Volume 270, number 1,2, 76-80 FEBS 08845 September 1990

Purification and characterization of the RNase H domain of HIV-1 reverse transcriptase expressed in recombinant *Escherichia coli*

S. Patricia Becerra[1], G. Marius Clore[2], Angela M. Gronenborn[2], Anders R. Karlström[3], Stephen J. Stahl[4], Samuel H. Wilson[1] and Paul T. Wingfield[4]

[1]*Laboratory of Biochemistry, NCI,* [2]*Laboratory of Chemical Physics, NIDDK,* [3]*Laboratory of Biochemistry, NHLB* and [4]*Protein Expression Laboratory, National Institutes of Health, Bethesda, MD 20892, USA*

Received 2 July 1990

The ribonuclease H (RNase H) domain of human immuno-deficiency virus (HIV-1) reverse transcriptase has been produced with the aim of providing sufficient amounts of protein for biophysical studies. A plasmid vector is described which directs high level expression of the RNase H domain under the control of the λ P_L promoter. The domain corresponds to residues 427–560 of the 66 kDa reverse transcriptase. The protein was expressed in *Escherichia coli* and was purified using ion-exchange and size exclusion chromatography. The purified protein appears to be in a native-like homogeneous conformational state as determined by ^{1}H-NMR spectroscopy and circular dichroism measurements. HIV-protease treatment of the RNase H domain resulted in cleavage between Phe-440 and Tyr-441.

HIV-1 reverse transcriptase; HIV-1 RNase H; HIV protease; Protein expression; Protein purification; Protein conformation

1. INTRODUCTION

Ribonuclease H (RNase H) activity resides in the C-terminal region of retroviral polymerases as shown by deletion experiments [1], linker insertion [2] and point mutagenesis [3,4]. The enzyme functions by cleaving the RNA portion of a DNA/RNA hybrid, a process which is required at several stages during reverse transcription of the viral genome within the host cell. Recent studies have established that the retroviral RNase H displays both endonuclease and $3' \rightarrow 5'$ exonuclease activity [5,6]. This activity is crucial for viral replication as mutant provirus, defective for RNase H function, fails to produce infectious virus particles [4,7]. Comparisons between retroviral reverse transcriptases and other known protein sequences revealed significant sequence similarity to *E. coli* RNase H [8], suggesting that the RNase H domain of reverse transcriptase and *E. coli* RNase H may have related three-dimensional structures.

Although recombinant HIV-reverse transcriptase of sufficient purity and abundance is available, high resolution structural studies have been prevented by the lack of crystals suitable for X-ray diffraction analysis. The crystals presently available are insufficiently ordered to allow a structure determination [9]. To study structure-function relationships of the RNase H portion of this protein, we have expressed the former in *E. coli* as a separate 15 kDa protein. The purified protein appears suitable for structural studies by both NMR spectroscopy and X-ray analysis, thereby providing an additional target in the search for therapeutic agents against the spread of HIV infection.

Correspondence address: P.T. Wingfield, Protein Expression Laboratory, National Institutes of Health, Bethesda, MD 20892, USA

2. MATERIALS AND METHODS

2.1. *Plasmid construction*

The coding sequence of the RNase H domain was derived from M13-RT (Becerra et al., submitted for publication), a subclone of the reverse transcriptase gene from HIV-1 strain HXB2, inserted between the EcoRI and *Hin*dIII cleavaged sites of M13mp18. The RNase H coding region is located on a 423 base-pair Asp718 - *Hin*dIII fragment, corresponding to nucleotides 3406–3809 of the HIV-1 sequence [10]. After removal of the internal Asp-718 site (an isoschizomer of *Kpn*I) by primer-directed mutagenesis changing the codon for Val-536 from GTA to GTT, the double-stranded RF form of this DNA was cleaved with Asp-718, the 5′ overhang was filled in with the Klenow fragment of DNA polymerase 1, and cut with *Hin*dIII. The resulting DNA fragment containing the coding region for the RNase H domain of reverse transcriptase was ligated to pEV-vrf1 [11], an ATG-containing expression vector that had been previously digested with *Eco*RI (the 5′ overhang filled in with Klenow DNA polymerase) and *Hin*dIII. This ligation mixture was used to transform *E. coli* strain RR1 [pRK248cIts]. The final construct p350-13 codes for the sequence Met-Asn-Glu-Leu followed by amino acids 427–560 of reverse transcriptase under control of the λ P_L promoter.

