Purification, characterization and crystallization of recombinant HIV-1 reverse transcriptase

Ramagauri Bhikhabhai, Thorleif Joelson, Torsten Unge, Bror Strandberg, Thomas Carlsson and Seved Lövgren

Department of Molecular Biology, Uppsala University, Box 590, S-751 24 Uppsala (Sweden)

ABSTRACT

The pol I gene from HIV-1 encoding the protease, reverse transcriptase (RT) and endonuclease has been expressed in *Escherichia coli*. By modifying the fermentation conditions and developing a new purification scheme, the yield of purified RT has been increased substantially compared with that obtained in an earlier procedure. The expressed RT was purified to homogeneity by ammonium sulphate fractionation followed by chromatography on DEAE Sepharose, Heparin Sepharose, S Sepharose and Poly(A)-Sepharose. The purified HIV-RT is a heterodimer (p66/p51) with an isoelectric point close to 8 and with a tendency to aggregate. The proteolytic product (p51), corresponding to the N-terminal end of the RT molecule, was isolated and identified, as were also some bacterial polypeptides that co-elute with HIV-RT during the early stages of the purification. The heterodimer was crystallized in several morphological forms using the vapour-diffusion hanging drop technique. To concentrate the protein and to change the buffer for crystallization, reverse-salt-gradient chromatography and micropreparative columns were used. The best crystals diffracted to 9 Å resolution. The best crystals of native RT diffracted to 9 Å resolution and in complex with nucleic acids to 4.5 Å resolution (using a rotating anode X-ray source).

INTRODUCTION

Reverse transcriptase (RT) is an RNA-dependent DNA polymerase (E.C. 2.7.7.49) [1]. This enzyme is a potential target for drugs designed to block retroviral infections, notably human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS) (for reviews, see refs. 2–4). RT is present in a few copies per retrovirus particle. Because of the small amount of protein and for security reasons, RT has been cloned and expressed in different organisms: *Escherichia coli* [5–7], *Bacillus subtilis* [8], yeast [9] and vaccinia virus [10]. The recombinant proteins have been shown to be identical or nearly identical with the viral RT with respect to size, composition and activity [11].

The HIV-RT is stable as a heterodimer. The two polypeptides have calculated molecular weights of 64 346 (p66) and 51 229 (p51), respectively. The p66 and p51 polypeptides have identical N-terminal sequences and p51 is the product of a cleavage of p66 into p51 and a small peptide p15, with the RNaseH activity [12,13]. The detailed three-dimensional (3-D) structure of the RNaseH polypeptide has been determined [14]. The interaction between p51 and p66 is much stronger than that between the two polypeptides of the homodimers [15]. The homodimer p51/p51 has polymerase activity but it is lower than those of the heterodimer and the homodimer p66/p66 [16]. This fact, along with site-directed mutagenesis data [3,4,17], localizes the polymerase activity to the N-terminal part of the RT molecule.

HIV contains single-stranded RNA as the genomic material. On infection, RT copies the RNA into a complementary DNA strand, using tRNA$_3^{Lys}$ as a primer [18]. The viral RNA is then degraded by the RNaseH domain of RT. The RT also has a DNA-dependent DNA polymerase activity, which

Correspondence to: Dr. R. Bhikhabhai, Department of Molecular Biology, Uppsala University, Box 590, S-751 24 Uppsala, Sweden.

is used to make a complementary strand to the single-stranded DNA [1].

To date, only one 3-D structure of a polymerase is known at high resolution, namely, that of the Klenow fragment [19], which is *E. coli* DNA polymerase I without the domain responsible for the 5'-3' exonuclease activity. A low-resolution 3-D structure, nominally at about 4 Å resolution, is available for phage T7 RNA polymerase [20]. HIV-RT and DNA polymerase I belong to different polymerase families with very low sequence homology. However, sequence alignments of a great number of polymerases from different families [21] have revealed a few conserved amino acids. On the basis of other experiments, some of these amino acids have turned out to be residues of vital importance for the catalytic acitivity. This is the case for the dNTP-binding amino acid residues LYS 758 and TYR 766 and for the probable metal-ion-binding residues ASP 705 and ASP 882 of the Klenow fragment [22]. These amino acids must therefore have nearly the same topological and 3-D positions in the Klenow fragment and in the HIV-RT. This could hardly happen without an appreciable similarity in tertiary structure between these parts of the two polymerase molecules.

Despite these similarities, we still need a detailed 3-D structural model of HIV-RT in order to design HIV-RT inhibitors with optimum binding and specificity. The necessary prerequisite for a good 3-D model is the availability of high-quality crystals of the enzyme. Lowe *et al.* [23] and Lloyd *et al.* [24] have reported morphologically different forms of HIV-RT crystals which diffract to 7–9 Å resolution. In our laboratory we have expressed the pol I gene in *E. coli* (the pol I polypeptide is processed to protease, RT and endonucleoase; RT is further processed to the heterodimer p66/p51) and we have reported earlier the expression, purification and crystallization of the heterodimer [25].

With the aim of producing larger amounts of RT and crystals of improved quality, we have substantially modified the fermentation conditions and the purification scheme, resulting in a reasonably high yield and very high purity. The isoelectric point has been estimated by chromatofocusing. The chromatographic properties of RT seem to change on association with other polypeptides or itself. Micropreparative columns were used to concentrate the protein and to carry out buffer exchanges for the crystallization experiments. We describe here numerous crystallization experiments and many different morphological forms of RT crystals. So far, these crystals diffract to about the same resolution as those reported earlier [23–25].

EXPERIMENTAL

Equipment

Fast Protein Liquid Chromatography (FPLC®) System, SMART System and PhastSystem™ were obtained from Pharmacia LKB Biotechnology (Uppsala, Sweden).

Materials

All chromatographic and electrophoretic gel media were obtained from Pharmacia LKB Biotechnology. Electrophoresis gels were PhastGel gel media. Chromatographic gels were DEAE-Sepharose Fast Flow, Heparin Sepharose CL-6B, Sepharose 4B and Poly(A)-Sepharose 4B.

The following prepacked columns were used: Hi-Load 26/10 S Sepharose HP, Mono Q HR 5/5, Phenyl Superose HR 5/5, HiTrap heparin (1 ml), Superdex 75 HR 10/30, HiLoad 16/10, Phenyl Sepharose HP, Mono S HR 5/5, Mono P HR 5/20 and Superose 12 HR 10/30. The micropreparative columns used were Phenyl Superose PC 1.6/5 and Mono S PC 1.6/5.

Microconcentrators were purchased from Amicon. Chemicals were obtained from Sigma and Merck and were of analytical-reagent grade.

Expression of RT in E. coli

RT was expressed in *E. coli* strain N4830-1 essentially as reported earlier [25], with a few modifications. A 1.4-kilobase DNA fragment containing a kanamycin resistance gene [26] was inserted into the Pst I site of the β-lactamase gene of pN10E-15. With the resulting plasmid pN10EK, N4830-1 was grown with 50 mg l^{-1} of kanamycin in a 100-l fermenter. The richer media, 2 × YT, was used instead of Luria broth (LB) [27].

Amino acid analysis

Before amino acid analysis, the sample was dialysed thoroughly against water and a spectrum was recorded over the range 230–340 nm. The RT was

hydrolysed for 24 h at 110°C in 6 M HCl containing 2 mg ml^{-1} of phenol. The hydrolysates were analysed using an LKB Alpha Plus amino acid analyser.

