AD_____

CONTRACT NUMBER: DAMD17-93-C-3122

TITLE: Isolation and Preliminary Characterization of a Recombinant TAT Protein from Human Immunodeficiency Virus

PRINCIPAL INVESTIGATOR: Linchun H. Walls, Ph.D.

CONTRACTING ORGANIZATION: Morgan State University
Baltimore, MD 21239

REPORT DATE: September 1996

TYPE OF REPORT: Final

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Final (24 May 93 - 23 Aug 96)

4. TITLE AND SUBTITLE
Isolation and Preliminary Characterization of a Recombinant TAT Protein from Human Immunodeficiency Virus

5. FUNDING NUMBERS
DAMD17-93-C-3122

6. AUTHOR(S)
Li chun H. Walls, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Morgan State University
Baltimore, MD 21239

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

A regulatory protein, Tat, from Human Immunodeficiency virus (HIV-1) was purified in large scale quantities. Tat was overexpresssed as a fusion protein, Rop-Tat, washed and sonicated to eliminate most *E. coli* contaminants, and Rop-Tat was then cleaved by cyanogen bromide. Tat, after incubation with ß-mercaptoethanol, was purified to near homogeneity by column chromatography. Tat was eluted from 0.5 M and 1 M salt of the S-Sepharose column. The 0.5 M elute was resolved into different molecular weight fractions by Sephacryl S-200 size columns. The 1 M elute contained Tat aggregates. Spectroscopic, fluorescence and circular dichroism studies were done. Indications are that Tat binds to Tar-RNA in a one to one ratio. The major species from purification steps has the smallest size, about 25,000 Daltons by size column. Refolding in the presence of Cd(II) resulted in one peak of about 40,000 Da. Refolding in the presence of Zn(II) resulted in two peaks. Both protein size and conformation are factors in protein **mobility** on the size column, so the 25 kD may be the same size as the 40 kD species. Mass spectrometry indicates the 25 kD species has a 9.7 kD size.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Human Immunodeficiency Virus (HIV) Regulatory Protein Tat	24
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRAC
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Lelan H. Wells 11/12/96

PI - Signature Date

TABLE OF CONTENTS

Summary

A regulatory protein, Tat, from Human Immunodeficiency virus (HIV-1) has been purified in large scale quantities. Tat was overexpressed as a fusion protein, Rop-Tat, and was extensively washed and sonicated to eliminate most *E. coli* contaminants. Rop-Tat was then cleaved by cyanogen bromide. The Tat protein, after incubation with ß-mercaptoethanol [BME], was purified to near homogeneity by fast flow Q and S-Sepharose column chromatography. Tat was eluted from 0.5 M and 1 M salt of the S-Sepharose column. The 0.5 M elute was further resolved into two to three different molecular weight fractions by Sephacryl S-200 size columns. The 1 M elute contained Tat aggregates. Possible causes for the different molecular sizes may be related to the oxidation of cysteine (Tat has 7 cysteine residues). The major species is the peak with the smallest molecular size, which has a molecular weight of about 25,000 Daltons judging by gel filtration. Refolding of this 0.5 M S-Sepharose elute in the presence of Cd(II) resulted in one peak with a molecular weight of about 40,000 Da by gel filtration. However, refolding in the presence of Zn(II) resulted in two peaks from the gel filtration column. Although the size of the protein plays a major factor in the **mobility** of the protein on the gel filtration column, the conformation of the protein also affects its mobility on the column. Therefore, the 25 kDa may have the same size as the 40 kDa species. Results obtained from the mass spectrometer indicate that the 25 kDa molecular weight species from the 0.5 M elute of the S-Sepharose actually has a 9.7 kD size. Spectroscopic, fluorescence and circular dichroism studies were done. Indications are that Tat binds to Tar-RNA in a one to one ratio.