2.2. *Protein expression and fermentation*

Plasmid bearing *E. coli* was grown in a 2.5 liter fermenter. Culture media and fermentation conditions were essentially as described previously [12]. When cultures had reached an absorbance of about 10 at 650 nm, the temperature was raised to 42°C. After 3 h at 42°C the cells were harvested to yield about 30 mg (wet weight cells) per ml of culture. Harvested cells were stored at −80°C prior to use.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

2.3. *Purification of HIV - RNase H*

Buffers used in protein purification were: buffer A, 50 mM Tris/HCl, pH 8.0; buffer B, 20 mM sodium acetate, pH 4.9; buffer C, 50 mM sodium phosphate, pH 7.6. All buffers contained 1 mM sodium azide. Cells (60 g wet weight) were resuspended in 170 ml of Buffer A and broken by passage through a French pressure cell at 124 mPa (18,000 lbf/in^2). The suspension was sonicated briefly to reduce the viscosity, diluted to a final volume of 300 ml with buffer A and centrifuged at 10 000 × *g* for 45 min. The supernatant was recentrifuged at 100 000 × *g* for 1 h, then applied at 200 ml/h to a DEAE-Sepharose column (5 cm diameter × 20 cm) equilibrated with buffer A. A 1 liter linear gradient (0–0.35 M) of NaCl in buffer A was applied. HIV-RNase H was eluted with about 50–100 mM NaCl, pooled fractions (170 ml) were dialyzed against three changes of 4 liter buffer B. The slightly cloudy solution (180 ml) was clarified by centrifugation at 5000 × *g* for 30 min and the supernatant applied to a CM-Sepharose column (5 cm diameter × 15 cm) equilibrated with buffer B. A 1 liter gradient (0–0.35 M) of NaCl was applied. A major peak of HIV-RNase was eluted with about 200 mM NaCl. In addition a minor peak of protein (10% of major peak) was eluted with about 150 mM NaCl. Both peaks were processed separately. Pooled fractions were adjusted to pH 7.5 with 1 M Tris base and concentrated to 15–20 ml by ultrafiltration using Diaflo YM5 membranes (Amicon) and applied at 40 ml/h to a column (2.5 cm diameter × 95 cm) of Spectra/Gel AcA 54 (Spectrum/Medical Industries, Inc.) equilibrated with buffer C. Fractions containing HIV-RNase H were pooled, concentrated by ultrafiltration to 10 mg/ml, sterile filtered using 0.22 μm pore size Millex GV filter units (Millipore) and stored at −80°C.

2.4. *Protein estimations*

Protein concentration was estimated by measuring the absorbance at 280 nm. The calculated molar absorbance coefficient (ε) used was 13.50 mM cm^{-1} (based on a molecular mass of 15,200).

2.5. *Enzymic assay*

The substrate for the assay, [^{32}P]φx174 DNA-RNA hydrid, was prepared as described previously [13] using α[^{32}P]UTP as the label. The assay buffer was 50 mM Tris/HCl, 50 mM NaCl, 5 mM $MgCl_2$, 0.1 mM DTT, pH 7.9 and the substrate concentration about 0.5 μg/ml. Purified HIV-1 RNase H domain was incubated at 37°C with the above components and aliquots from the reaction were removed at various times and spotted onto DEAE filters (Whatman DE81). After washing 3 × 15 min in 5% Na_2HPO_4 and drying, the amount of undigested RNA remaining on the filter was determined by scintillation counting.

2.6. *Analytical methods*

SDS/PAGE, isolectric focusing and N-terminal sequence analysis were all performed as previously described [14].

