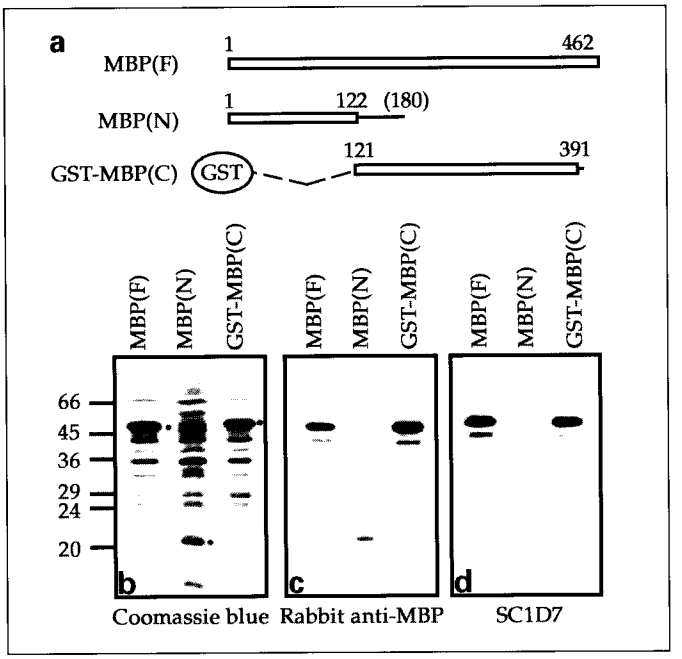

Fig. 1. Mapping the binding site of SC1D7 on MBP. **a** Schematic of MBP fragmentation. **b** Coomassie blue staining patterns of bacterial lysates run on SDS-polyacrylamide gel. Dots mark overexpressed proteins. Proteins in SDS gels parallel to that in **b** were transferred to nitrocellulose membranes and reacted with different antibodies: rabbit polyclonal antibodies to MBP (**c**), or MAb SC1D7 (**d**).

		314						320
MBP:		K	D	P	R	I	A	A
Consensus sequence:		X	D	X	R	V/I/L	P	X
M13 phage display	#1	P	Q	D	P	R	R	I
	#2	A	F	D	P	T	I	P
	#3	H	Q	D	P	R	L	P
	#4	Y	Y	D	P	R	I	A
	#5		F	D	L	R	V	P Q
	#6		K	D	V	R	I	A P
	#7		K	D	S	R	I	P T
	#8		F	D	L	R	V	P V
	#9		F	D	P	R	V	P L
	#10		G	D	A	R	V	P S
	#11		A	D	P	R	V	A K
	#12			D	L	R	V	P T K
	#13			D	W	R	I	P T L
	#14			D	L	R	V	P L S
	#15			D	L	R	I	P F S
	#16			D	L	R	V	P L S
	#17			D	Y	R	I	P I Y

Fig. 2. Epitope characterization using random peptides displayed on M13 phages. Phages bound on SC1D7 were randomly picked out and determined for the pIII sequence. The N-terminal residues of pIII were aligned, and identical residues or those appearing as conservative amino acid changes are indicated with bold letters. These residues are summarized as a consensus sequence (shown on top) where X denotes residues with no preferred biophysical property. A sequence best matched to the consensus sequence in MBP (residues 314–320) is also aligned for comparison.

be visualized. To confirm that these proteins are the correct MBP derivatives, Western blotting using anti-MBP polyclonal antisera was performed (fig. 1c). The fact that all protein reacted positively with anti-MBP antibodies indicated that these proteins were properly assigned. These proteins were then reacted with SC1D7, and the Western blotting results (fig. 1d) indicated that GST-MBP(C) and MBP(F), but not MBP(N), were detected. Therefore, the binding site of SC1D7 was mapped to C-terminal residues 121–391 of MBP.