Introduction

HIV-1 encodes several trans-activating regulatory proteins. One of these proteins, Tat (trans-activator of transcription), acts in trans to control viral gene expression (1). Tat interacts with the cis-acting element, TAR, which is located in the viral long terminal repeat (LTR), and is essential for virus replication (2,3). Tat is reported to be an RNA-binding protein, binding to TAR RNA rather than to TAR DNA (4). Tat is composed of 86 amino acids, is a nuclear protein, and contains at least three distinct functional domains (5). These are (a) the amino-terminal end which might participate in Tat's transactivation function, (b) the basic region of the protein (8 lysines and arginines), which may be involved in the binding of TAR RNA in addition to serving for nuclear localization (5), and (c) the cysteine-rich region between residues 22 and 37 which participates in Tat transactivation in addition to possible involvement in metal-linked dimer (6). Although there is a wealth of information regarding Tat's transactivation function, very little is known about the protein's secondary and tertiary structure. For example, is the protein a zinc protein? Is it a dimer or a monomer? Frankel (6) reported that Tat is a zinc-dimer. Rice *et al.*, (7) and others reported that the Tat protein is a monomer when expressed. Tat protein can be considered an promising target for antiviral drug design. Sim (8) notes that the Tat protein has a pivotal role in the emergence of HIV from the latent state, and suggests that antagonists of Tat may be able to suppress viral replication in blatantly infected cells while at the same time allowing them to function normally.

The fusion protein Rop-Tat (sequence shown in Figure 1) used in this paper was developed by Huang *et al.* (9). In this system the Rop-Tat protein accumulates in large quantity as membrane-bound inclusion bodies. The inclusion bodies are then extensively washed and sonicated to eliminate most *E. coli* contaminants. Rop-Tat is then cleaved by cyanogen bromide. The Tat protein is then purified by ion-exchange and size columns. The purified protein may be used for spectroscopic studies.

Materials - All buffers and cyanogen bromide are from Sigma Chemical Co. Q and S-Sepharoses and Sephacryl S-200 are from Pharmacia Biotech. All amino acid sequence analyses were done commercially.

Procedure:

I. Preparation of Tat Protein

1. Inclusion Bodies (IB) Preparation

About 10-20 grams of over-expressed cell culture (9) is suspended in 100-200 ml of 0.05 M tris, 50 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 8% sucrose, 5% triton X-100, 5 mM ß-mercaptoethanol at pH 8. 50-100 mg of lysozyme is added and the solution is incubated on ice for 2 hours. 400 ml of the same triton X-100 buffer is added and sonicated for 10 minutes (600 watt sonicator with 3/4" high gain probe, at 50% output).

500 ml of the same buffer is added (total buffer volume 1 liter) and centrifuged at 10,000 rpm for 30 minutes (Sorvall GSA rotor).

The pellet is resuspended in 1 liter volume of the above triton X-100 buffer, sonicated in the same manner as above, and the suspension is centrifuged at 10,000 rpm for 30 minutes.

The pellet is suspended in 1 liter of the same buffer except that triton X-100 is omitted. The suspension is sonicated in the same manner as above, and centrifuged in the same way.

The pellet is resuspended in the same 1 liter, no triton X-100 buffer and sonicated at 100% output (an increase from 50%) for this last washing for 10 minutes and centrifuged.

2. Rop-Tat Protein Preparation from the Inclusion Bodies(IB)

The pellet(IB) is dissolved in 150-200 ml 6 M urea, 0.05 M tris, 5 mM ß-mercaptoethanol pH 7.8. It is then sonicated with 25% output for 10 minutes. The solution is then centrifuged at 12,000 rpm for 20 minutes. The supernatant solution is assigned as the first urea solution. An aliquot of this solution is used to run gel electrophoresis.

200 ml of the above urea, tris solution is added to the pellet of the first urea solution above. The solution is sonicated with 100% output for 10 minutes and centrifuged and assigned as the second urea solution.

200 ml of urea buffer is added to the above pellet, sonicated and centrifuged, and assigned to be the third solution.

