Biochemistry

 Volume 24, Number 22 October 22, 1985

Accelerated Publications

Clustering of Glycine and N^G,N^G-Dimethylarginine in Nucleolar Protein C23†

Michael A. Lischwe,*,‡ Richard G. Cook,§ Young Soo Ahn,‡ Lynn C. Yeoman,‡ and Harris Busch‡

Department of Pharmacology and Howard Hughes Medical Institute and Department of Microbiology and Immunology, Baylor College of Medicine, Houston, Texas 77030

Received July 24, 1985

ABSTRACT: Protein C23 (M_r 110 000, pI = 5.5), a major phosphoprotein in the nucleolus of mammalian cells, has been shown to contain 1.3 mol % of N^G,N^G-dimethylarginine (DMA) [Lischwe, M. A., Roberts, K. D., Yeoman, L. C., & Busch, H. (1982) *J. Biol. Chem. 257*, 14600–14602]. A tryptic peptide from protein C23 that contains DMA has been isolated and sequenced. Its sequence is Gly-Glu-Gly-Gly-Phe-Gly-Gly-*DMA*-Gly-Gly-Gly-*DMA*-Gly-Gly-Phe-Gly-Gly-*DMA*-Gly-Gly-Gly-*DMA*-Gly-Gly-*DMA*-Gly-Gly-Phe-Gly-Gly-*DMA*-Gly-*DMA*-Gly-Gly-Phe-Gly-Gly-*DMA*-Gly-Gly-Phe-*DMA*-Gly-Gly-*DMA*-Gly-Gly-Gly-Gly-Asp-Phe-Lys. This peptide contains 34 glycine, 10 DMA, and 6 phenylalanine residues and has clusters of glycine and N^G,N^G-dimethylarginine interspersed with phenylalanine residues. A similar domain has been found at the amino terminus of a nucleolar protein of M_r 34 000, pI = 8.5. This sequence array may represent a conserved domain characteristic of a certain class of nuclear proteins. All of the methylated arginine residues in protein C23, the 34-kilodalton protein, and myelin basic protein [Carnegie, P. R. (1971) *Biochem. J. 123*, 57–67] have at least one adjacent glycine. Access of certain arginine methylases to arginine residues may be sterically possible because of the lack of a side chain on the adjacent glycine residue(s).

Arginine methylation is catalyzed by a group of enzymes classified as protein methylase I [*S*-adenosyl-L-methionine: protein (arginine) *N*-methyltransferase, EC 2.1.1.23] (Paik & Kim, 1980). These enzymes can modify arginine residues to N^G-mono-, N^G,N'^G-di-, and N^G,N^G-dimethylarginines (Paik & Kim, 1980). The enzymes are primarily cytosolic and use *S*-adenosylmethionine as the methyl donor (Paik & Kim, 1980; Baldwin & Carnegie, 1971; Farooqui et al., 1985; Crang & Jacobson, 1982). The methylation of arginine residues in proteins appears to be an irreversible posttranslational modification (Paik & Kim, 1980; DesJardins & Morell, 1983). A methyltransferase modifies arginine-107 in human myelin basic protein to N^G-monomethylarginine or N^G,N'^G-dimethylarginine (Baldwin & Carnegie, 1971; Crang & Jacobson, 1982; Carnegie, 1971). It has been postulated that methylation of this residue aids in the transfer of this region of the protein into the nonpolar environment within myelin and in maintaining the integrity of myelin (Baldwin & Carnegie, 1971). The activity of this methylase increases as the myelin sheath is formed during development and has, accordingly, been designated a myelination-related enzyme (Crang & Jacobson, 1982).

The high-mobility groups 1 and 2 (Boffa et al., 1979), heterogeneous nuclear RNA (hnRNA)[1] binding proteins (Beyer et al., 1977; Karn et al., 1978; Thomas et al., 1983), a 34-kDa nucleolar protein (Lischwe et al., 1985), and nucleolar protein C23 (Lischwe et al., 1982) are nuclear proteins known to contain methylarginine. The 34-kDa (M_r 34 000, pI = 8.5) nucleolar protein is the most highly arginine-methylated protein found thus far in mammalian cells

†These studies were supported by Cancer Research Center Grant CA-10893, P1, awarded by the National Cancer Institute, Department of Health and Human Services, U.S. Public Health Service, Human Tumor Nucleolar Antigen Grant CA-27534, the DeBakey Medical Foundation, the Davidson Fund, the Pauline Sterne Wolff Memorial Foundation, the H. Leland Kaplan Cancer Research Endowment, the Linda and Ronnie Finger Cancer Research Endowment Fund, and the Sally Laird Hitchcock Fund.