2.7. *^{1}H-NMR spectroscopy*

The sample for NMR spectroscopy contained 0.6 mM RNase H in 50 mM sodium phosphate, pH 7.7 in 90% H_2O/10% D_2O. The spectrum was recorded on a Bruker AM600 600 MHz spectrometer at 35°C and the water resonance was suppressed by low power coherent presaturation. Chemical shifts are expressed relative to 4,4-dimethyl-4-silapentane-1-sulfonate.

2.8. *Circular dichroic spectra*

Spectra were recorded on a Jasco J-600 spectropolarimeter. Measurements in the far ultraviolet region (180–240 nm) were made using a protein solution of 1.0 mg/ml, a 0.01 cm pathlength cell and 1 nm bandwidth. Protein was buffered with 50 mM sodium phosphate, pH 7.5. The spectrum was the average of 4 scans with the base-line subtracted. The mean residue molecular mass was calculated as 110.3. Secondary structure content was estimated by the method of Provencher and Gloeckner [15].

2.9. *HIV-Protease cleavage of RNase H domain*

The HIV-Protease was produced in *E. coli* as a fusion product with a modified IgG binding domain derived from protein A and purified as described previously [16]. HIV-RNase H (2.3 mg/ml) was incubated with protease (0.15 mg/ml) in 50 mM sodium acetate, 0.1 mM EDTA, 10% glycerol, 5% ethylene glycol and 1 mM dithiothreitol pH 5.8, for 5 h at 37°C. The reaction was terminated by acidification with trifluroacetic acid. Products were separated on a Vydac C18 column run at 1 ml/min using a gradient from 0 to 70% acetonitrile in 0.05% trifluroacetic acid over 35 min (2%/min). A Tyr-containing peptide eluted with 35% acetonitrile, while a Trp con-

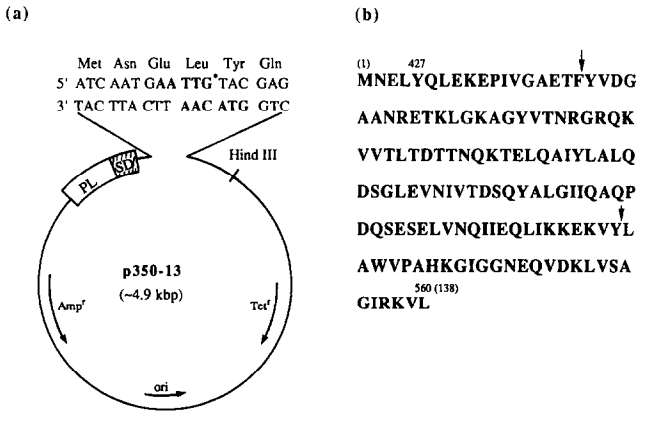

Fig. 1. (a) Expression plasmid for HIV-1 RNase H. The asterisk (*) refers to nucleotide 3407 of the HIV-1 gene sequence [10] and is the first base in the expression vector for the authentic reverse transcriptase. Bases in bold face refer to the junction region where overhangs were filled in with Klenow polymerase. PL, λ P_L promoter; SD, Shine-Delgarno ribosome binding site; Amp^r, ampicillin resistance gene; ori, origin of replication; Tet^r, tetracyclin resistance gene. Further details are given in the section 2. (b) Amino acid sequence of the HIV-1 RNase H domain. Tyr at position 427 is the first amino acid from the reverse transcriptase sequence. The arrows indicate the HIV-1 protease processing sites observed in this study.