The amino acid sequence obtained from the DNA sequence was used to calculate the number of tryptophan and tyrosine residues. Tryptophan was also determined by determining the molar absorptivity of the protein in connection with the amino acid analyses and using the molar absorptivities of tyrosine and tryptophan at neutral pH. This was done to compare the experimental and calculated molar absorptivities.

Determination of protein concentration

Protein concentrations were determined either by the Bradford assay technique (Bio-Rad Labs.) or spectrophotometrically at 280 nm. The calculated molar absorptivity at 280 nm is $2.52 \cdot 10^5$ l mol^{-1} cm^{-1}.

Analysis by electrophoresis

Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE), native PAGE and isoelectric focusing (IEF) were performed with the PhastSystem using the appropriate PhastGel gel media. PAGE with acidic buffer and 6 M urea was carried out using the reversed polarity electrode assembly. Two-dimensional electrophoresis was run using either IEF or native PAGE in the first dimension and SDS-PAGE in the second dimension. The protein bands were revealed by the silver staining method with the following modification: 20% trichloroacetic acid was used as the first step for fixation with all kinds of gel media; and washing steps with 10% ethanol and 5% acetic acid were omitted. All PAGE experiments and staining were otherwise done as recommended in the instruction manual for the PhastSystem.

Other analytical methods

Protein sequencing, western blotting using antibodies directed toward RT and assay of RT activity were performed as described by Unge et al. [25]. GENEPRO, a sequence data base (Riverside Scientific Enterprises), was used to identify N-terminally sequenced polypeptides.

Analytical chromatography

Gel filtration. The molecular weight of RT was determined by gel filtration on Superose 12 (30 × 1 cm I.D.) and Superdex 75 columns (30 × 1 cm I.D.) in 50 mM Tris–HCl (pH 8.0)–0.15 M NaCl–1 mM dithiothreitol (DTT). Blue Dextran, thyroglobulin, albumin, ovalbumin, trypsin inhibitor and lysozyme were used as molecular weight markers.

Chromatofocusing. The purified RT [after Poly (A)-Sepharose chromatography] was dialysed against 25 mM ethanolamine–10% glycerol (pH 9.4) and applied to a Mono P column (20 × 0.5 cm I.D.) which was then eluted with 30 ml of 10% Polybuffer 96 (pH 6.0). A partially purified sample of RT [before Poly(A)-Sepharose chromatography] was also chromatofocused using similar elution conditions. The experiments were done at room temperature.

Chromatography on Mono Q. The Mono Q column was equilibrated with buffer D, pH 9.5 (10 mM ethanolamine–10% glycerol–20 mM KSCN–5 mM DTT) and was eluted with a linear gradient from 0 to 0.5 M NaCl in buffer D.

Purification procedure

All chromatographic experiments were performed using the FPLC system at 4°C. The purification scheme was optimized on small columns and the production was done on larger columns as described below. During the early stages of optimization, the enzyme activity assay was used to detect RT in different chromatographic steps. When pure RT was assayed, there was a linear increase in the amount of DNA, measured in terms of radioactive incorporation of ^{3}H dTTP for 24 h. When the activity assay was done on the fractions from the crude extract and from the previous purification steps, the impurities present interfered with the activity measurement. Therefore, it was not possible to evaluate the purification of RT in different steps in terms of specific acitivity. SDS-PAGE was used to detect RT in each chromatographic step. Purified heterodimer of RT was used as the marker to identify fractions containing RT.

Preliminary steps. The bacterial cells from the fermenter were harvested by centrifugation, washed twice with 20 mM Tris HCl–75 mM NaCl–1 mM EDTA (pH 8.2) and then suspended in the lysis buffer (pH 8.2) [20 mM Tris–HCl–75 mM NaCl–5

mM EDTA-5 mM DTT-1 mM phenylmethylsulphonyl fluoride (PMSF)-2 mM benzamidine-1 mM MgSO$_4$-DNase-RNase]. After disruption of the cells using a French pressure cell, the lysate was centrifuged at 23 000 g for 10 min to remove cell debris. The supernatant was subjected to ammonium sulphate fractionation. Saturated ammonium sulphate solution adjusted to pH 8 with NaOH was used. The protein that precipitated with ammonium sulphate in the interval 30-70% saturation was dissolved in buffer A (pH 8.2) (20 mM Tris-HCl-1 mM EDTA-1 mM DTT-2% glycerol) and dialysed for at least 1 h against the same buffer.

Anion-exchange and affinity chromatography in tandem. After dialysis, the conductivity of the solution was adjusted to 1.5 mS by dilution with buffer A. A solution corresponding to 100 g of bacteria was then applied to a column of DEAE Sepharose which was connected in tandem to a column of Heparin Sepharose. After application of the solution the columns were washed with buffer A, the DEAE Sepharose column was disconnected and the Heparin Sepharose column was eluted stepwise with buffer A (pH 7.9) containing 0.15 M NaCl in the first step and 0.35 M NaCl in the second step. The RT-containing fractions eluted in 0.35 M NaCl and were pooled.

Cation-exchange chromatography. The RT pool from the Heparin Sepharose column was diluted and adjusted to pH 6.5 by addition of a solution of 2-morpholinoethanesulphonic acid (MES) and applied to an S Sepharose column pre-equilibrated with buffer B (pH 6.5) (20 mM MES-1 mM DTT-1 mM EDTA-2% glycerol). The column was eluted with a linear gradient from 0 to 0.3 M NaCl in buffer B (total volume = 12 × bed volume). RT elutes at about 0.18 M NaCl.

Affinity chromatography on Poly(A)-Sepharose 4B. The fractions containing the heterodimer RT from the previous step were pooled and dialysed against buffer C, pH 8.2 (20 mM Tris-HCl-10% glycerol-1 mM DTT-1 mM EDTA) and applied to a column of Poly(A)-Sepharose 4B equilibrated with buffer C. Elution was carried out with a linear gradient from 0 to 0.2 M NaCl in buffer C (total volume = 6 × bed volume).

Hydrophobic interaction chromatography. The fractions containing RT from the step above were pooled and the concentration of ammonium sul-

phate was adjusted to 1.5 M and applied to a Phenyl Sepharose HP column equilibrated with 1.5 M ammonium sulphate-50 mM HEPES (pH 7.8)-1 mM DTT-10 mM MgCl$_2$. RT was eluted with a linear gradient from 1.5 to 0 M ammonium sulphate in three column volumes of buffer.

Concentration of RT using different chromatographic methods

The following steps were performed to concentrate the protein further and for buffer exchange. The SMART System, a micropreparative chromatography system, was used for concentration of RT on Phenyl Superose and Mono S. The small size of the the columns (100 μl) and the collection of fractions in small volumes are the most important features of the SMART system. The FPLC System was used for the other columns mentioned below to concentrate the protein.

Phenyl Superose. A Phenyl Superose column (100 μl) was equilibrated with 1 M ammonium sulphate at the desired pH. RT was applied to the column and was eluted directly with a buffer of low ionic strength which contained the appropriate additives for crystallization. RT eluted in small volumes and at the concentrations required for crystallization. When crystallization were performed in the presence of mercury(II) salts or sodium sulphite (in these instances the presence of DTT is not desired), the following steps were carried out to remove DTT while keeping the RT reduced: the sample applied was completely reduced by adding 10 mM DTT; the washing and eluting buffers (without DTT) were made oxygen free by passing nitrogen through the solution. The eluted fractions were used for the hanging drop crystallization experiments.