‡Department of Pharmacology.

§Howard Hughes Medical Institute and Department of Microbiology and Immunology.

[1] Abbreviations: MMA, N^G-monomethylarginine; DMA, N^G,N^G-dimethylarginine; hnRNA, heterogeneous nuclear ribonucleic acid; PTH, phenylthiohydantoin; kDa, kilodalton; HPLC, high-performance liquid chromatography; RNP, ribonucleoprotein.

0006-2960/85/0424-6025$01.50/0

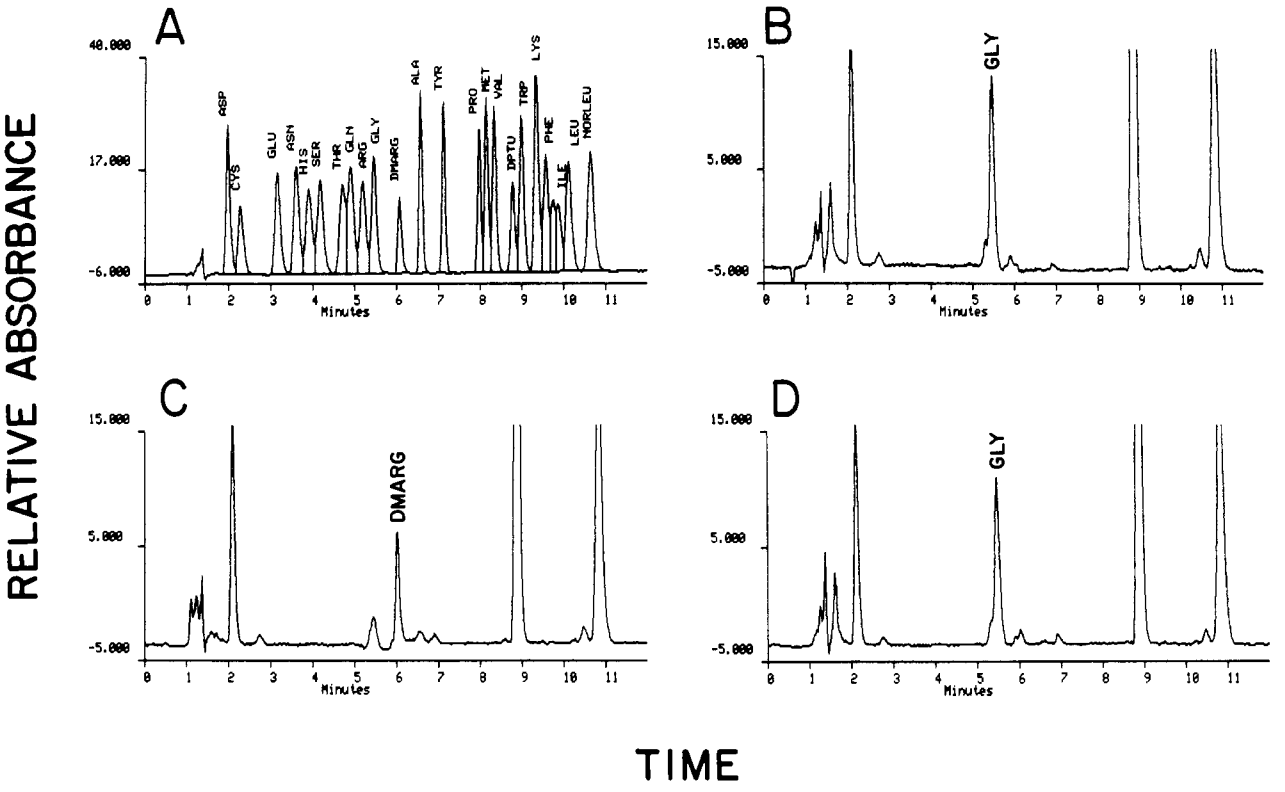

FIGURE 1: HPLC analysis of the PTH-amino acids. A tryptic peptide was purified from protein C23 and sequenced as described under Materials and Methods. The PTH-amino acids were identified and quantitated on a Waters HPLC system as described under Materials and Methods. The elution positions for the standards (250 pmol) including PTH-N^G,N^G-dimethylarginine are shown in panel A. The chromatographs obtained for steps 7–9 are shown in panels B–D. DMARG is N^G,N^G-dimethylarginine.