200 ml of urea buffer is added to the third pellet, sonicated for 30 minutes with 100 % output and centrifuged. This solution is used for the Tat preparation below, the fourth solution.

Solid ammonium sulfate is added to the fourth urea solution for a final concentration of 65%. The solution is centrifuged at 12,000 rpm for 20 minutes after sitting overnight in the refrigerator.

3. Cyanogen Bromide Cleavage of Rop-Tat

The ammonium sulfate precipitated pellet is dissolved into 75 ml (the absorbance should not exceed 2 for a 1 to 10 fold dilution) 6 M guanidine hydrochloride, 50 mM sodium phosphate, 5 mM EDTA pH 2.

1.7 grams (this amount is determined by multiplying the protein O.D. reading by a factor of 2) of cyanogen bromide, 1.57 ml concentrated HCl (final HCl concentration is 0.25 M) is flushed with nitrogen gas for 20 minutes.

After gentle stirring for 3 to 3½ hours, the solution is incubated in a dark bottle for a total of 24 hours at room temperature.

The solution is then transferred into an evaporating dish and flushed with nitrogen gas for one hour to get rid of the cyanogen bromide.

After adding solid tris to raise the pH to neutral, the solution is dialyzed against 1 liter of 6 M urea, 0.05 M tris pH 7.8 with one change of the buffer solution.

4. Isolation of Tat Protein by Ion-Exchange Column Chromatography

A. Q-Sepharose Anion Exchange Column Chromatography

After two days of dialysis, the solution was made to a final concentration of 0.6% ß-mercaptoethanol, flushed with nitrogen and incubated at 56°C for 3 hours. The sample was then immediately applied onto a 2.5 x 12 cm fast flow Q-Sepharose column previously equilibrated

with 6 M urea, 0.05 M tris pH 7.8. The column was eluted batchwise with increasing NaCl concentration in the 6 M urea, 0.05 M tris pH 7.8 buffer.

B. S-Sepharose Cation Exchange Column Chromatography

The runthrough fraction from the above Q-Sepharose was applied onto a 2.5 x 12 cm fast flow S-Sepahrose column previously equilibrated with 6 M urea, 0.05 M sodium acetate, pH 5.0 buffer. The column was eluted batch wise by using increasing concentration of NaCl in equilibrated buffer as indicated on flow chart on Figure 2.

C. Sephacryl S-200 Size Column Chromatography

The 0.5 M NaCl elute from S-Sepharose column was concentrated by ultrafiltration using an 5 kD membrane Amicon Ultrafiltration device to 5 ml and applied onto a 2 x 42 cm Sephacryl S-200 column. Alternatively, to determine the size of the protein, the sample was applied onto a 1.5 x 42 cm Sephacryl S-200 HR calibrated with protein standards. Both colums were run in 2 M urea, 0.05 M sodium acetate, 1 M NaCl at pH 5.0 buffer. The standards were also dissolved into the same buffer. The column buffer for the Sephacryl S-200 column was later changed to 0.5 M arginine at pH 7.5 in order to avoid using the denaturant, urea.

5. Alternate procedure to prepare Cd(II)-Tat:

The cultured cells were sonicated according to the above mentioned sonication process except the buffer was changed to 50 mM sodium acetate, 100 mM NaCl, 5 mM BME, 2 mM cadmium chloride, 5% triton X-100, 8% sucrose, 0.1 mM PMSF at pH 5. The cells were then washed with same buffer without triton.

The pellet after centrifugation was directly dissolved into one of two solutions, either (1) 6 M guanidine hydrochloride, 50 mM sodium phosphate at pH 2, with HCl and cyanogen bromide added as above, or (2) the pellet was dissolved into 70% formic acid with addition of 2 fold weight (to protein) of cyanogen bromide. The solution was incubated at room temperature for 22-24 hours. After incubation, the solution was flushed with nitrogen for 1 hour. Water was then added and the solution was lyophilized to dryness. The protein was dissolved in 6 M urea, 0.05 M tris at pH 7.8, and dialyzed against the same buffer for 48 hours. The dialysis solution was applied directly to Q-Sepharose without incubation with BME at 56°C. The unbound fraction was applied to S-Sepharose as above.