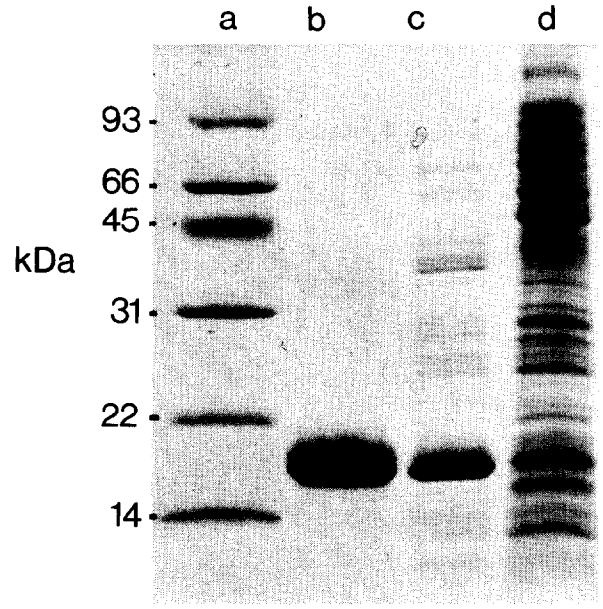

Fig. 2. SDS/PAGE of HIV-1 RNase H. Lane (d), cell extract after centrifugation; lane (c), pooled fraction after DEAE-Sepharose stage; lane (b) purified protein; and lane (a) M_r markers, the molecular masses of which are indicated in kDa. All samples were run on a 20% homogeneous Phast gel stained with Coomassie blue.

taining peptide eluted at 42% acetonitrile. Due to some impurities in the Trp containing peptide this was rerun with an acetonitrile gradient at 1%/min.

3. RESULTS AND DISCUSSION

The strategy used in the construction of the vector for overproduction of the RNase H domain of HIV-1 reverse transcriptase is summarized in Fig. 1. Using the host/vector system described, high level expression of the HIV RNase H domain was achieved. The protein constituted about 5–10% of the soluble cytoplasmic *E. coli* proteins (Fig. 2, lane d). The protein was purified to homogeneity (Fig. 2, lane b) using a combination of ion-exchange and gel filtration chromatography. Small amounts of singly deaminated and dimerized protein were removed at the cation ion-exchange and the gel filtration stage, respectively. 1 g wet weight of *E. coli* cells yielded approximately 1.8 mg of purified protein.

Assessment of enzymic activity during the purification was complicated by the presence of *E. coli* RNase H. This was especially problematic as the *E. coli* enzyme exhibits a much higher specific activity than the RNase H activity associated with the HIV-reverse transcriptase (M. Krug and S. Berger, personal communication). RNase H activity was clearly detected up to the cation exchange chromatography stage but was not detected in subsequent stages. *E. coli* RNase H is predicted to be a basic protein (p*I* 8.7) and would, therefore, be expected to bind fairly tightly to the cation exchanger under the conditions used for the purification. The removal of RNase H activity at this stage is consistent with most of the measured activity being due to contaminating *E. coli* enzyme. Using modified conditions for the enzyme assay, the RNase H domain protein appears to have some detectable, albeit, very low activity. This activity does not stem from trace amounts of contaminating *E. coli* RNase H on the basis of kinetic and stability measurements. This work will be described in detail elsewhere (M. Krug and S. Berger, personal communication). A possible cause for the low specific activity of the isolated HIV-1 RNase H domain may be the requirement for the N-terminal 51 kDa portion of the reverse transcriptase for efficient binding of the substrate. Attempts to demonstrate substrate bin-