(Lischwe et al., 1985). The amino-terminal region of this 34-kDa protein contains Gly and DMA clusters interspersed with Phe residues (Lischwe et al., 1985). We report here that a similar domain exists in the nucleolar acidic phosphoprotein C23.

Materials and Methods

Nucleoli and Protein C23. Nucleoli were isolated from Novikoff hepatoma ascites cells as previously described (Rothblum et al., 1977). Protein C23 was isolated by DEAE-cellulose and Bio-Rad AG3-X4A column chromatography as reported (Lischwe et al., 1981).

Tryptic Peptides of Protein C23. Protein C23 (2.2 mg) was digested with trypsin (22 μg) for 4 h at 37 °C in 0.1 M NH_4HCO_3, pH 7.8. The reaction products were dried and suspended in 0.1% trifluoroacetic acid. The peptides were applied to a Waters μBondapak C_{18} column equilibrated in 0.1% trifluoroacetic acid. The peptides were eluted from the column with a 0–60% gradient of acetonitrile in 80 min. The flow rate on the Varian Model 5000 HPLC was 2 mL/min, and 2-mL fractions were collected. The fractions were dried, hydrolyzed in 5.7 N HCl for 2 h in vacuo, and analyzed on a Beckman 121 MB amino acid analyzer. The peptides that contained DMA were purified by rechromatography on the C_{18} column with an extended acetonitrile gradient of 15–40% in 60 min. The peptide was judged to be pure by the symmetry of the HPLC peak and by amino acid sequencing.

Amino Acid Sequencing. The tryptic peptide from protein C23 was sequenced on the Applied Biosystems Model 470A protein sequencer using the O2NRUN program with trifluoroacetic acid conversion to the phenylthiohydantoin (PTH) derivatives. PTH-amino acids were identified by HPLC on a Waters Nova Pak C_{18} column using a Waters HPLC system consisting of a Model 840 control station, a Model 710B WISP automatic sample processor, a Model 490 detector, and Model 510 pumps. Norleucine was added to each sample and served as an internal standard. Solvent A was 84% sodium acetate (50 mM, pH 5.0)/16% acetonitrile; solvent B was 60% 2-propanol/40% water. Both solvents were blanketed under helium. The column was maintained at 40 °C and PTH-amino acids were eluted at a flow rate of 1 mL/min with the following gradient: 0–0.5 min, 0% solvent B; 0.5–3.5 min, convex gradient to 38% solvent B (Waters' curve 5); 3.5–12 min, isocratic elution at 38% solvent B. All PTH-amino acids were eluted from the column by 11 min at 38% solvent B. Absorbance was monitored at 265 and 313 nm (for Ser and Thr); background absorbance caused by the gradient was subtracted from each chromatogram by the Model 490 detector. The sequence was determined twice on samples of 160 and 320 pmol with an average repetitive yield of 94%.

Results

Protein C23 was purified from Novikoff hepatoma cells as described earlier (Lischwe et al., 1981). The protein was digested with trypsin, and the resulting peptides were purified by HPLC on a μBondapak C_{18} column. Three peptides were found to contain DMA. One of these peptides was sequenced on the Applied Biosystems Model 470A gas-phase sequencer. Separation of the standard PTH-amino acids on the Nova Pak C_{18} column is shown in Figure 1, panel A. PTH-N^G,N^G-dimethylarginine clearly separated from the common PTH-amino acids. PTH-N^G-monomethylarginine is also resolved on this system and elutes immediately after PTH-glycine (data not shown). The chromatographs of the PTH-amino acids obtained from steps 7–9 are shown in Figure 1, panels B–D. The quantitative sequence results are illustrated in Figure 2. The peptide that is 53 amino acids long contained 34 Gly, 10 DMA, and 6 Phe residues and 1 Glu, 1 Asp, and 1 Lys residue. No N^G-monomethylarginine or arginine was detected in any of the steps containing DMA. Chymotryptic peptides, rep-