Heparin Sepharose or Mono S. The pure RT was applied to either a Hitrap Heparin Sepharose column (1 ml) or a Mono S column (100 μl). RT was eluted in the required buffer with a high salt concentration. The peak fractions contained the optimum concentrations of protein and salt and were used directly for crystallization. For example, RT was eluted in 0.75 M ammonium sulphate when 1.5 M ammonium sulphate was used as the precipitant.

Reverse salt gradient chromatography on Sepharose 4B [28]. In this instance the protein was precipitated on Sepharose 4B packed into an HR 5/5 column (1 ml) which was equilibrated with 1.5 M

ammonium sulphate. RT was eluted with a linear gradient of six column volumes from 1.5 to 0 M ammonium sulphate. The saturated solution of RT was used directly for crystallization. Similar reverse salt gradients were used to obtain the proper concentrations of other precipitants for the crystallization experiments.

Crystallization

RT was concentrated to 10 mg ml^{-1} either in a Centricon microconcentrator with a molecular weight cut-off of 10 000 or by the chromatographic techniques described above. Crystallization was done by the hanging drop vapour diffusion technique [29]. In a typical experiment, a 5 μl drop of RT was mixed with an equal volume containing 1.5 M ammonium sulphate, buffer and various additives such as DTT, MgCl$_2$ and methylmercury acetate. The drop was equilibrated against 1.5 M ammonium sulphate (10 ml) containing the same buffer, in a plastic Petri dish (9 cm diameter) sealed with parafilm. The time for crystallization varied from 2 days to 2 weeks.

X-Ray diffraction

X-Ray diffraction photographs were taken both with rotating anode and synchrotron X-ray sources. In both instances Enraf–Nonius Arndt–Wonacott oscillation cameras were used. The rotating anode generator was an Elliot GX-6 instrument operating at 35 kV and 50 mA. In this instance, the wavelength (λ) was 1.54 Å (Cu Kα), the crystal–film distance 100 mm, the oscillation angle 0.25° and the oscillation time 120 000 seconds per degree (total exposure time ca. 8.3 h). The conditions when using the synchrotron X-ray source at the SERC Synchrotron Radiation source (Daresbury, UK) were λ = 0.895 Å, crystal–film distance 145 mm, oscillation angle 1.0° and oscillation time 1000 seconds per degree (total exposure time ca. 16.7 min).

RESULTS AND DICUSSION

The expression and purification of recombinant HIV-RT have been reported by several laboratories [6–9, 18, 25, 30–32]. RT is expressed in fairly small amounts when the whole HIV-pol I gene is used to express the authentic RT, protease and endonuclease. The low yield might be due to the proteases

TABLE I

VOLUME, CONCENTRATION, YIELD AND PURIFICATION FACTOR OF RT (FROM 250 g WET CELL WEIGHT) AT DIFFERENT STAGES OF PURIFICATION

Purification step	Volume (ml)	Total protein (mg)	Purification factor	
			A[a]	B[b]
Lysate after centrifugation	600	30 000		
Ammonium sulphate preciptate[c]	5500	16 000	2	0.18
Heparin Sepharose	410	410	39	7
S Sepharose	70	37	11	75
Poly(A)-Sepharose	40	28	1.3	100

[a] A = Purification factor of protein between the two steps.

[b] B = Percentage of RT in each step calculated by using the final yield of RT in relation to the amount of total protein after each step.

[c] 30–70% ammonium sulphate precipitate dissolved and diluted.

in E. coli [33] and/or a reduction in translational efficiency due to, for example, secondary structural elements and short half-life of the mRNA [34]. In order to achieve the high homogeneity which is required to obtain well diffracting protein crystals, the purification scheme used earlier in our laboratory [25] was modified. To increase the yield of RT, we also modified the expression plasmid and the medium used in the fermenter as described under Experimental. Using the new protocol, 25–35 mg of RT, purified to homogeneity, were obtained from 200–250 g of cells. The purification scheme consists of salt fractionation and the following chromatographic principles: anion-exchange, affinity and cation-exchange chromatography. The purified RT was concentrated by hydrophobic interaction chromatography.

We have been using this scheme for more than one year and many interesting observations concerning the properties of RT have been made. RT has a strong tendency to aggregate and to associate with small polypeptides which seem to change the chromatographic properties of the protein.

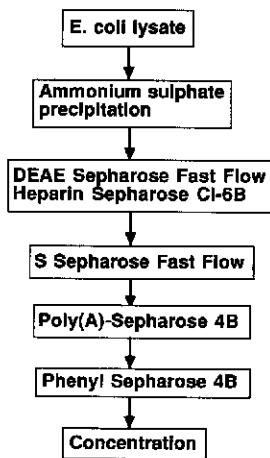Chromatofocusing was performed for the determination of the isoelectric point of p66/p51. For

Fig. 1. Flow diagram for the purification and concentration of HIV-reverse transcriptase.

crystallization of the heterodimer in various buffers, salts and precipitants, micropreparative columns (100 μl) were used both for concentration of the protein and for change of buffers.

Purification of the heterodimer RT (p66/p51)

The details of the purification are described under Experimental. The purifications scheme is summarized in Table I and Fig. 1.

To obtain a high yield and high homogeneity of RT, it is essential to wash the cells thoroughly and to start the ammonium sulphate fractionation immediately after lysis and centrifugation of the cells. If this procedure is not followed, RT seems to become degraded, resulting in a lower yield and a less homogeneous enzyme. This might be attributed to the activity of bacterial proteases or to the HIV-1 protease, which is co-expressed with RT. Many proteins in *E. coli*, being acidic, bind to DEAE Sepharose, whereas the basic RT molecules are recovered in non-adsorbed fractions from DEAE Sepharose. RT binds to Heparin Sepharose at the pH and ionic strength used to run the DEAE Sepharose column. The columns of DEAE and Heparin Sepharose were coupled in tandem in order to reduce the purification time. Approximately 90% of the host contaminant proteins bound to DEAE Sepharose. About 1-2% of the total soluble protein in the lysate was eluted with 0.35 *M* NaCl from the Hepa-
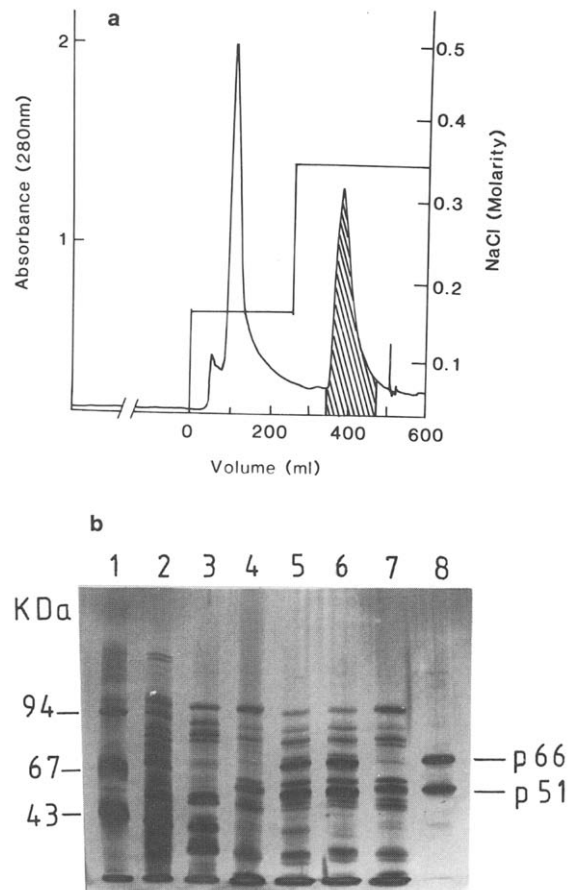

