0273-2289/96/6101-02—0097$07.75

# HIV-I Protease

## Cloning, Expression, and Purification

NATALYA I. DERGOUSOVA,* ALEXANDER YU. AMERIK, ALLA M. VOLYNSKAYA, AND LEV D. RUMSH

*Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, Moscow, 117871, Russia*

## ABSTRACT

A new method for obtaining HIV-I protease was suggested. Fusion proteins composed of the N-terminal fragment of human γ-interferon and HIV-I protease connected with $(Asp)_4Lys$ (protein I) or Asp-Pro (protein II) linkers were expressed in *Escherichia coli* cells. The fusion proteins were produced as insoluble inclusion bodies in the 20% yield of total cell protein. Protein I was cleaved by enterokinase. The solubility of protein I was increased by treating with Na-sulfite/Na-tetrathionate under denaturing conditions.

Optimal conditions for efficient acidic hydrolysis of protein II at Asp-Pro bond were found. The hydrolysis products were separated by reversed-phase FPLC. The amount of tryptophan and cysteine residues in the enzyme obtained was estimated. The activity of HIV-I protease was determined using the chromogenic peptide AlaArgVal NleNphGluAlaNle$NH_2$ and a high-mol-wt substrate consisting of β-galactosidase and a fragment of *gag* proteins, including p17-p24 processing site.

**Index Entries:** HIV-1 protease; fusion protein; enterokinase; Asp-Pro bond; protein folding.

**Abbreviations:** Tris, tris(hydroxymethyl) aminomethane; MES, 2-(*N*-morpholino)ethanesulfonic acid; DTT, dithiothreitol; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis, IPTG, isopropyl-β-thiogalactopyranoside; TFA, trifluoroacetic acid.

*Author to whom all correspondence and reprint requests should be addressed. E-mail: nd@enzyme.siobc.ras.ru.

## INTRODUCTION

Human immunodeficiency virus protease (HIV-I) plays a key role in the virus life cycle. This enzyme cleaves the precursors of *gag* and *gag-pol* polyproteins with formation of structural proteins and enzymes of mature virus (*1,2*). Therefore, it represents an attractive target for therapeutic intervention in the treatment of acquired immunodeficiency syndrome (AIDS). These studies require considerable amounts of the enzyme.

HIV-I protease is an aspartic protease. According to the crystallography data, the enzyme is a homodimer containing two aspartic acid residues in its active site (*3–5*).

A considerable number of methods have been developed for obtaining HIV-I protease that include the expression of its gene in bacterial cells and total peptide synthesis of the enzyme (for review, *see 6*). In most cases, purification of HIV-I protease was complicated by a low yield of the enzyme and its insolubility.

In this article, we describe a new effective method for obtaining HIV-I protease as a fusion protein with the N-terminal fragment of human γ-interferon. Subsequent hydrolysis led to obtaining high amount of active HIV-I protease.

## MATERIALS AND METHODS

### Cloning, Expression, and Purification of Hybrid Proteins

The gene encoding HIV-I protease was cloned into pIFN-prot, the pIFN-γ-*trp*-2-derived vector (*7*) that contained a synthetic DNA fragment coding for $(Asp)_4Lys$ pentapeptide. The $(Asp)_4Lys$ linker was replaced by Asp-Pro using oligonucleotide-directed mutagenesis (*8*). The DNA sequence was determined by the Maxam-Gilbert solid-phase method (*9*).

*Escherichia coli* cells MH 1 bearing pIFN-prot were grown overnight in the Luria-Bertani medium containing tetracycline (5 μg/mL) and tryptophan (20 μg/mL) at 37°C. The overnight culture was diluted with the M9 medium containing 0.2% casamino acids and tetracycline (5 μg/mL) at a culture–medium ratio of 1:100, and the cells were grown further at 37°C for 3 h (to an absorbance at 600 nm of 0.3). Indole-3-acrylic acid was then added (20 μg/mL). Four hours after induction, the culture was chilled to 10°C; the cells were harvested by centrifuging and stored at −70°C. Inclusion bodies were isolated as described elsewhere (*10*).