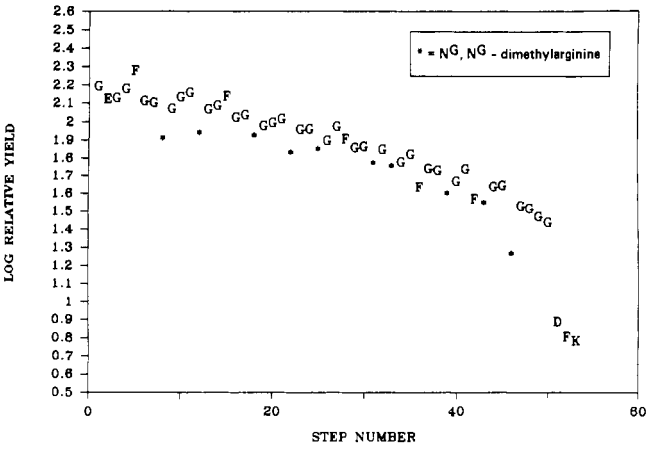

FIGURE 2: Quantitative amino acid sequence results of a tryptic peptide from protein C23. The yield of the PTH-amino acids was normalized on the basis of the recovery of the internal standard (norleucine). The logarithm of the yield, in picomoles, of the PTH-amino acids is plotted vs. the step number. The asterisks denote N^G,N^G-dimethylarginine. No arginine or N^G-monomethylarginine was detected at the steps containing DMA.

resenting residues 29–36 and 43–52 of this peptide, have also been isolated and sequenced (data not shown).

Discussion

The site specificity for the protein arginine methylases is not clearly understood (Paik & Kim, 1980). Position 107 in the myelin basic protein sequence was found to be a mixture of arginine, monomethylarginine, and dimethylarginine (Baldwin & Carnegie, 1971; Carnegie, 1971). In the in vitro labeling experiment using *S*-adenosylmethyl-L-[^{14}C]-methionine, a crude arginine methylase preparation, and human myelin basic protein as a substrate, only residue 107 of this protein was radiolabeled (Baldwin & Carnegie, 1971). There are a total of 18 arginine residues in human myelin basic protein. The sequence from residues 105 through 109 in this protein is

DMA
Lys-Gly-MMA-Gly-Leu
Arg

(Baldwin & Carnegie, 1971). In protein C23, nine of the ten methylated arginine residues that have been sequenced are flanked by glycine residues and the other methylarginine has two adjacent glycine residues on the carboxyl side (Table I). In the 34-kDa nucleolar protein, three of the six DMA residues are flanked by glycine and the other three DMA residues have an adjacent glycine (Table I). A single adjacent glycine to the arginine residue appears to be required for methylation. The lack of a side group on the glycine residue(s) may facilitate access of the methylase to the arginine residue.

In a computer search of published arginine-containing sequences, one protein was found to contain Gly and Arg clusters. It is the 100-kDa adenovirus late protein, which has the sequence *GRGG*ILGQS*GRG*F*GRGGG*D from residue 726 to residue 745 (Kruijer et al., 1981; Galibert et al., 1979). Since these sequence data were derived from the DNA sequence, it is not known whether this protein contains methylated arginines.

Protein C23 comprises 9.5% of the total nucleolar proteins of Novikoff hepatoma cells (Lischwe et al., 1982). It has been immunochemically localized to the nucleolus, found to be concentrated in the fibrillar region of the nucleolus, and present on the nucleolus organizer region of metaphase chromosomes (Lischwe et al., 1981; Ochs et al., 1983; Spector et al., 1984). The purified protein was found to be highly phosphorylated

Table I: Arginine Methylation Sites[a]

Nucleolar Protein C23

Gly-Glu-Gly-Gly-Phe-Gly-Gly-DMA-Gly-Gly-Gly-DMA-Gly-Gly-Phe-Gly-Gly-DMA-Gly-Gly-Gly-DMA-Gly-Gly-DMA-Gly-Gly-Phe-Gly-Gly-DMA-Gly-DMA-Gly-Gly-Phe-Gly-Gly-DMA-Gly-Gly-Phe-DMA-Gly-Gly-DMA-Gly-Gly-Gly-Asp-Phe-Lys

Nucleolar 34 kDa protein (unpublished experiments)