Fig. 2. (a) Affinity chromatography using Heparin Sepharose. Stepwise elution were done at 0.15 *M* NaCl and 0.35 *M* NaCl in buffer A (pH 7.9) (20 m*M* Tris–HCl–2% glycerol–1 m*M* EDTA–1 m*M* DTT). The shaded peak shows the fractions containing RT. Column size: 5 × 5 cm I.D. (b) SDS-PAGE (7.5%) analysis on PhastGel homogeneous 7.5. Samples are: lane 1 = molecular weight marker in kilodalton, (KDa); 2 = starting material applied on DEAE and Heparin Sepharose; 3 = flowthrough fraction of Heparin Sepharose; 4 = 0.15 *M* NaCl pool; 5 and 6 = 0.35 *M* NaCl pool; 7 = 1.0 *M* NaCl pool; 8 = RT as a marker.

rin Sepharose column (Fig. 2a and b). As shown in Fig. 3a and b, RT eluted from the S Sepharose column in a symmetrical peak. In this step, RT was concentrated about 100-fold.

The chromatogram from the Poly(A)-Sepharose column is presented in Fig. 4a. The purified RT from this step had a very high specific acitvity [35,36]. The Poly(A)-Sepharose column could be
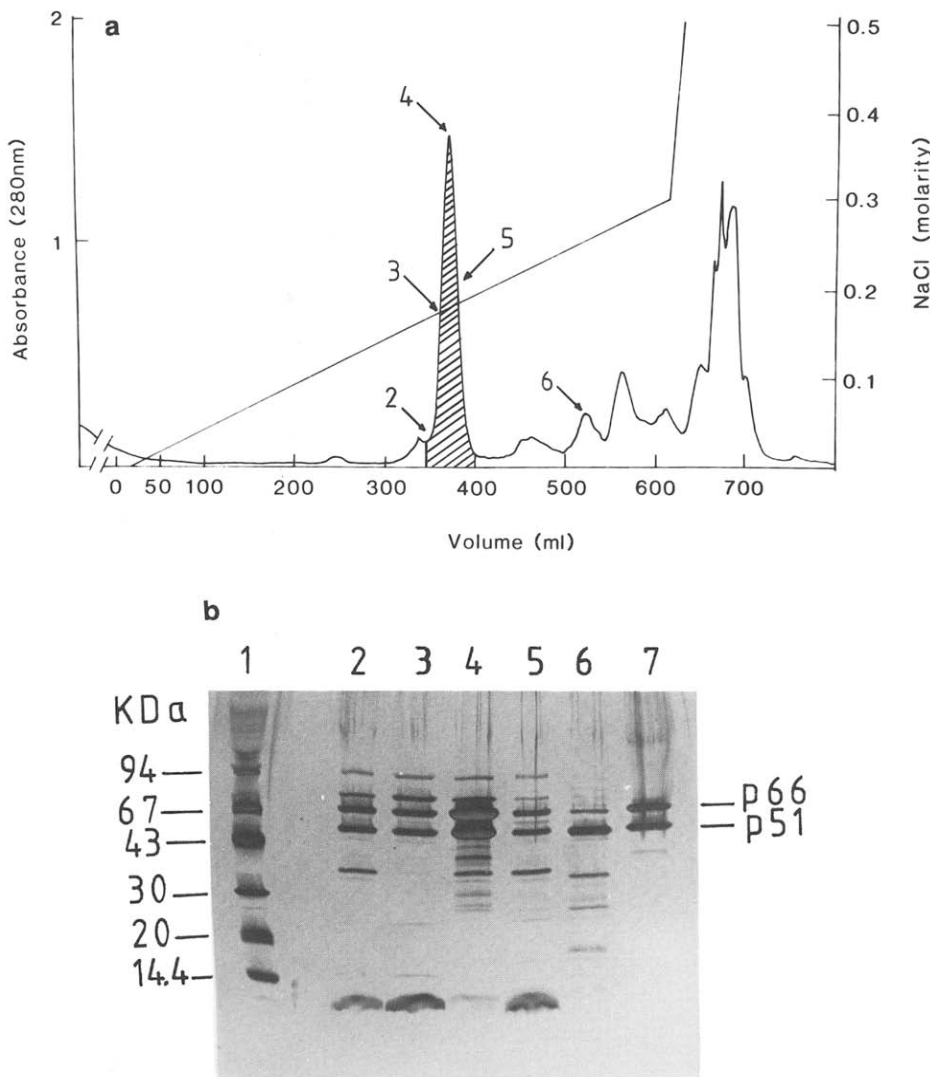
Fig. 3. (a) Cation-exchange chromatography using HiLoad 26/10 S Sepharose HP. Sample, 0.35 *M* pool from Heparin Sepharose; column size, 26 × 10 cm I.D.; fraction volume, 10 ml. Elution was performed with a linear gradient (600 ml) of 0 to 0.3 *M* NaCl in buffer B, pH 6.5 (20 m*M* MES–2% glycerol–1 m*M* EDTA–1 m*M* DTT). The shaded peak shows the fractions containing the heterodimer RT. (b) SDS-PAGE (8–25%) analysis on PhastGel gradient 8–25. Samples are: lane 1 = molecular weight markers; 2–6 = fractions indicated by arrows 2–6, respectively, in Fig. 3a; 7 = RT as a marker.