To better characterize the epitope, we tested the binding of SC1D7 on random heptapeptides displayed on M13. After panning 3 times, MAb-bound phages were eluted. From them, 17 phages were picked out, and the peptide sequences displayed on their pIII molecules were determined. These peptides were aligned, and the results yielded a consensus sequence, X-D-X-R-I/V/L-P-X (fig. 2), where X represents a residue with no preferred biophysical properties. The most notable residues in this sequence are Asp and Arg separated by a noncharged residue, an Ile exchangeable with Val/Leu, and a Pro substitutable with Ala. When this consensus sequence is aligned with that of the parental immunogen, the best-matched region was found in residues 314–319 of MBP. In this region, Asp315, Arg317 and Ile318 matched the consensus sequence, whereas Lys314, Pro316 and Ala319 were found twice or more in the deduced phage peptides. Therefore, these phage-mapping results are consistent with the data obtained from the MBP fragment-binding experiment.

The above X-D-X-R-I/V/L-P-X peptides are located at the N-terminus of pIII or in the middle of MBP. To test whether those sequences located at the C-terminus of a protein are still available for SC1D7 binding, we constructed Pin-M and expressed K-D-S-R-I-P-T at the protein C-terminus. This K-D-S-R-I-P-T peptide is the one displayed by phage 7 (fig. 2) and is named the M epitope. The M epitope followed by a translation stop codon was inserted into a site close to the C-terminus of Pin (the parental ORF encoded by the expression vector PinPoint-Xa-1) but not at the end. As a result, Pin has an expected size of 15.5 kd, whereas Pin-M is expected to be 14.3 kd. Both proteins should be detected by avidin-HRP because of a de novo biotinylated residue at Lys89 [20]. Host bac-

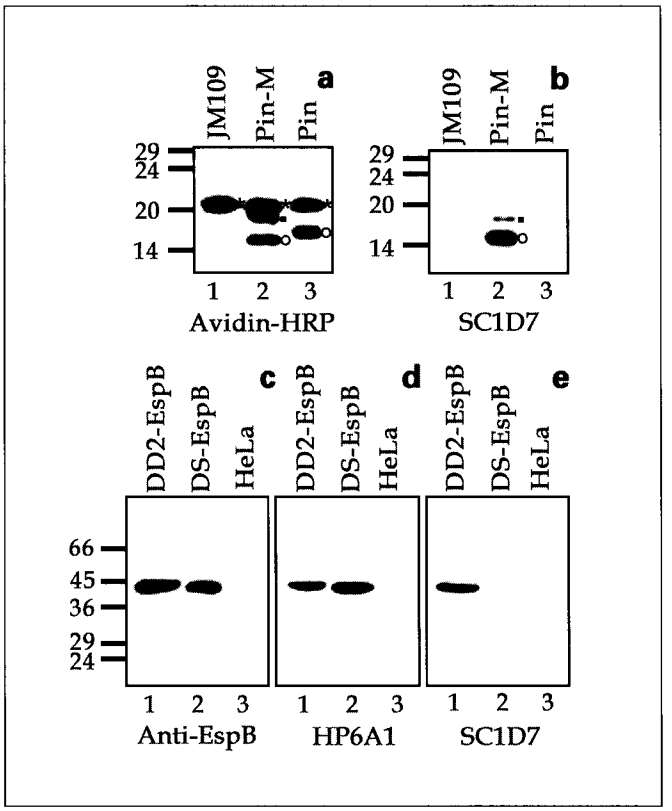

Fig. 3. Western blot analysis of proteins with the M epitope attached. The sequence coding for the M epitope was inserted into different expression plasmids (see Materials and Methods). Proteins from lysates of different bacteria were separated by SDS-PAGE and transferred onto nitrocellulose membranes for analysis with avidin-HRP (**a**) or SC1D7 (**b**). Proteins de novo biotinylated and detected by avidin-HRP are endogenous 22-kd host protein (asterisks), proteins encoded by the expression vectors (open circles), and proteins encoded by pPin-M but whose translation presumably terminates at the next stop codon 25 residues downstream (dots). **c–e** Analyses were carried out similarly to those described for **a** and **b** except that proteins were from lysates of plasmid-transfected HeLa cells that expressed different EspB constructs.