### Cloning, Expression and Purification of High-Mol-Wt Substrate

The *Hind*III/*Hind*III fragment (628 bp) that contained a fragment of the *gag* gene and the native p17-p24 processing site was cloned into the

pUR290 plasmid vector (*11*). *E. coli* cells XL1-Blue bearing the recombinant plasmid were grown at 37°C overnight in the Luria-Bertani medium containing ampicillin (100 μg/mL). The overnight culture was diluted with the fresh medium at a ratio of 1:100, and the cells were further grown for 3 h (to an absorbance at 600 nm of 0.4). Then, IPTG (1 m*M*) was added. The cells were grown for 3 h, and the culture was then chilled. The cells were harvested by centrifuging and stored at −70°C. Inclusion bodies were isolated as above.

The isolated inclusion bodies were dissolved in 0.01*M* Tris-HCl (pH 7.5) containing 1 m*M* EDTA, 0.1% Triton X-100, 10 m*M* DTT, and 8*M* urea, and applied on a CL-Sepharose 4B column (2 × 70 cm) equilibrated with the same buffer. Proteins were eluted at a flow rate of 5 mL/h. Fractions containing *β-gal-gag* were pooled and dialyzed against 20 m*M* MES (pH 6.25) containing 5 m*M* 2-mercaptoethanol, 4 m*M* EDTA, and 1*M* NaCl (buffer A).

## Hydrolysis of Fusion Protein I by Enterokinase

Inclusion bodies were dissolved (200 μg/mL) in 20 m*M* Tris-HCl (pH 8.0) containing 8*M* urea and dialyzed against the same buffer that contained no urea. Enterokinase was added to the protein I solution (0.2 mg/mL) at a protein–enzyme ratio of 90:1. The reaction mixture was incubated at 37°C overnight.

## Acidic Cleavage of Hybrid Protein II

Inclusion bodies were dissolved in 75% formic acid (5 mg/mL) and incubated under argon at 70°C for 2 h. The hydrolysis products were lyophilized and then dissolved in 0.05*M* MES (pH 4.5) containing 1 m*M* DTT, 1 m*M* EDTA, and 8*M* urea, and separated by reversed-phase FPLC on a ProRPC HR 5/10 column (Pharmacia) in a 24–55% linear gradient of acetonitrile/0.1% TFA over 60 min at a flow rate of 0.5 mL/min. Protease-containing fractions were pooled and lyophilized.

## Analysis of Tryptophan and Cysteine Residues

To determine the amount of tryptophan residues, HIV-I protease was dissolved ($10^{-5}M$) in 50% acetic acid. The fluorescence spectrum was registered in the range 310–400 nm and compared to those of tryptophan solutions of different concentrations.

The amount of SH- groups was determined by the Ellman method (*12*) or by the ammetric titration using silver nitrate (*13*).

## Refolding of HIV-I Protease

Lyophilized HIV-I protease was dissolved (0.1 mg/mL) in 0.02*M* Tris-HCl, pH 8.0, containing 25 m*M* NaCl, 2% 2-mercaptoethanol, and

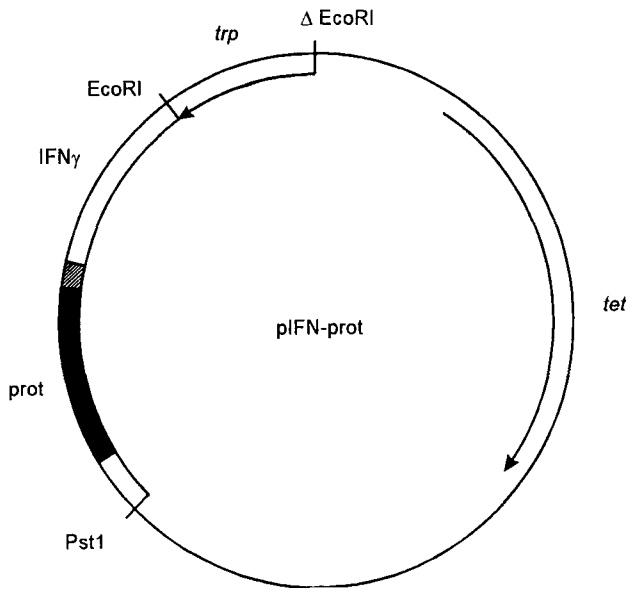


Fig. 1. Structure of the pINF-prot expression vector. The region corresponding to the linkers $(Asp)_4Lys$ and Asp-Pro in proteins I and II, respectively, is dashed.