used for four or five runs. The reduced binding capacity might be due to hydrolysis of the poly r(A). For further concentration of the protein a Phenyl Sepaharose column was used. SDS-PAGE of purified RT from the peak fraction revealed a few faint bands of molecular weights between 40 000 and 50 000 in the silver-stained gel (Fig. 4b). When the samples were diluted 10-, 50- and 100-fold these contaminating bands disappeared. We calculate that the protein must be at least 99% pure. Analysis of purified RT, by PAGE in acidic buffer and 6 *M* urea, revealed two sharp bands. As seen from the dilution series of RT on SDS-PAGE, the ratio between p66 and p51 is very close to unity.
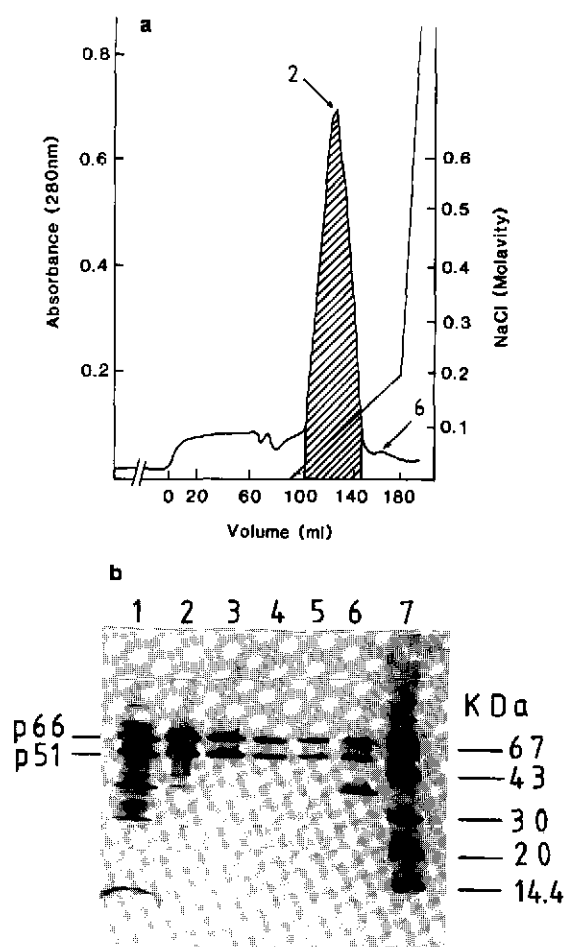
Fig. 4. (a) Elution profile from Poly(A)-Sepharose. Sample, dialysed RT fractions from S Sepharose; column size, 10 × 1.6 cm I.D.; fraction volume, 5 ml. Elution was performed with a linear gradient (six bed volumes) from 0 to 0.2 M NaCl in buffer C'(pH 8.2); 20 mM Tris–HCl–10% glycerol–1 mM EDTA–1 mM DTT. The peak containing RT is shaded. (b) SDS-PAGE (8–25%) analysis on Phastgel gradient 8–25. Samples in lane 1 = starting material; 2 = peak fraction indicated by arrow 2 in the chromatogram; 3, 4 and 5 = 10 ×, 50 × and 100 × dilutions of the peak fraction, respectively; 6 = fraction indicated by arrow 6 in the chromatogram; 7 = molecular weight marker.

Further analysis of polypeptide p51

The heterodimer RT eluted in a symmetrical peak from the S Sepharose column. The other peak fractions from S Sepharose were analysed on SDS-PAGE (Fig. 2b). The p51 polypeptide, along with other small polypeptides, were spread out over the

chromatogram. The chromatographic pattern on S Sepharose reflected the degree of proteolytic degradation. The relative heights of the peaks containing p51 increased when (i) instead of direct lysis after the cell harvest, the bacterial cells were frozen before the purification, and (ii) the time taken to reach this purification step happened to be unusually long. The p51 band was identified as part of RT as it reacted positively in western blot analysis using antibodies raised against a synthetic peptide with the N-terminal sequence of RT. The N-terminal sequence analysis of p51 confirmed its identity as the N-terminal part of p66. An attempt to purify p51 further either by anion-exchange chromatography on Mono Q or by gel filtration on Superdex 75 resulted in fractions containing the p51 band co-eluting with a few small polypeptides. The p51 polypeptide bound strongly to Phenyl Superose and could only be eluted with 6 M urea. This indicates that p51, in comparison with the heterodimer, displays new hydrophobic surfaces.

Characterization of RT

There are no disulphide bridges between p66 and p51, as SDS-PAGE under non-reducing conditions resulted in two distinct bands. The aparent molecular weight of the heterodimer is 110 000 as determined by gel filtration experiments. The N-terminal amino acid sequences of p66 and p51 are identical, as shown by N-terminal sequence determination. The amino acid composition of the purified heterodimer corresponds to the sum of those of p66 (1–560) and p51 (1–440) and to that reported by Mizrahi et al. [6]. The molar absorptivity (ε) obtained from the experimental data agreed with the calculated value ($\varepsilon = 2.52 \times 10^5$ l mol^{-1} cm^{-1}).

Isoelectric point of RT

Although the purification of recombinant RT has been reported by several groups, none has determined the isoelectric point. The isoelectric point of the protein was calculated from the pK_a values of the amino acids contributing to the charge. For both p66 and p51, the calculated isoelectric point is 9.1. To determine the pI experimentally, pure RT was analysed on a IEF gel in the pH range 3–9. The protein did not focus in a sharp band but gave a smear in the pH range 6.5–8. The same result was obtained in the presence of 6 M urea. The partially
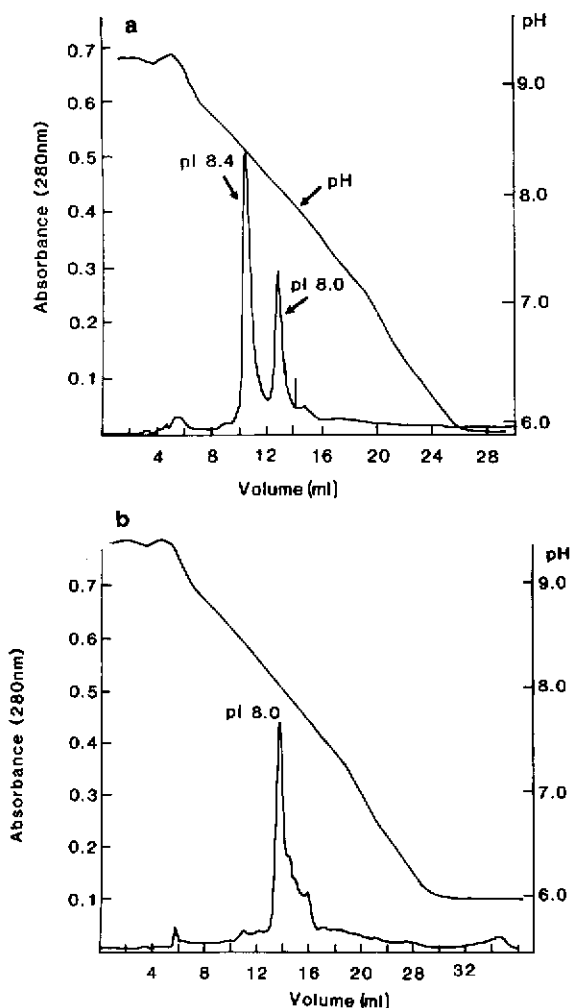
Fig. 5. Chromatofocusing on Mono P 5/20. The column was equilibrated with 25 mM ethanolamine–10% glycerol (pH 9.4). A pH gradient from 9.4 to 6.0 was created by eluting the column with 10% Polybuffer 96 (pH 6.0). (a) Sample: purified RT [after Poly(A)-Sepharose chromatography]. (b) Sample: partially purified RT [before Poly(A)-Sepharose chromatography].

purified RT was analysed by two-dimensional electrophoresis. In the first dimension on IEF gels, a patch and a number of bands between pH 5 and 9 were seen. The second dimension using SDS-PAGE revealed that the heterodimer was present in all of the bands.

As the results from IEF could not be interpreted,

chromatofocusing was attempted on a Mono P column to determine the exact pI. Purified RT showed a tendency to elute in a broad pH range as in isoelectric focusing gels. When DTT was added to the sample solution, two sharp peaks at pH 8.0 and 8.4 were observed (Fig. 5a) with some trailing of RT in the pH range 7.0–8.5. Several chromatofocusing runs with different concentrations of RT were performed using identical elution conditions. These experiments showed that the smearing of the protein was not due to overloading. The ratio between the two peaks, which appeared in all runs at pH 8.0 and 8.4, varied among the experiments. When these peaks were re-run on the chromatofocusing column under similar conditions, RT was found to be evenly distributed over a broad pH range. Dynamic light scattering investigations of these fractions showed aggregates of different sizes (results not shown). It was not feasible to perform chromatofocusing in 6 M urea because RT precipitated in 6 M urea at pH 9.4, the starting pH of the column. When partially purified RT (the fractions containing RT from the S Sepharose chromatographic step) was chromatofocused, a distinct peak eluted at pH 8.0 and very little protein was found elsewhere in the chromatogram (Fig. 5b). Furthermore, the elution of RT from the ion-exchange column in a single symmetrical peak precludes charge heterogeneity. The data above indicate that the true isoelectric point is close to 8. It seems likely that the broadening of the peak in chromatofocusing of the pure RT is caused by aggregation.