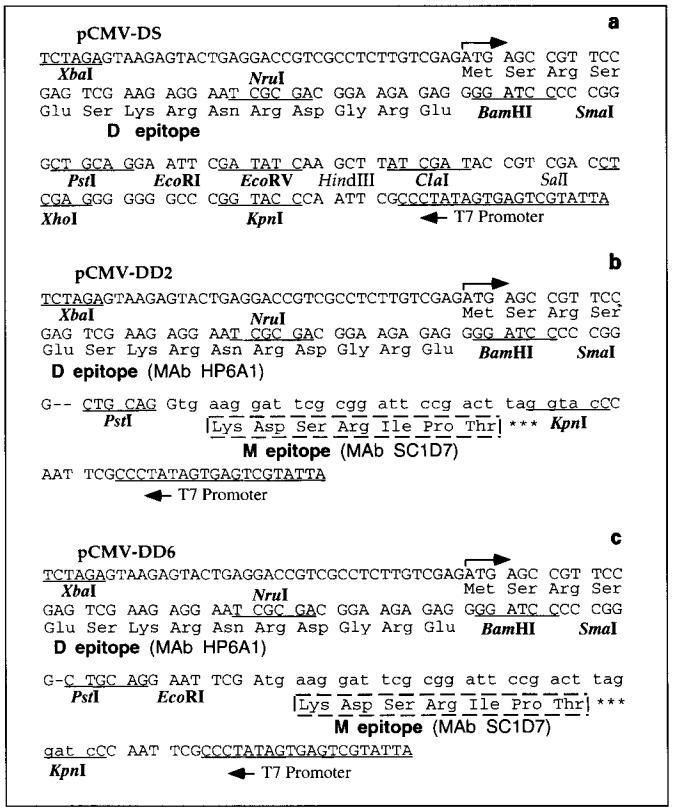

Fig. 4a–c. Cloning and expression regions of plasmids pCMV-DS, pCMV-DD2 and pCMV-DD6. Note that while the D epitope is in the same frame as the initiation codon (bent arrow), the M epitope is in the third and second frames in pCMV-DD2 and pCMV-DD6, respectively.

teria also have an intrinsic biotinylated protein at 22 kd. Figure 3a shows that lysates from host *Escherichia coli* JM109 and the other transformants all share a commom 22-kd avidin-HRP-detected protein. Thus, this 22-kd protein must be the host intrinsic biotinylated protein. The construct of Pin-M was readily detected as a 14.3-kd product (fig. 3a, lane 2), which is smaller than the Pin product (lane 3) expressed from the plasmid control. Figure 3b shows that the product of Pin-M (lane 2), but not Pin, was detected by SC1D7. Therefore, these data further confirmed that the M epitope is specifically recognized by SC1D7 and suggest that the M epitope is antibody-accessible when it is attached to the protein C-terminus.

In the sample of Pin-M, a third biotinylated protein was detected by both avidin-HRP (fig. 3a) and SC1D7 (fig. 3b). Its electrophoretic mobility is faster than that of the host intrinsic protein (22 kd) and slower than those of Pin-M (14.3 kd) and Pin (15.5 kd). This protein retains the M epitope and has a size (of about 17 kd) that is consistant with the product of Pin-M extending to the next stop codon. Presumably, this band could be a translational read-through product [12] of Pin-M.

The above data were obtained from different procaryotic expression systems. To demonstrate that the M epitope is an appropriate tag in the eucaryotic system, we constructed two plasmids derived from pCMV-DS (fig. 4a) that contains a 'D' epitope defined by MAb HP6A1 [7]. Plasmids pCMV-DD2 (fig. 4b) and pCMV-DD6 (fig. 4c) differ from pCMV-DS by small DNA segments inserted between the *Pst*I and *Kpn*I sites. These