8*M* urea, and dialyzed against 20 m*M* phosphate buffer (pH 7.0) containing 25 m*M* NaCl, 2% 2-mercaptoethanol, and 10% glycerol at 4°C for 24 h (Appelt, K., unpublished data). The enzyme solution was stored at −70°C.

## Protease Assay

HIV-I protease (10 ng) was added to solution of the high-mol-wt substrate (40 ng) in 100 μL of buffer A. The reaction mixture was incubated at 37°C overnight. The products were separated by electrophoresis in 15% PAGE.

The hydrolysis of LysAlaArgValNleNphGluAlaNleNH$_2$ (kindly provited by J. W. Erikson, National Cancer Institute, USA) was performed in buffer A at 37°C as described elsewhere (*14*). Kinetic parameters were calculated using the program ENZFITTER (*15*).

# RESULTS

## Cloning and Expresion of HIV-I Protease

The gene encoding HIV-I protease was cloned into the modified pIFN-γ-*trp*-2 vector under control of the *trp* promoter (Fig. 1). Synthetic oligonucleotide encoding $(Asp)_4Lys$ linker was introduced at an *Asu*II restriction site of the γ-interferon gene. The pentapeptide $(Asp)_4Lys$ represents a site of cleavage by enterokinase from bovine duodenum (EC 3.4.21.9). *E. coli*

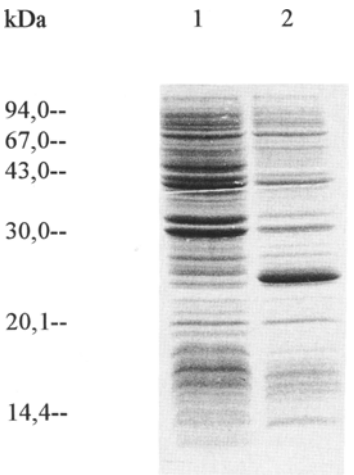


Fig. 2. SDS-PAGE of total proteins from *E. coli* cells containing pINF-prot: (1), without induction; (2) induction by indole-3-acrylic acid.

cells MH 1 were transformed with the obtained recombinant pIFN-prot vector. The expression of the fusion protein was induced by growing the transformed cells in a tryptophan-depleted medium in the presence of indole-3-acrylic acid. The calculated molecular mass of fusion protein I consisting of the N-terminal fragment of human γ-interferon and HIV-I protease was 22.5 kDa. Extracts from the induced and noninduced cells bearing the expression vector pIFN-prot were analyzed by SDS-PAGE. The induction by indole-3-acrylic acid yielded a protein of expected 23 kDa (Fig. 2, lane 1). The fusion protein was produced as insoluble inclusion bodies with a 20% yield that considerably exceeded the yields of fusion proteins. This considerably facilitated its purification.

## Cloning, Expression, and Purification of High-Mol-Wt Substrate

The high-mol-wt substrate consisting of β-galactosidase and a fragment of *gag* polyprotein was obtained to determine the activity of HIV-I protease. The fusion protein contained the native p17-p24 site of processing.

The expression of high-mol-wt substrate was analyzed by SDS-PAGE (Fig. 3). We suggested that the presence of β-galactosidase in the extracts from the cells transformed with the recombinant pUR-GAG plasmid (lane 4) may be owing to cleavage of the recombinant *β-gal-gag* protein within the C-terminal fragment of β-galactosidase by unidentified bacterial protease. Thus, a partial degradation product comigrating with β-galactosidase in SDS-polyacrylamide gels was also observed by other authors in a human apolipoprotein AI-β-galactosidase fusion protein (*16*).

The fusion protein was produced as insoluble inclusion bodies and was purified by gel filtration on a CL-Sepharose 4B column in a buffer containing 8*M* urea.