Polyacrylamide gel electrophoresis

The electrophoretic mobility of RT heterodimers in native PAGE at pH 8.8 was very high compared with that of many other proteins of similar size. This is surprising, as PAGE was performed at a pH close to the isoelectric point. Two diffuse bands or smears appeared when the gel was stained. Two-dimensional electrophoresis with native PAGE in one dimension and SDS-PAGE in the second dimension revealed that both bands contained the heterodimer, p66/p51. This behaviour of RT in both IEF and native electrophoresis gels is probably due to aggregation. PAGE in acidic buffer containing 6 M urea resulted in two sharp bands as in SDS-PAGE.

TABLE II

N-TERMINAL AMINO ACID SEQUENCES OF THE POLYPEPTIDES (1–5) WHICH CO-ELUTED WITH RT IN THE FIRST THREE CHROMATOGRAPHIC STEPS AND WERE ISOLATED ON POLY(A)-SEPHAROSE

The approximate sizes of the polypeptides were estimated from SDS-PAGE.

No.	Contaminating peptides
1	MW 20 000
	[56]Lys–Ala–Asn–Arg–Val–Thr–Lys–Pro–Glu–Ala–Gly–X–Phe...
	Ribosomal protein L3 *E. Coli*
2	MW 16 000
	[3]Lys–Lys–Asp–Ile–His–Pro–Lys–Tyr–Glu...
	Ribosomal protein L31 *E. coli*
3	MW 42 000
	Ala–Glu–Glu–Met–Leu–Arg–Lys–Ala–Val–Gly–Lys–X–Ala–Tyr–Gly...
	Unidentified
4	MW 23 000
	Phe–Ser–Ile–Asp–Asp–Val–Ala–Lys–Gln–Ala...
	Unidentified
5	MW 14 000
	Met–X–Pro–Met–Leu–Asn–Ile–Ala–Val...
	Unidentified
6	MW 12 000
	Met–Asn–Lys–Thr–Gln–Leu–Ile–Asp–Val–Ile...
	DNA-binding protein NS2 *E. coli*

Investigation of contaminating polypeptides

During optimization of the conditions for the Poly(A)-Sepharose chromatography, wherein RT was eluted with a linear gradient (ten column volumes) from 0 to 0.5 *M* NaCl, polypeptides of low molecular weight were eluted together with the heterodimer RT. Often, one of the contaminating polypeptides was dominant (Fig. 4b, lane 6), but the molecular weight of the contaminant varied between 12 000 and 45 000 in different preparations. In no case were the polypeptides degradation products of RT, as shown by western blot analysis. To eliminate the contaminating polypeptides, the elution on the Poly(A)-Sepharose column was done using a shallow gradient as described under Experimental. Under these conditions the polypeptide eluted immediately after the heterodimer RT. The N-terminal amino acid sequences of these polypeptides from a few preparations (Table II, 1–5) were determined. Two of these polypeptides were identified as parts of the ribosomal proteins L3 and L31, whereas for the others no homologous sequences were found in the data base. These ribosomal proteins L3 [37] and L31 [38] are basic and L3, at least, binds to the ribosomal RNA. The calculated iso-electric points of these proteins are similar to that of RT. This may explain the co-elution of these polypeptides in all the purification steps.

The fractions containing the heterodimer RT from the S Sepharose chromatography were analyzed further on a Mono Q column. All of the fractions were analysed by SDS-PAGE under non-reducing conditions. Most of the heterodimer RT adsorbed to the column at pH 9.5. However, the nonbound fractions showed three bands on the PAGE gel, two of which coincided with the heterodimeric marker (p66 and p51). The third polypeptide was identified by the N-terminal amino acid sequence determination as the DNA binding protein NS2 from *E. coli*. When SDS-PAGE was run without boiling the samples, two bands occurred very close to each other at a position around MW 66 000 and the small peptide was not observed. One explanation for this observation might be that the association of the NS2 protein with RT might prevent the binding of RT to Mono Q at pH 9.5.