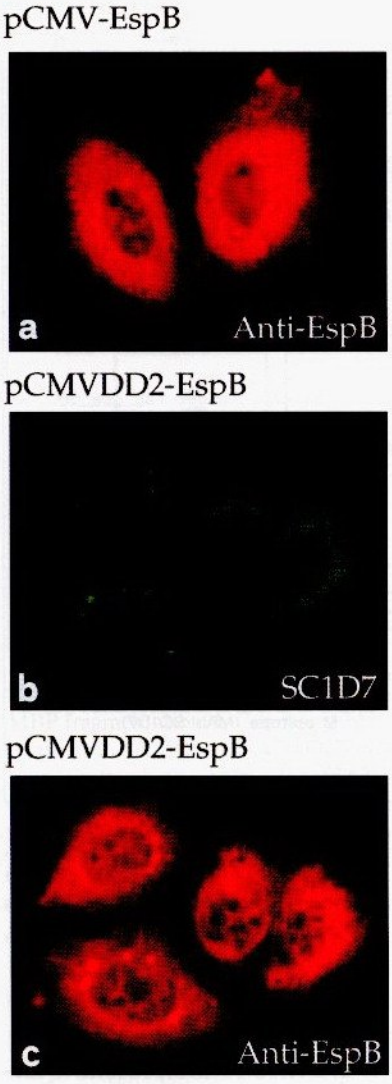

Fig. 5. Immunofluorescence analysis of EspB expressed in plasmid-transfected HeLa cells using different antibodies. HeLa cells were transfected to express EspB as indicated and immunostained 48 h after transfection. Cells were transfected with pCMV-EspB and stained with rabbit anti-EspB (**a**). Cells were transfected with pCMVDD2-EspB and doubly stained with SC1D7 (**b**) and rabbit anti-EspB (**c**). Rabbit antibodies were indirectly visualized with rhodamine-labeled goat anti-rabbit IgG, whereas MAb was detected with FITC-labeled goat anti-mouse IgG.

small fragments encode the M peptide at different reading frames. By appropriate insertion in front of the M peptide-coding region, the product of an ORF driven by a CMV promoter is possibly detected by both HP6A1 and SC1D7.

To test the feasibility of the above system, a virulence gene of enterohemorrhagic *E. coli*, *espB* [16, 17], was cloned into pCMV-DD2, with a frame correctly matched to the M peptide. The resulting plasmid pDD2-EspB was transfected into HeLa cells and analyzed for protein expression by Western blotting. The doubly tagged DD2-EspB product has an estimated size of 40 kd. In Western blotting reactions with rabbit anti-EspB (fig. 3c), HP6A1 (fig. 3d) and SC1D7 (fig. 3e), the cell lysates containing DD2-EspB (lane 1) all gave a single product of the expected size. In contrast, when EspB was cloned into pCMV-DS to generate DS-EspB, similarly prepared cell lysates had a protein detected by anti-EspB and HP6A1, but not by SC1D7 (lane 2, fig. 3c–e). These results are consistent with the expectation that DD2-EspB is labeled at both ends, whereas DS-EspB is only N-terminally tagged.

It is important that tags do not alter the properties of the questioned gene products. However, proteins vary among themselves, and it is difficult to address every single case. Since the original design of our system was to facilitate subcellular protein localization of unknown ORFs, we simply compared the results of the model protein, EspB, in the presence or absence of tags. The distribution patterns of EspB in cells transfected with pDS-EspB or pDD2-EspB were examined by immunofluorescence staining using anti-EspB, HP6A1 and SC1D7. In all cases observed, the staining pattern of the epitope-tagged EspB was similar to that without tags. Typical examples shown are EspB expressed from pCMV-EspB and stained with anti-EspB polyclonal antibodies (fig. 5a) and DD2-EspB expressed from pDD2-EspB and doubly stained with SC1D7 (fig. 5b) and rabbit anti-EspB (fig. 5c). These immunostaining results all suggest that EspB is mainly located in the cytoplasmic region, regardless of whether it is tagged or not. Thus, our data are consistent with the previous observation obtained by staining EspB in cells infected with *E. coli* [17] or in EspB-expressing cells [16]. These data also reveal that our tagging system does not adversely affect the destination of protein sorting. Therefore, this system appears to be an ideal alternative for tracking the subcellular localization of ORFs that have not been previously characterized.

Acknowledgments

We thank Dr. C.K. Chou, Dr. T.J. Chang and Dr. C.W. Chi for their continuous encouragement. This work was supported in part by grants 89-B-FA22-2-4 from the Education Department, ROC, NSC 89-2315-B-010-005 MH from the National Science Council, ROC, and VGH-89-385-4 from Veterans General Hospital Medical Research Advancement Foundation in memory of Dr. Chi-Shuen Tsou.

References

1 Chen MR, Huang H, Fen CH, Chen JY. A novel EBNA-1 tag system for high level expression and efficient detection of fusion proteins in vitro and in vivo. J Virol Methods 85:35–41; 2000.