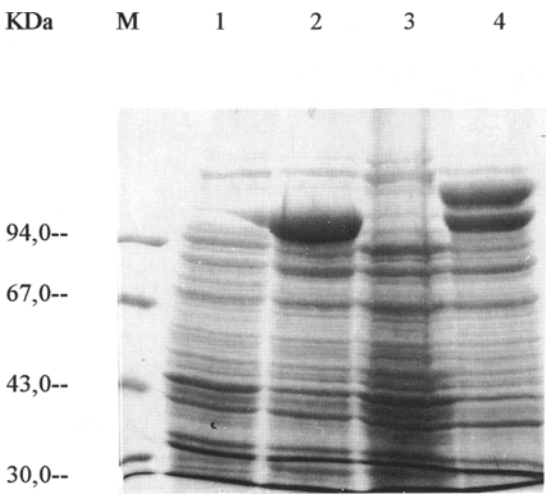


Fig. 3. Expression of β-galactosidase and *β-gal-gag* fusion protein. SDS-PAGE of total proteins from *E. coli* cells containing the pUR290 (1, 2) and pUR-GAG (3, 4) expression vectors:(1, 3), without induction; (2, 4) induction by IPTG.

## Cleavage of Protein I by Enterokinase

Fusion protein I was cleaved by enterokinase. Prior to the cleavage, the protein was dissolved in a buffer containing 8*M* urea and dialyzed against 20 m*M* Tris-HCl (pH 8.0). The solubility of protein I at neutral and slightly basic pH (pH optimum for enterokinase) did not exceed 0.2 mg/mL. The results of hydrolysis are presented in Fig. 4.

Low solubility of protein I appeared to be owing to the presence of disulfide bonds that were not reduced on the isolation of the inclusion bodies (*17*). To break these bonds, protein I was treated with Na-sulfite/ Na-tetrathionate under the denaturing conditions (*18*). This considerably increased the solubility of the protein at slightly basic pH. However, at the concentration of 4 mg/mL, only 10% of protein I was hydrolyzed by enterokinase. The contradicted the data on the efficient hydrolysis of highly soluble proteins containing the $(Asp)_4Lys$ linker (*19*). Probably, the hydrolysis was impaired by the incorrect exposure of the linker on surface of the protein globule.

## Acidic Cleavage of Protein II

HIV I protease contains proline residue at its N-terminus. It is known that the Asp-Pro bond can be cleaved under mild acidic conditions. Acidic hydrolysis of the Asp-Pro bond was used for obtaining various recombinant proteins (20–22). Using oligonucleotide-directed mutagenesis, we replaced $(Asp)_4Lys$ by the Asp-Pro linker. The fusion protein II was produced as insoluble inclusion bodies in a high yield. Various reaction conditions were tested including hydrolysis in acetic or formic acids in the

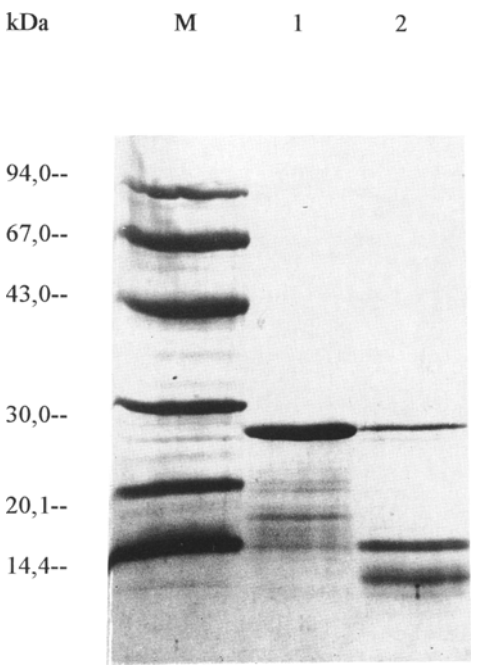


Fig. 4. Hydrolysis of IFN–prot I hybrid protein by enterokinase: (1), control, incubation in the absence of enterokinase; (2), hydrolysis products.

presence or in the absence of guanidine chloride at different temperatures. The maximum efficiency of hydrolysis was observed when the reaction was performed in 75% formic acid at 70°C for 2 h. Under these conditions, almost complete hydrolysis of protein II occurred. Considering different hydrophobicity of the HIV-I protease and γ-interferon molecules, the reaction products were separated with high efficiency by reversed-phase FPLC on a ProRPC HR 5/10 column in a gradient of acetonitrile concentration (Fig. 5).