Crystallization

Several groups have attempted to crystallize RT but so far none has reported crystals diffracting to
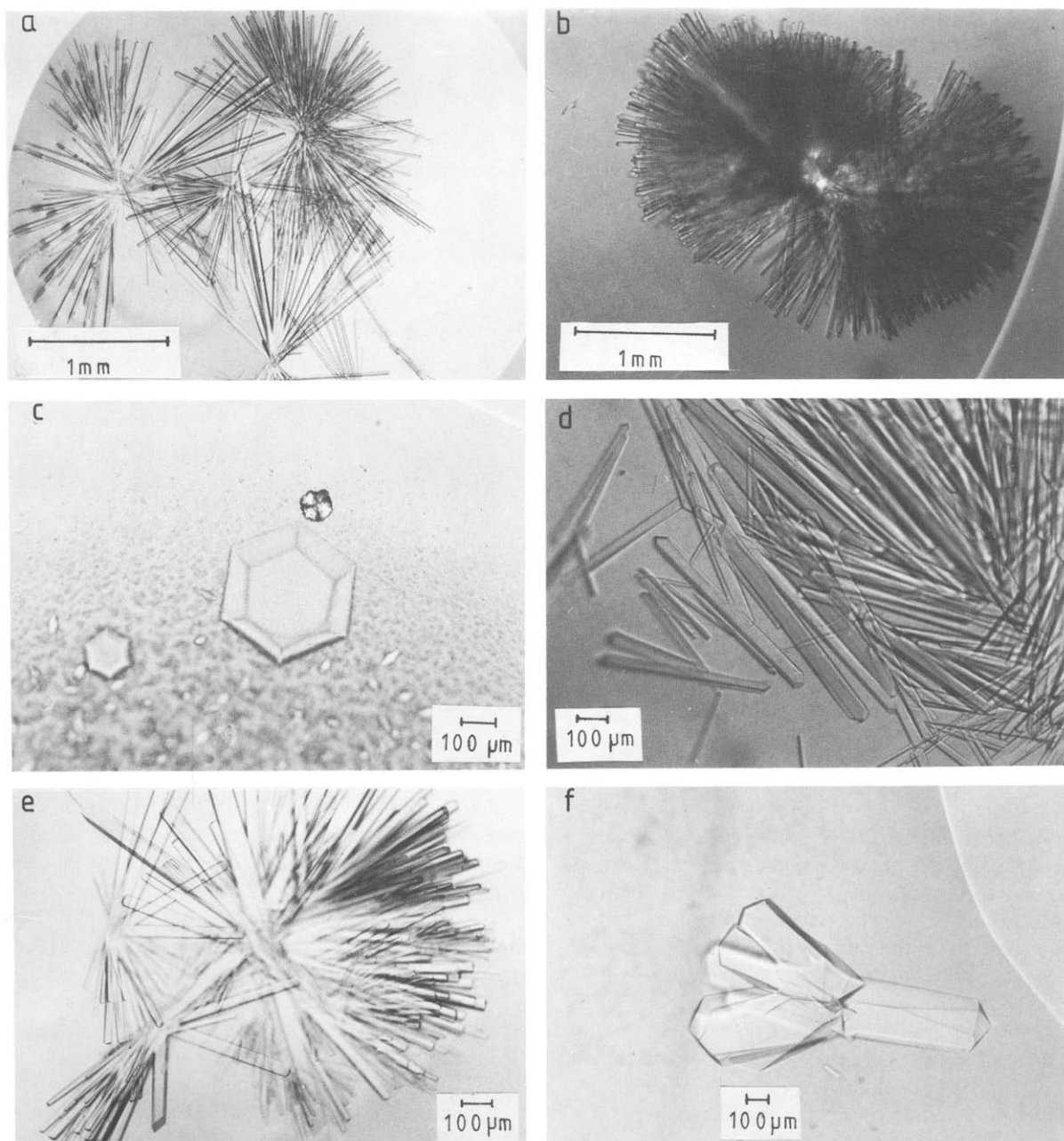
Fig. 6. Photographs showing different morphological forms of some of the crystals obtained in the laboratory. Crystallization was done by the hanging-drop vapour diffusion method with 1.5 M ammonium sulphate–50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid (HEPES)–10 mM MgCl$_2$–1 mM DTT–50 mM NaCl (pH 7.8) as precipitant at 4°C and with protein concentration of 10 mg ml^{-1}. Size calibration is given by bars in each photograph. (a) Needles and rods. These are typical crystals which appeared within 3–5 days. The material was obtained directly from those experiments where small columns were used for concentration and buffer exchange. (b) Crystals forming sea urchins were obtained under similar conditions as in (a). (c) Truncated hexagonal pyramid crystals grown with 100 mM mercaptoethanol instead of 1 mM DTT in the buffer. Spindle-shaped crystals appeared first. Some of them were transformed into hexagonal pyramids. (d) Rod-shaped crystals obtained with the addition of 50 mM sodium pyrophosphate. Similar crystals could be obtained by the addition of phosphate, guanidine sulphate, spermidine and spermine (10–100 mM). (e) Crystals obtained in the presence of a synthetic deoxyoligonucleotide (TGC AGG GCA CGT). In this instance the rods had more defined ends. (f) Crystals obtained in the presence of 1 mM p-chloromercuribenzenesulphonic acid.
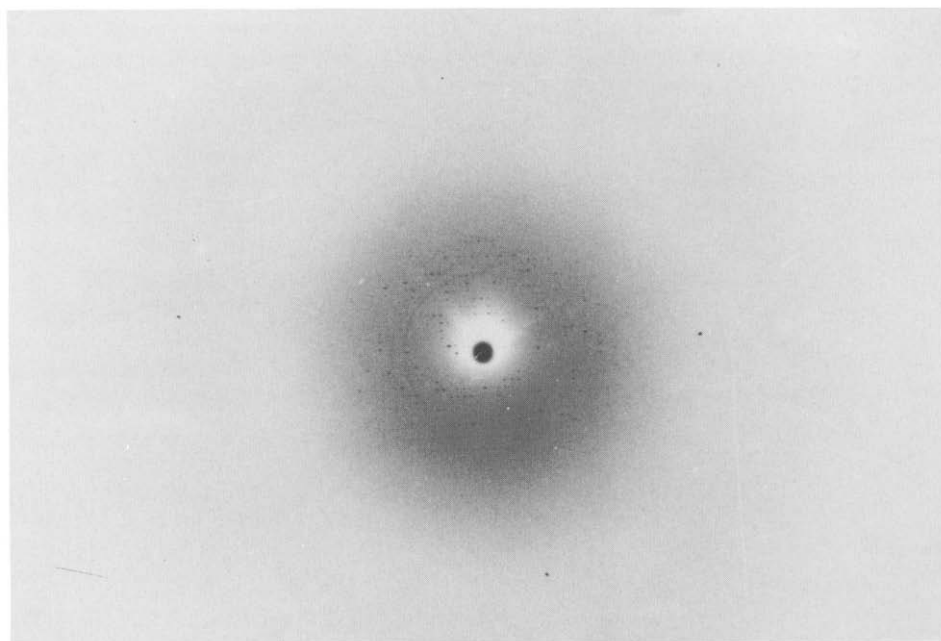
Fig. 7. X-ray diffraction pattern from a crystal of heterodimer RT. The X-ray photograph was obtained by the use of Enraf–Nonius Arndt–Wonacott camera and an X-Ray synchrotron source (at the SERC Synchrotron Radiation source, Daresbury, UK) with the following conditions: wavelength (λ), 0.895 Å; crystal–film distance, 145 mm; oscillation angle, 1.0°C; and oscillation time, 1000 s per degree (total exposure time ca. 16.7 min). The diffraction limit is about 9 Å resolution and the X-ray reflections are well suited for intensity measurements.

better than about 7 Å resolution [23–25]. We investigated a variety of crystallization conditions, e.g., different pH, ionic strength and temperature, and also stabilization with synthetic deoxyoligonucleotides and inhibitors. Attempts to co-crystallize RT with some of the polypeptides listed in Table II did not lead to better crystals. Further, we used micropreparative columns to exchange buffers and increase the RT concentrations at the same time. The reverse salt gradient chromatographic technique on Sepharose 4B [28] was used to produce saturated solutions of RT for crystallization. The reverse salt gradient method, when performed using different precipitants, not only gave a saturated solution of the protein but also enabled us to determine the appropriate concentration of the precipitant in the reservoir. The crystals appeared within two days when ammonium sulphate was used as the precipitant.

Fig. 6 shows a number of different morphological forms of RT crystals grown in our laboratory. The growth conditions are described in the legend. In most instances the crystals started to grow as needles or rods. The growth of the crystals terminated after some time. Instead, more needles appeared or needles started to branch off, resulting in feathery fans or sea urchin structures. In some instances the crystals had the shape of well defined truncated hexagonal pyramids. These hexagonal crystals sometimes disappeared and needles were formed instead. Both crystal forms could also occur in the same drop.

The best heterodimer crystals (Fig. 6e and f) diffracted to about 9 Å resolution, with both rotating anode and synchrotron X-ray sources (Fig. 7).

NOTE ADDED IN PROOF

Recently we have obtained crystals of heterodimer RT in complex with nucleic acid which diffract to 4.5 Å resolution (using a rotating anode X-ray source).

SUMMARY

The HIV-1 pol I gene proteins (protease, reverse transcriptase and endonuclease) expressed in *E. coli* by the use of an inducible expression vector have been reported by Unge *et al.* [25]. Authentic reverse transcriptase (HIV-RT) was purified in a stable heterodimeric form (p66/p51) exhibiting a high degree of purity and high activity. By modifiying the fermentation conditions and developing a new purification scheme, the yield of purified RT has been substantially increased compared with that obtained in our earlier procedure [25]. The purification procedure consists of ammonium sulphate fractionation followed by chromatography on DEAE Sepharose, Heparin Sepharose, Poly(A)-Sepharose and Penyl Sepharose.

The heterodimer has an isoelectric point close to 8 (as determined by the chromatofocusing technique) and shows a tendency to aggregate (as indicated by dynamic light scattering).