FPLC (Pharmacia Biotech) system of the Superose 12 column was used to determine the molecular weight of the two fractions from the Sephacryl S-200 column SDS gel electrophoresis, and amino acid sequence determination was used to determine the purity of the Tat protein.

II. Refolding of Tat after either the S-Sepharose or Sephacryl S-200 Columns

Three different methods of refolding were used as follows:

1. The Christie et al. (10) procedure is followed, briefly described here. Tat protein from 0.5 M elute of S-Sepharose column was either precipitated from urea buffer, or the protein solution is dialyzed against 10 mM HCl and then lyophilized to dryness. The protein is then dissolved into 8 M urea, 0.1 M DTT to a final concentration of 5 mg/ml. The pH is adjusted to about 8 by adding tris, and the protein solution is incubated at 56°C for 2 hours. The solution is then passed through a Sephadex G-25 column (1.5 x 45 cm) previously equilibrated with 10 mM HCl, which eliminates the excess DTT. The protein peak from the Sephadex G-25 column is gradually diluted during the course of 3 hours to make a final solution with a concentration of 0.016 M sodium phosphate at pH 6.6, 2 mM $MgCl_2$, and 40 μM $ZnCl_2$. After shaking the solution 3 hours in the buffer, it is then made to 1 mM DTT and incubated at 37°C for 1 hour. Tat is then dialyzed for a few days against 20 mM tris at pH 7, 50 mM NaCl, 2 mM $MgCl_2$, 0.1 mM DTT, and 40 μM $ZnCl_2$ under nitrogen. The folded protein was then concentrated and run on a non-denaturing gel (11) electrophoresis for basic proteins and also on a Sephacryl S-200 column.

2. This method involves using an oxidative regeneration system in the presence of oxidized and reduced glutathione. Either Peak 2 or Peak 3 of Tat from Sephacryl S-200 column (Figure 4-A) is dialyzed against 10 mM HCl and the lyophilized to dryness. The protein is then dissolved into 8 M urea, 0.1 M DTT, 0.1 M tris, pH 8 and incubated at 56°C for two hours. The solution is diluted 1 to 100 fold into a buffer containing 0.1 M tris, 0.1 mM EDTA, 0.5 mM oxidized glutathione and 1.0 mM reduced glutathione. The solution is concentrated by Amicon Ultrafiltration and the sample is applied onto a 1.5 x 42 cm Sephacryl HR-200 column (column running buffer is 2 M urea, 0.05 M sodium acetate, 1 M NaCl pH 5).

3. . The protein is refolded in the presence of Zn(II) and Cd(II). This procedure is also used to prepare the Zn(II), Cd(II)-Tat proteins. The 0.5 M elute and also the 1 M elute from S-Sepharose column are concentrated by Amicon ultrafiltration to an absorbance at 278 nm of about 1 unit. Solid guanidine chloride is added to make a final concentration of 6 M. 1 M tris pH 9 buffer is added to make a final concentration of 0.1 M tris. The solution is then made to 0.1 M DTE and incubated at room temperature for 2-4 hours. The solution is then acidified to pH 3-4 by using dilute hydrochloride acid. It is dialyzed against 6 M guanidine hydrochloride, 500 μM cadmium chloride (or zinc chloride) pH 5 for 24 hours against one change of buffer. The dialysis **solution** is changed to 0.5 M arginine at pH 7.5, 40 μM cadmium chloride (or zinc chloride), 1 mM cystamine, 5 mM cysteamine. An aliquot (250 μl) is withdrawn at 24 hour intervals and chromatography done on a Superose 12 HR 10/30 gel filtration (Pharmacia Biotech.) in 0.5 M arginine pH 7.5 with a flow rate of 0.5 ml/min.