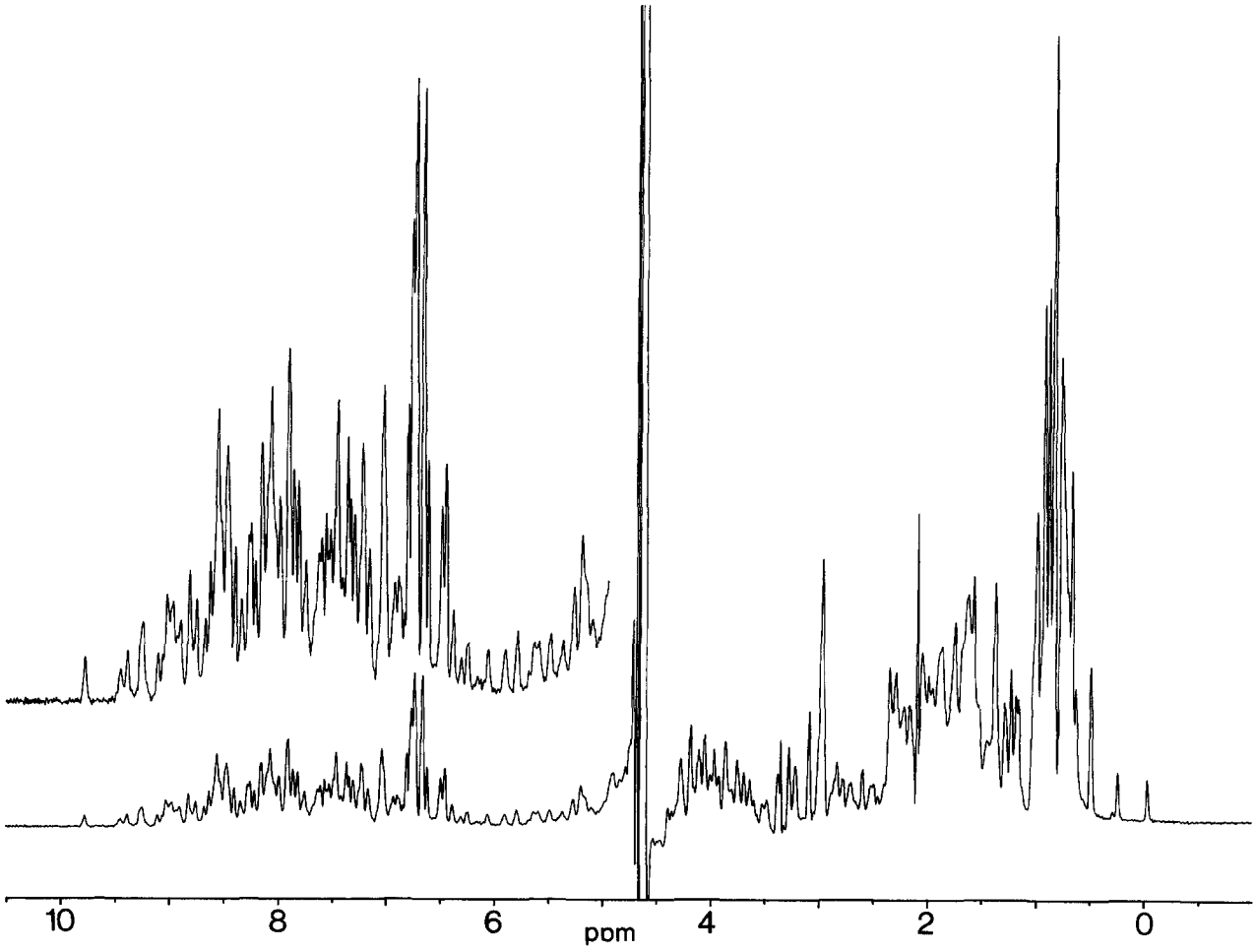

Fig. 3. 600 MHz ^{1}H NMR spectrum of recombinant HIV-1 RNase H. The spectum was recorded at 35°C on a sample containing 0.6 mM protein in 50 mM sodium phosphate, pH 7.7 dissolved in 90%H_2O/10% D_2O. The upper trace shows the aromatic region of the spectrum plotted with a four-fold magnification.

ding to the HIV-1 RNase H domain alone using U.V. induced cross linking have so far been unsuccessful (S. Stahl, unpublished results).

The molecular mass estimated by SDS-PAGE (Fig. 2, lane b) is about 15 000, consistent with that predicted from the cDNA sequence (15 200). Under native conditions the protein appears to be monomeric as judged by gel filtration experiments (data not shown). The isoelectric point of the protein (pI ~5.3) is also close to that predicted from the cDNA sequence. The first cycle of N-terminal sequencing indicated that methionine, derived from initiating *N*-formyl methionine, had not been processed. This is not surprising as the second residue (Asn) has a relatively large radius of gyration [17]. Further sequencing (19 cycles) confirmed the sequence of the first 20 amino acid residues.