2 Chen XS, Brash AR, Funk CD. Purification and characterization of recombinant histidine-tagged human platelet 12-lipoxygenase expressed in a baculovirus/insect cell system. Eur J Biochem 214:845–852;1993.

3 Chen YT, Holcomb C, Moore HP. Expression and localization of two low molecular weight GTP-binding proteins, Rab8 and Rab10, by epitope tag. Proc Natl Acad Sci USA 90:6508–6512;1993.

4 Enomoto S, Chen G, Berman J. Vectors for expressing T7 epitope- and His6 affinity-tagged fusion proteins in *Saccharomyces cerevisiae*. Biotechniques 24:782–788;1998.

5 Fan H, Villegas C, Chan AK, Wright JA. Myc-epitope tagged proteins detected with the 9E10 antibody in immunofluorescence and immunoprecipitation assays but not in Western blot analysis. Biochem Cell Biol 76:125–128;1998.

6 Gloor S, Pongs O, Schmalzing G. A vector for the synthesis of cRNAs encoding Myc epitope-tagged proteins in *Xenopus laevis* oocytes. Gene 160:213–217;1995.

7 Hsu SC, Lin HP, Wu JC, Ko KL, Sheen IJ, Yan BS, Chou CK, Syu WJ. Characterization of a strain-specific monoclonal antibody to hepatitis delta virus antigen. J Virol Methods 87:53–62;2000.

8 Hsu SC, Yan BS, Pan JM, Syu WJ. A monoclonal antibody reacts with maltose-binding protein of *Escherichia coli* and related enteric bacteria. J Immunol Methods 204:169–174; 1997.

9 Knappik A, Pluckthun A. An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. Biotechniques 17:754–761;1994.

10 Larsson M, Graslund S, Yuan L, Brundell E, Uhlen M, Hoog C, Stahl S. High-throughput protein expression of cDNA products as a tool in functional genomics. J Biotechnol 80:143–157;2000.

11 Lin HP, Hsu SC, Wu JC, Sheen IJ, Yan BS, Syu WJ. Localization of isoprenylated antigen of hepatitis delta virus by anti-farnesyl antibodies. J Gen Virol 80:91–96;1999.

12 MacBeath G, Kast P. UGA read-through artifacts – when popular gene expression systems need a pATCH. Biotechniques 24:789–794; 1998.

13 Nakajima K, Yaoita Y. Construction of multiple-epitope tag sequence by PCR for sensitive Western blot analysis. Nucleic Acids Res 25: 2231–2232;1997.

14 Slootstra JW, Kuperus D, Pluckthun A, Meloen RH. Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5. Mol Diversity 2:156–164;1997.

15 Tanaka S, Mikhailov M, Roy P. Synthesis of bluetongue virus chimeric VP3 molecules and their interactions with VP7 protein to assemble into virus core-like particles. Virology 214: 593–601;1995.

16 Taylor KA, Luther PW, Donnenberg MS. Expression of the EspB protein of enteropathogenic *Escherichia coli* within HeLa cells affects stress fibers and cellular morphology. Infect Immun 67:120–125;1999.

17 Taylor KA, O'Connell CB, Luther PW, Donnenberg MS. The EspB protein of enteropathogenic *Escherichia coli* is targeted to the cytoplasm of infected HeLa cells. Infect Immun 66: 5501–5507;1998.

18 Tseng CP, Verma AK. Functional expression and characterization of the mouse epitope tag-protein kinase C isoforms, alpha, beta I, beta II, gamma, delta and epsilon. Gene 169:287–288; 1996.

19 Yan BS, Lee KM, Liu SH, Syu WJ. Characterization of monoclonal antibodies to the 26-kDa glutathione S-transferase of *Schistosoma japonicum*. Hybridoma 15:429–433;1996.

20 Yan BS, Tam MH, Syu WJ. Self-association of the C-terminal domain of the hepatitis C virus core protein. Eur J Biochem 258:100–106; 1998.

21 Zhang F, Robbins DJ, Cobb MH, Goldsmith EJ. Crystallization and preliminary X-ray studies of extracellular signal-regulated kinase-2/MAP kinase with an incorporated His-tag. J Mol Biol 233:550–552;1993.