III. Biochemical Characterization of Tat Protein by Circular Dichroism and Binding Study of Tar RNA to Tat by Fluorescence Spectroscopy

1. Gel Electrophoresis

SDS-PAGE was carried out according to standard Laemmli procedures using 15% gels (12). Unless otherwise stated, a two-fold concentrated sample buffer was added to an equal volume of sample. The samples were then boiled for 5 min prior to loading. The standards (Pharmacia) were phosphorylase (97,000), bovine serum albumin (66,000), ovalbumin (43,000), carbonic anhydrase (30,000), trypsin inhibitor (20,100), and ∝-lactalbumin (14,400). The native PAGE discontinuous buffer of basic proteins for running in an acidic pH on a 7% gel was also run according to the standard procedures (11).

2. Circular Dichroism Spectroscopy

Far UV CD measurements were made on a Jasco-720 spectropolarimeter operating at 25°C (circulating water bath) in a 1 mm path length cell. The spectra were recorded with a time constant of 2 s and a spectral band width of 1.0 nm. Scan speed was 20 nm/min and step resolution was 0.1 nm. A total of seven scans were acquired and were averaged. The estimated percentages of secondary structure were calculated from the CD spectra according to the manufacturer's program based upon the method of Yang et al. (13). All protein samples were dialyzed against 10 mM sodium phosphate, 40 μM $CdCl_2$, pH 7.0 for Cd(II)-Tat. Other non-cadmium samples were dialyzed in 25 mM sodium phosphate pH 7 buffer and were diluted with water to 10 mM sodium phosphate before CD measurement.

3. Fluorescence Spectroscopy

Study of Tar-RNA binding to Tat by using Fluorescene Spectroscopy. The Tat binding to Tar-RNA is limited to the region around the bulge area (1,14). Since it was reported that the region from +19 to +42 of Tar-RNA is the minimal domain required for transactivation (15), the 27 nucleotides from +18 to +44 Tar-RNA (the number is relative to the start of transcription from the HIV long-terminal repeat) is to be used for the binding experiment. The RNA, 5'p-GGAGAUCUGAGCCUGGGAGCUCUCUCC (the two CG base pairs at the base of the stem have been reversed to GC (+18 and +19 CC changed to GG and +26 and +29 changed to CC from GG, to improve stability of RNA) (14), was commercially synthesized by National Biosciences, Inc. (NBI). It was purified by PAGE electrophoresis. The procedure for binding this Tar-RNA to the Tat protein follows McKenna et. al. (16), briefly described here. The Pk 3 of Tat protein from Sephecryl S-200 of Figure 4 -A is concentrated to an absorbance at 278 nm of 0.3 and then dialyzed against a buffer solution of 0.05 M sodium acetate pH 4, 100 μM $ZnCl_2$, 10 mM DTT. 0.4 ml of this dialyzed protein solution is added to 0.4 ml of a buffer solution consisting of 10 mM potassium phosphate pH 7, 100 mM NaCl. The Tar-RNA is preincubated in 10 mM potassium phosphate, 100 mM NaCl, pH 7 at 80°C for 15 minutes. An absorbance number of 1.3 is used as an equivalence for 1 mg/ml of Tat protein (17). An amount of Tat protein is then added to the Tar-RNA in its 0.4 ml solution to make up an equal 1:1 molar ratio of the two components. 0.4 ml of Tat protein solution is added to 0.4 ml of Tar-RNA solution and the mixture is incubated for 30 minutes at room temperature. This solution is then dialyzed under nitrogen against buffer consisting of 0.05 M sodium acetate pH 4, 100 μM $ZnCl_2$, 10 mM DTT overnight. The fluorescence spectroscopy was performed with a Perkin-Elmer LS-50B

Luminescence spectrometer equipped with a xenon lamp. Excitations are at 280 and 295 nm. Slit width was 2.5 nm. To determine the stoichemical ratio of binding between Tar-RNA and Tat protein, titration is performed by adding a 211.2 μg/ml (0.024 μmole/ml) of Tar-RNA in a 0.1 ml volume into a 0.8 ml of Tat protein (with an absorption of 0.3) sample followed by reading the emission spectrum (excitation at 295 nm). The process is repeated with another 0.1 ml of Tar-RNA until a total volume of 1 ml has been added.