5 15
Phe-Ser-Pro-DMA-Gly-Gly-Gly-Phe-Gly-Gly-DMA-Gly-Gly-Phe-Gly-Asp-DMA-
25
Gly-Gly-DMA-Gly-Gly-Gly-DMA-Gly-Gly-DMA

Human Myelin Basic Protein (Baldwin & Carnegie, 1971)

DMA*
105 MMA
Lys-Gly-Arg-Gly-Leu-Ser-Leu-Ser-Arg-Phe

[a] The asterisked DMA in myelin basic protein is N^G,N'^G-dimethylarginine, whereas DMA in protein C23 and the 34-kDa protein is N^G,N^G-dimethylarginine.

and to contain an acidic tryptic phosphopeptide (Mamrack et al., 1979). Protein C23 has DNA binding properties; it was reported to show a preference for DNA sequences upstream from the genes for ribosomal RNA (Olson et al., 1983). Several reports suggest that protein C23 is a component of the preribosomal RNP particles (Prestayko et al., 1974; Olson & Thompson, 1983). The 34-kDa nucleolar protein is, apparently, an RNA-associated protein. Sera of several scleroderma patients contain antibodies that react with this protein (Lischwe et al., 1985). These patients' sera and affinity-purified antibodies to the 34-kDa protein precipitated U3 RNA containing particles (Lischwe et al., 1985). The clusters of glycine and DMA may be characteristic of a certain class of RNA-associated proteins.

Several nucleoplasmic hnRNA-associated proteins have also been found to be rich in glycine (≈20 mol %) and to contain N^G,N^G-dimethylarginine (Beyer et al., 1977; Karn et al., 1978; Thomas et al., 1983). These basic, RNA-associated proteins may also contain clusters of glycine and DMA.

The functional role of these glycine, DMA clusters interspersed with phenylalanines is currently not clear. This domain is highly basic at physiological pH and, thus, may represent a nucleic acid binding region in protein C23 and the 34-kDa protein. It can be speculated that the interspersed phenylalanine residues serve as a hydrophobic link between the core of the protein and this basic domain. The positively charged DMA residues are likely exposed, thus being free to interact with nucleic acid. The relative positions of the DMA residues in this domain may determine specificity for certain nucleic acid molecules.

Registry No. DMA, 30315-93-6; Gly, 56-40-6; L-Arg, 74-79-3.

References

Baldwin, G. S., & Carnegie, P. R. (1971) *Science (Washington, D.C.) 171*, 579–581.

Beyer, A. L., Christensen, M. E., Walker, B. W., & LeStourgeon, W. M. (1977) *Cell (Cambridge, Mass.) 11*, 127–138.

Boffa, L. C., Sterner, R., Vidali, G., & Allfrey, V. G. (1979) *Biochem. Biophys. Res. Commun. 89*, 1322–1327.

Carnegie, P. R. (1971) *Biochem. J. 123*, 57–67.

Crang, A. J., & Jacobson, W. (1982) *J. Neurochem. 39*, 244–247.

DesJardins, K. C., & Morell, P. (1983) *J. Cell Biol. 97*, 438–446.

Farooqui, J. Z., Tuck, M., & Paik, W. K. (1985) *J. Biol. Chem. 260*, 537–545.
Galibert, F., Herisse, J., & Courtois, G. (1979) *Gene 6*, 1–22.
Karn, J., Boffa, L. C., Vidali, G., & Allfrey, V. G. (1978) *J. Biol. Chem. 252*, 7307–7322.
Kruijer, W., van Schaik, F. M. A., & Sussenbach, J. S. (1981) *Nucleic Acids Res. 9*, 4439–4457.
Lischwe, M. A., Richards, R. L., Busch, R. K., & Busch, H. (1981) *Exp. Cell Res. 136*, 101–109.
Lischwe, M. A., Roberts, K. D., Yeoman, L. C., & Busch, H. (1982) *J. Biol. Chem. 257*, 14600–14602.
Lischwe, M. A., Ochs, R. L., Reddy, R., Cook, R. G., Yeoman, L. C., Tan, E. M., Reichlin, M., & Busch, H. (1985) *J. Biol. Chem.* (in press).
Mamrack, M. D., Olson, M. O. J., & Busch, H. (1979) *Biochemistry 18*, 3381–3386.
Ochs, R., Lischwe, M. A., O'Leary, P., & Busch, H. (1983) *Exp. Cell Res. 146*, 139–149.
Olson, M. O. J., & Thompson, B. A. (1983) *Biochemistry 22*, 3187–3193.
Olson, M. O. J., Rivers, Z. M., Thompson, B. A., Kao, W. Y., & Case, S. T. (1983) *Biochemistry 22*, 3345–3351.
Paik, W. K., & Kim, S. (1980) *Protein Methylation*, Wiley, New York.
Prestayko, A. W., Klomp, G. R., Schmoll, D. J., & Busch, H. (1974) *Biochemistry 13*, 1945–1951.
Rothblum, L. I., Mamrack, P. M., Kunkle, H. M., Olson, M. O. J., & Busch, H. (1977) *Biochemistry 16*, 4716–4721.
Spector, D. L., Ochs, R. L., & Busch, H. (1984) *Chromosoma 90*, 139–148.
Thomas, J. O., Glowacka, S. K., & Szer, W. (1983) *J. Mol. Biol. 171*, 439–455.