Some bacterial polypeptides, *e.g.*, the ribosomal proteins L3 and L31 and the DNA binding protein NS2, co-eluted with HIV-RT during the early stages of purification, indicating similar properties (L3 and L31) and, in the case of NS2, association with the heterodimer.

A proteolytic product of the heterodimer was separated and identified as the protein p51 (the N-terminal end of the p66), which appears to be more hydrophobic than the heterodimer.

The HIV-RT has been crystallized in different morphological forms using the hanging drop vapour diffusion technique. The use of reverse salt gradient chromatography and micropreparative columns resulted in increased control of the crystallization process.

With the aim of obtaining crystals of higher quality, we are devoting considerable efforts towards improving the expression systems, modifying the RT molecule, achieving complex formation between RT and Fab fragments of monoclonal antibodies and other measures.

ACKNOWLEDGEMENTS

REFERENCES

1 H. Varmus and R. Swanstrom, in R. A. Weiss, N. M. Teich, H. E. Varmus and J. M. Coffin (Editors), *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 381–384.

2 L. A. Evans and J. A. Levy, *Biochim. Biophys. Acta*, 989 (1989) 237–254.

3 S. H. Hughes, A. L. Ferris and A. Hizi, in W. G. Laver and G. M. Air (Editors), *Use of X-Ray Crystallography in the Design of Antiviral Agents*, Academic Press, New York, 1990, pp. 297 307.

4 D. K. Stammers, K. L. Powell, B. A. Larder, G. Darby, D. J. M. Purifoy, M. Tisdale, D. M. Lowe, D. I. Stuart, E. Y. Jones, G. L. Taylor, E. F. Garman, R. Griest and D. C. Phillips, in W. G. Laver and G. M. Air (Editors) *Use of X-Ray Crystallography in the Design of Antiviral Agents*, Academic Press, New York, 1990, pp. 309–317.

5 B. Müller, T. Restle, S. Weiss, M. Gautel, G. Sczakiel and R. S. Goody, *J. Biol. Chem.*, 264 (1989) 13975–13978.

6 V. Mizrahi, G. M. Lazarus, L. M. Miles, C. A. Meyers and C. Debouck, *Arch. Biochem. Biophys.*, 273 (1989) 347–358.

7 M. R. Deibel Jr., T. J. McQuade, D. P. Brunner and W. G. Tarpley, *AIDS Res. Hum. Retrovirus*, 6 (1990) 329–340.

8 S. F. J. Le Grice, V. Beuck and J. Mous, *Gene*, 55 (1987) 95–103.

9 I. C. Bathurst, L. K. Moen, M. A. Lujan, H. L. Gibson, P. H. Feucht, S. Pichuantes, C. S. Craik, D. V. Santi and P. J. Barr, *Biochem. Biophys. Res. Commun.*, 171 (1990) 589–595.

10 V. Karacostas, K. Nagashima, M. A. Gonda and B. Moss, *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 8964–8971.

11 W. G. Farmerie, D. D. Loeb, N. C. Casavant, C. A. Hutchison, III, M. H. Edgell and R. Swanstrom, *Science*, 236 (1987) 305–308.

12 J. Hansen, T. Schulze, W. Mellert and K. Moelling, *EMBO J.*, 7 (1988) 239–243.

13 M. C. Graves, M. C. Meidel, P. E. Yu-Ching, M. Manneberg, H.-W. Lahn and F. Gruninger-Leitch, *Biochem. Biophys. Res. Commun.*, 168 (1990) 30–36.

14 J. F. Davies, II, Z. Hostomska, Z. Hostomsky, S. R. Jordan and D. A. Matthews, *Science*, 252 (1991) 88–90.

15 T. Restle, B. Müller and R. Goody, *J. Biol. Chem.*, 265 (1990) 8986–8988.

16 V. R. Prasad and S. P. Goff, *Proc. Natl. Acad. Sci. U.S.A.*, 86 (1989) 3104–3108.

17 D. M. Lowe, V. Parmar, S. D. Kemp and B. A. Larder, *FEBS Lett.*, 282 (1991) 231–234.

18 M.-L. Sallafranque-Andreola, D. Robert, P. J. Barr, M. Fournier, S. Litvak, L. Sarih-Cottin and L. Tarrago-Litvak, *Eur. J. Biochem.*, 184 (1989) 367–374.

19 D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong and T. A. Steitz, *Nature (London)*, 313 (1985) 762–766.

20 Y. J. Chung, R. Sousa, J. P. Rose, E. Lafer and B. C. Wang, in F. Y.-H. Wu and C.-W. Wu (Editors), *Structure and Function of Nucleic Acid and Proteins*, Raven Press, New York, 1990, pp. 55–59.

21 M. Delarque, O. Poch, N. Tordo, D. Moras and P. Argos, *Protein Eng.*, 3 (1990) 461–467.

22 C. M. Joyce, *Curr. Opinion Struct. Biol.*, 1 (1991) 123–129.

23 D. M. Lowe, A. Aitken, C. Bradley, G. K. Darby, B. A. Larder, K. L. Powell, D. J. M. Purifoy, M. Tisdale and D. K. Stammers, *Biochemistry*, 27 (1988) 8884–8889.

24 L. F. Lloyd, P. Brick, L. Mei-Zhen, N. E. Chayen and D. M. Blow, *J. Mol. Biol.*, 217 (1990) 19–22.

25 T. Unge, H. Ahola, R. Bhikhabhai, K. Bäckbro, S. Lövgreen, E. M. Fenyö, A. Honigman, A. Panet, G. S. Gronowitz and B. Strandberg, *AIDS Res. Hum. Retrovirus*, 6 (1990) 1297–1303.

26 J. Vieira and J. Messing, *Gene*, 19 (1982) 259–268.

27 T. Maniatis, E. F. Fritsch and J. Sambrook, *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 3rd ed., 1989.

28 K. Gulewicz, D. Adamiak and M. Sprinzl, *FEBS Lett.*, 189 (1985) 179–182.

29 A. J. McPherson, *Methods Biochem. Anal.*, 23 (1979) 249–345.

30 A. Hizi, C. McGill and S. H. Hughes, *Proc. Natl. Acad. Sci. U.S.A.*, 85 (1988) 1218–1222.

31 B. Larder, D. Purifoy, K. Powell and G. Darby, *EMBO J.*, 6 (1987) 10 3133–3137.

32 S. F. J. Le-Grice and F. Grüninger-Leitch, *Eur. J. Biochem.*, 187 (1990) 307–314.

33 D. E. Danley, K. F. Geoghegan, K. G. Scheld, S. E. Lee, J. R. Merson, S. J. Hawrylik, G. A. Rickett, M. J. Ammirati and P. M. Hobart, *Biochem. Biophys. Res. Commun.*, 165 (1989) 1043–1050.

34 P. Balbas and F. Bolivar, *Methods Enzymol.*, 185 (1990) 14–37.

35 J. S. Gronowitz, M. Neümuller, J. Lennerstrand, R. Bhikhabhai, T. Unge, H. Weltman and C.. F. R. Källander, *Biotechnol. Appl. Biochem.*, 13 (1991) 127–142.

36 M. Neümüller, A. Karlström, C. Källander and S. Gronowitz, *Appl. Biochem.*, 12 (1989) 34–56.

37 P. O. Ollins and M. Nomura, *Cell*, 26 (1981) 205–211.

38 J. Brosius, *Biochemistry*, 17 (1978) 501–507.