Acidic hydrolysis is known to be accompanied by a number of side reactions (i.e., disulfide exchange, partial destruction of tryptophan residues, and so on). HIV-I protease contains two tryptophan residues. Hence, it seemed important to elucidate whether these residues were affected by the acidic hydrolysis. We compared the fluorescence spectrum of the isolated enzyme and that of the tryptophan solution of an equimolar concentration. This comparison suggested that the tryptophan residues are stable under these conditions.

The HIV-I proteinase molecule (99 amino acid residues) contains two cysteine residues (Cys67 and Cys95). The modification of Cys67 completely inactivated the enzyme (23). Using the Ellman reagent and the ammetric titration using silver nitrate, we demonstrated that 89% of SH-groups were free in the isolated enzyme.

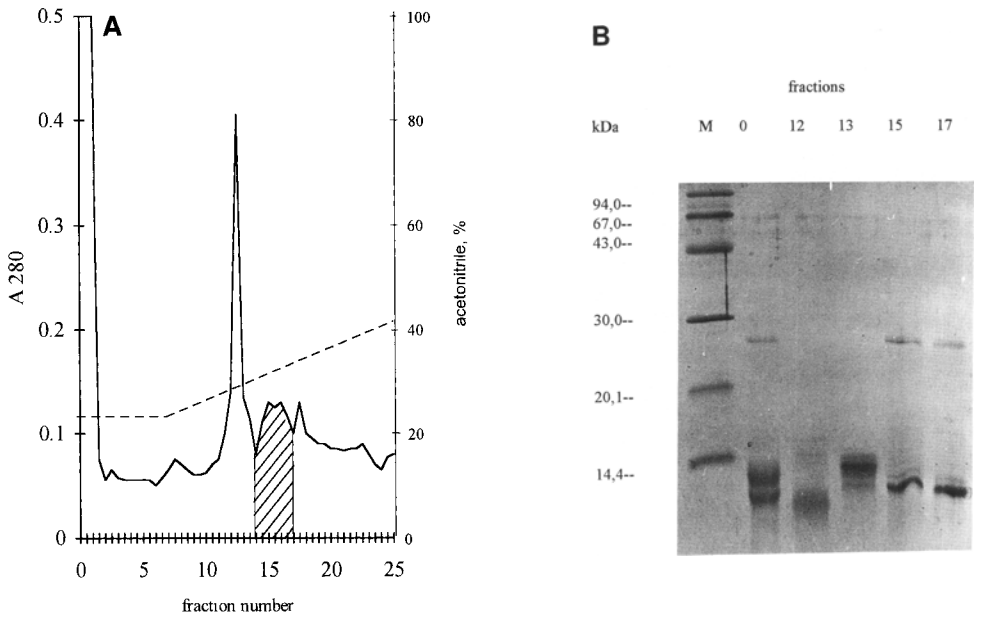


Fig. 5. Purification of HIV-I protease: **A,** reversed-phase FPLC on a ProRPC HR 5/10 column.; **B,** SDS-PAGE analysis of fractions. Dashed line, gradient of acetonitrile in 0.1% TFA solution in water; protease-containing fractions are dashed. Fr. 0, products of hydrolysis by Asp-Pro bond.

## Refolding of HIV-I Protease and Determination of Protease Activity

The purified HIV-I protease displayed no activity. We attempted to reactivate it by refolding using the procedures described earlier (*21,24*). However, the most efficient refolding was observed when the protein solution in 20 m*M* Tris-HCl buffer (pH 8.0) containing 8*M* urea was dialyzed against 20 m*M* phosphate buffer (pH 7.0).

The activity of HIV-I protease was analyzed using the high-mol-wt substrate *β-gal-gag* (Fig. 6).

The activity of HIV-I protease was assayed using synthetic chromogenic substrate LysAlaArgValNleNphGluAlaNleNH$_2$. The hydrolysis of the substrate was followed by monitoring a decrease in absorbance at 300 nm (pH 6.5; ionic strength 1*M*). The kinetic parameters were determined ($k_{cat}$ = 40 $s^{-1}$; $K_m$ = 90 $\mu M$), which correlated well with the published data (*14*).