Due to the detection of only very low levels of enzymic activity, it was important to establish that the HIV RNase H domain protein had a stable folded three-dimensional structure and did not exhibit a random coil conformation. As is illustrated in Fig. 3 the protein gives rise to a ^{1}H NMR spectrum expected for a folded protein. This is evidenced by the excellent chemical shift dispersion, both for the amide proton region as well as for the aliphatic protons. In particular, it is likely that the protein contains a high amount of β-sheet structure as a large number of well resolved resonances just downfield of the water peak are visible between 5 and 6 ppm. These commonly arise from downfield shifted C$^\alpha$H protons in β-sheets.

Thus, it is clear that the purified protein has a compact and well defined tertiary structure. An estimate of the secondary structure using the far ultraviolet circular dichroic spectrum (Fig. 4) indicates a high β-sheet content. The protein appears to contain 15–20% α-helix and 60–68% β-sheet. This is consistent with the ^{1}H NMR spectrum which was recorded at a 10-fold higher protein concentration. The far ultraviolet circular dichroic spectrum of the HIV-RNase H domain is qualitatively similar to that recently reported for the *E. coli* RNase H [18]. Similar to the HIV protein, the *E. coli* RNase H(M_r = 18 000) exhibits a substantial β-sheet content (41%). The CD evidence combined with the known sequence similarities and conservation of critical residues adds further weight to the view that the main chain fold of these two proteins is similar.

HIV-1 reverse transcriptase is initially produced in virus infected cells as a polyprotein. Upon processing by the *pol* encoded viral protease, it is found as a 66 kDa and a 51 kDa polypeptide which form a heterodimer [19]. A similar p51/p66 heterodimer can be prepared from purified p66 reverse transcriptase by α-chymotrypsin cleavage [9]. As the coding sequence used in our construction of the RNase H domain encompasses several amino acids preceding the proteolytic cleavage site around residues 430–440 of the HIV-1 reverse transcriptase [9], we proceeded to use our

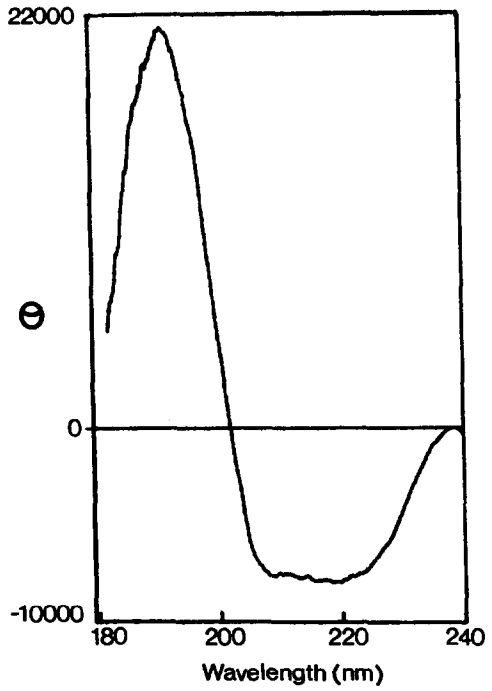

Fig. 4. Far ultraviolet circular dichroic spectrum of HIV-1 RNase H. The spectum was recorded at 25°C as described in section 2. The mean residue ellipticity θ has the unit: deg · cm^2 · dmol^{-1}.

purified domain as a substrate for the HIV protease. Treatment of the RNase H domain with purified HIV-protease generated two main peptides. A Tyr containing peptide eluting at 35% acetonitrile had the N-terminal sequence: YVDGAANRETKLGK. This sequence indicated cleavage between Phe-440 and Tyr-441 (Fig. 1b). This result was corroborated whilst this paper was in preparation, when it was reported that in vitro HIV-protease cleavage of homodimeric 66 kDa HIV-1 reverse transcriptase also results in C-terminal cleavage at Phe-440, generating a 66/51 kDa heterodimer [20].

A second minor Trp containing peptide was generated by protease treatment of the RNase H domain protein (see Materials and Methods). N-terminal sequence analysis of this peptide revealed the following sequence: LAWVPAHKGIGGN. This result indicates protease cleavage between Tyr-532 and Leu-533 (Fig. 1b). Processing of the reverse transcriptase at this site has not been previously reported. It is of interest to note, however, that a purified C-terminal deletion mutant of reverse transcriptase, terminating with Pro-537 has no detectable RNase H activity, whereas reverse transcriptase activity is not affected (S. Stahl and P. Wingfield, unpublished data).