Articles

Amino Acid Sequence of an Active Fragment of Rabbit Skeletal Muscle Myosin Light Chain Kinase†

Koji Takio,‡ Donald K. Blumenthal,§ Arthur M. Edelman,§ Kenneth A. Walsh,∥ Edwin G. Krebs,‡ and Koiti Titani*,∥

Howard Hughes Medical Institute and Departments of Biochemistry and Pharmacology, University of Washington, Seattle, Washington 98195

Received April 2, 1985

ABSTRACT: The amino acid sequence of a 368-residue segment at the carboxyl-terminus of rabbit skeletal muscle myosin light chain kinase (MLCK) has been determined. The sequence was derived primarily from analysis of two complementary sets of fragments obtained by cleavage at methionyl and arginyl bonds in S-carboxymethylated MLCK. The segment included a 360-residue fragment produced by limited tryptic digestion of MLCK. This fragment was both catalytically active and dependent on Ca^{2+}–calmodulin. Unique structural features of MLCK have been identified, and a likely calmodulin interaction site is suggested. Sequence comparisons of MLCK to other protein kinases indicate close structural relationships in spite of marked differences in physicochemical properties, enzymatic characteristics, and regulatory response among these enzymes.

Protein kinases perform important regulatory roles in response to both intracellular and extracellular signals [cf. reviews by Krebs & Beavo (1979), Flockhart & Corbin (1982), Klee & Vanaman (1982), Nishizuka (1984), and Stull et al. (1985)]. Specific protein kinases are thought to control a variety of cellular functions including glycogen metabolism, muscle contraction, and growth. The catalytic activities of the various protein kinases are regulated by agents as diverse as hormones, Ca^{2+}–calmodulin, cyclic nucleotides, growth factors, and diacylglycerol. Many protein kinases are key enzymes in cascade systems where they regulate, and are regulated by, other protein kinases. Thus, knowledge of the structure of these enzymes is important in understanding their function and their regulation.

Shoji et al. (1981, 1983) reported the amino acid sequence of the catalytic subunit of cAMP-dependent protein kinase, and subsequently Barker & Dayhoff (1982) found that the protein was homologous to an oncogene product ($pp60^{src}$).[1] The latter is also a protein kinase that phosphorylates tyrosine, rather than serine, residues of target proteins (Hunter & Sefton, 1980). Several other oncogene products and hormone

† This work was supported in part by a grant from the National Institutes of Health (GM15731).
* Address correspondence to this author.
‡ Howard Hughes Medical Institute.
§ Department of Pharmacology.
∥ Department of Biochemistry.

[1] Abbreviations: cGK, cGMP-dependent protein kinase; DTT, dithiothreitol; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid; EDTA, ethylenediaminetetraacetic acid; MOPS, 3-(*N*-morpholino)propanesulfonic acid; CM, *S*-carboxymethyl; HPLC, high-performance liquid chromatography; LBTI, lima bean trypsin inhibitor; MLCK, myosin light chain kinase; PbK, phosphorylase *b* kinase; $pp60^{src}$, transforming phosphoprotein of Rous sarcoma virus; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; kDa, kilodalton(s); $NaDodSO_4$, sodium dodecyl sulfate.

0006-2960/85/0424-6028$01.50/0