## DISCUSSION

HIV-I protease is not easily expressed in *E. coli* because of its cytotoxic properties (*25,26*) and low solubility. Several groups have described bacterial production of HIV-I protease, which was produced by autocatalytic processing of a larger precursor (*27–30*) or by fusions with either β-galac-

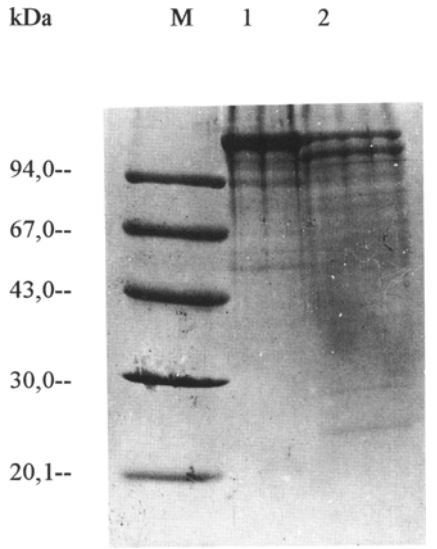


Fig. 6. Hydrolysis of *β-gal-gag* by HIV-I protease; (1), control, incubation in the absence of HIV-I protease; (2), hydrolysis products.

tosidase (*31,32*), dihydrofolate reductase (*25*), β-lactamase (*33*), maltose-binding protein (*34*), or the IgG-binding domain of protein A (*35*). In most cases, the expression levels were low, and the protein could be detected only by immunoblotting. It is also important to note that subsequent cleavage of the fusion proteins was often complicated by unspecific proteolysis.

In this work, we describe a new method for obtaining HIV-I protease in a fusion with the N-terminal fragment of human γ-interferon. We suggested two linkers for connecting the fragments of the fusion protein. The linker $(Asp)_4Lys$ represents an enterokinase cleavage site. Enterokinase possesses an extremely high specificity compared to proteases used for hydrolysis of fusion proteins. We also used the acid-labile Asp-Pro linker, which allowed us to obtain the mature form of the enzyme.

We obtained two fusion proteins that consisted of the N-terminal fragment of human γ-interferon and HIV-I protease connected by the $(Asp)_4Lys$ linker (protein I) or the Asp-Pro linker (protein II). The corresponding genes were expressed in *E. coli* cells with a high yield. The expression of the fusion proteins as insoluble inclusion bodies considerably facilitated their purification.

For unknown reasons, the enzymatic cleavage of protein I by enterokinase was ineffective, but acid hydrolysis of protein II led to it complete cleavage. After separation of the hydrolysis products and subsequent refolding, high amounts of the active protease were obtained. Its kinetic parameters did not differ from those published earlier (*14*). The obtained pINF-prot vector may be used for site-directed mutagenesis of HIV-I proteinase and searching for new inhibitors.

## ACKNOWLEDGMENT

This work was supported by grant from the State Program "Novel Methods for Bioengineering: Protein Engineering" (No 7).