4. CONCLUSIONS

Using the methods described, large amounts of the reverse transcriptase RNase H domain can be produced

as a soluble monomeric protein in *E. coli*. The purified recombinant protein appears conformationally homogeneous and well suited for structural studies by NMR spectroscopy, as well as for crystallization attempts. The determination of the HIV-protease processing site will allow the production of a minimal length RNase H domain and this work is currently underway.

Acknowledgements: We thank Dr. Joseph Shiloach for the bacterial fermentations and Dallas Lee for technical assistance. This work was supported by the Intramural AIDS Targeted Program of the Office of the Director of the National Institutes of Health. A.R.K. thanks the Swedish Medical Research Council for their support.

REFERENCES

[1] Kotewicz, M.L., Sampson, C.M., d'Alessio, J.M. and Gerard, G.F. (1988) Nucleic Acids Res. 16, 265–277.
[2] Tanese, N. and Goff, S.P. (1988) Proc. Natl. Acad. Sci. USA 85, 1777–1781.
[3] Repaske, R., Hartley, J.W., Karlick, M.F., O'Neill, R.R. and Austin, J.B. (1989) J. Virol. 63, 1460–1465.
[4] Schatz, O., Cromme, F.V., Grüninger-Leitch, F. and LeGrice, S.F.J. (1989) FEBS Lett. 257, 311–314.
[5] Krug, M.S. and Berger, S.L. (1989) Proc. Natl. Acad. Sci. USA 86, 3539–3543.
[6] Schatz, O., Mous, J. and LeGrice, S.F.J. (1990) EMBO J. 9, 1171–1176.
[7] Schatz, O., Cromme, F., Naas, T., Lindemann, D., Mous, J. and LeGrice, S.F.J. (1990) in: Gene Regulation and AIDS Papas, T.S. (ed.), Portfolio, Houston.
[8] Johnson, M.S., McClure, M.A., Feng, D.-F., Gray, J. and Doolittle, R.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7648–7652.
[9] Lowe, D.M., Aitken, A., Bradley, C., Darby, G.K., Larder, B.A., Powell, K.L., Purifoy, D.J.M., Tisdale, M. and Stammers, D.K. (1988) Biochemistry 27, 8884–8889.
[10] Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal, F. (1985) Nature 313, 277–384.
[11] Crowl, R., Seamans, C., Lomedico, P. and McAndrew, S. (1985) Gene 38, 31–38.
[12] Allet, B., Payton, M., Mattaliano, R.J., Gronenborn, A.M., Clore, G.M. and Wingfield, P.T. (1988) Gene 65, 259–268.
[13] Crouch, R.J. (1974) J. Biol. Chem. 249, 1314–1316.
[14] Wingfield, P.T., Payton, M., Tavenier, J., Barnes, M., Shaw, A., Rose, K., Simona, M.G., Demczuck, S., Williamson, K. and Dayer, J.M. (1986) Eur. J. Biochem. 160, 491–497.
[15] Provencher, S.W. and Gloeckner, J. (1981) Biochemistry 20, 33–37.
[16] Boutelje, J., Karlström, A.R., Hartmanis, M.G.N., Holmgren, E., Sjogren, A. and Levine, R.L. (1990) Arch. Biochem. Biophys. (in press).
[17] Serman, F., Steward, J.W. and Tsunassasawa, S. (1985) Bioassays 3, 27–31.
[18] Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., Ohtsuka, E. and Ikehara, M. (1990) J. Biol. Chem. 265, 4615–4621.
[19] Di Marzo Veronese, F., Copeland, T.D., DeVico, A.L., Rahman, R., Oroszlan, S., Gallo, R.C. and Sarngadharan, M.G. (1986) Science 231, 1289–1291.
[20] Graves, M.C., Meidel, M.C., Pan, Y.-C.E., Manneberg, M., Lahm, H.-W. and Grüninger-Leitch, F. (1990) Biochem. Biophys. Res. Comm. 168, 30–36.