## REFERENCES

1. Kohl, N., Emini, E., Schlif, W., Davis, E., Helmbach, J., Dixon, R., Scolaick, E., and Sigal, I. (1988), *Proc. Natl. Acad. Sci. USA* **85,** 4686–4690.
2. Peng, C., Ho, B., Chang, T., and Chang, N. (1989), *J. Virol.* **63,** 2550–2556.
3. Pearl, L. and Taylop, W. (1987), *Nature* **329,** 351–354.
4. Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., Rosenberg, M., Moore, M. L., Magaard, V. W., and Debonck, C. (1989), *Proc. Natl. Acad. Sci. USA* **86,** 1841–1845.
5. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schnreider, J., and Kent, S. B. H. (1989), *Science* **245,** 616–621.
6. Stebbins, J. and Debouck, C. (1994), *Methods Enzymol.* **241,** 3–16.
7. Sverdlov, E. D., Tsarev, S. A., Krykbaev, R. A., Chernov, I. P., and Rostapshov, V. M. (1987), *FEBS Lett.* **212,** 233–236.
8. Foss, K. and Mc Clain, W. H. (1987), *Gene* **59,** 285–290.
9. Chuvpilo, S. A. and Kravchenko, V. V. (1985), *FEBS Lett.* **179,** 34–36.
10. Lin, K. and Chang, S. (1991), *BioTechniques* **11,** 748–752.
11. Ruether, U. and Mueller-Hill, B. (1983), *EMBO J.* **2,** 1791–1794.
12. Riddles, P. W., Blakeley, R. L., and Zerner, B, (1970), *Methods Enzymol.* **91,** 49–60.
13. Gevondyan, N. M., Gevondyan, V. S., and Modyanov, N. N. (1993), *Biochem. Mol. Biol. Int.* **29,** 327–337.
14. Richards, A. D., Phylip, L. H., Farmerie, W. G., Scarborough, P. E., Alvarez, A., Dunn, B. M., Hirel, P. -H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V., and Kay, J. (1990), *J. Biol. Chem.* **265,** 7733–7736.
15. Leatherbarrow, R. J. (1988), *Enzfitter.*
16. Loreuzetti, R., Sidoli, A., Palemba, R., Monaco, L., Martineau, D., Lappi, D., and Soria, M. (1986), *FEBS Lett.* **194,** 343–346.
17. Fischer, B., Sumner, I., and Goodenough, P. (1993), *Biotechol Bioeng.* **41,** 3–13.
18. Bailey, J. L. and Cool, R. D. (1959), *J. Biol. Chem.* **234,** 1733–1739.
19. Mikhailova, A. G., Shibanova, E. D., Rumsh, L. D., and Antonov, V. K. (1994), *Bioorg. Khim.* **20,** 883–893. (in Russ.).
20. Landon, M. (1977), *Methods Enzymol.* **47,** 145–149.
21. Boutelje, J., Karlstrom, A. R., Hartmanis, M. B. N., Holmgren, E., Sjogren, A., and Levine, R. L. (1990), *Arch. Biochem. Biophys.* **283,** 141–149.
22. Gaskin, D. J. H. and Wilderspin, A. F. (1992), *Biochem. Soc. Trans.* **20,** 162. S.
23. Karlstrom, A. R., Shames, B. D., and Levine, R. L. (1993), *Arch. Biochem. Biophys.* **304,** 163–169.
24. Hui, J. O., Tomasselli, A. G., Reardon, I. M., Lull, J. M., Brunner, D. P., Tomich, C. C., and Heinrikson, R. L. (1993), *J. Protein Chem.* **12,** 323–327.
25. Hostomsky, Z., Appelt, K., and Ogden, R. (1989), *Biochem. Biophys. Res. Commun.* **161,** 1056–1063.
26. Baum, E., Bebernitz, G., and Gluzman, J. (1990), *Proc. Natl. Acad. Sci. USA* **87,** 5573–5577.
27. Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., and Rosenberg, M. (1987), *Proc. Natl. Acad. Sci. USA* **84,** 8903–8906.

28. Hansen, J., Billich, S., Schulze, T., Sukrow, S., and Moelling, K. (1988), *EMBO J.* **7,** 1785–1791.
29. Strickler, J. E., Gorniak, J., Dayton, B., Meek, T., Moore, M., Magaard, V., Malinowski, J., and Debouck, C. (1989), *Proteins* **6,** 139–154.
30. Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A., and Moss, B. (1993), *Virology* **193,** 661–671.
31. Giam, C. -Z. and Boros, I. (1988), *J. Biol. Chem.* **263,** 14,617–14,620.
32. Valverde, V., Lemay, P., Masson, J. -M., Gay, B., and Boulanger, P. (1992), *J. Gen. Virol.* **73,** 639–651.
33. Korant, B. D. and Rizzo, C. J. (1990), *Biol. Chem. Hoppe-Seyler* **371,** 271–275.
34. Louis, J. M., Donald, R. A., Nashed, N. T., Wondrak, E. M., Jerina, D. M., Oroszlan, S., and Mora, P. T. (1991), *Eur. J. Biochem.* **199,** 361–369.
35. Boutelje, J., Karlstrom, A. R., Hartmanis, M. G. N., Holmgren, E., Sjogren, A., and Levine, R. L. (1990), *Arch. Biochem. Biophys.* **283,** 141–149.