Results

I. Preparation of Tat Protein

The over-expressed Rop-Tat protein (Lane 1, Figure 5) in the inclusion bodies of *E. coli* cell pellets is purified to about 80% homogeneity prior to cyanogen bromide cleavage by washing and sonicating with a large volume of buffer containing Triton X-100, followed by buffer without Triton X-100 (Lanes 2 of Figure 5). The effectiveness of the washing procedure probably is due to *E.coli* contaminant proteins being considerably more soluble in the washing buffers than the Rop-Tat fusion protein.

The Tat protein was separated from Rop protein (after cyanogen bromide cleavage) by Q-Sepharose (see Figure 3A). The flow-through fraction contains the Tat protein (Lane 10, Figure 5) and the 0.25 M elute has the Rop protein (Lane 7, Figure 5). Note that since the Rop has an internal methionine residue, Rop protein will lack the first 11 residues from the N-terminal end due to cyanogen bromide action (Figure 1). The 0.1 M elute of Q-Sepharose (Figure 3A) appears to be slightly below the Rop-Tat fusion protein (see Lane 6 of Figure 5). This Rop-Tat protein appears to be a different species from the Rop-Tat fusion protein obtained from over-expressed *E. coli*. The fusion protein Rop will have Rop as its N-terminal residue. This 0.1 M elute has both Tat and Rop as its N-terminal amino acids, shown by N-terminal sequence analysis of this fraction. This indicates that the 0.1 M elute with Rop attached to Tat is resistant to both cyanogen bromide and BME cleavage. The attachment between these two residues is neither a methionine nor a disulfide bond.

From the results of SDS gel electrophoresis and N-terminal amino acid sequence analysis, the Tat protein is found contained in two fractions, the 0.5 M and 1 M elutes of S-Sepharose (Lanes 5 & 11 of Figure 5). The yield of Tat protein for 10 g of cell culture is about 20 mg for the 0.5 M salt elute, and 12 mg for the 1 M elute. However, this amount is only from one of the four urea solutions of the above procedure. Therefore, not counting the first urea solution (which is discarded), the recovery of the Tat protein from a 10 g cell culture will be about 40 mg for the 0.5 M elute of the S-Sepharose. Without the 0.6% BME (56°C, 3 hours), the fraction from 0.5 M elute of S-Sepharose contains equal mixtures of Ala.Arg.Phe.Iso.Arg-(from Rop) and Glu.Pro.Val.Asp.Pro-(from Tat). With the 0.6% BME incubation, the 0.5 M elute only contains Tat at 90-95% pure. The need to incubate the sample with a reducing agent in order to get Tat indicates that the Rop protein is still covalently attached to Tat by disulfide bonds even after cyanogen bromide cleavage. The Rop protein has two cysteine residues, which must participate

in the disulfide linkage with Tat. The 0.5 M elute of S-Sepharose was further resolved into two peaks (Peaks 2 and 3) by Sephacryl S-200 gel filtration (Figure 4A). However, Peak 2 in some preparations was two peaks in Superose 12 FPLC column. Another time Peak 2 was one shoulder in the same column.

Alternate procedure to prepare Cd(II)-Tat:

Attempts to introduce Cd(II) into Tat at the sonication step have mixed results. The process eliminates the 1 M elute of S-Sepharose of Figure 3, indicating no aggregates of Tat were formed. This was achieved without incubation with 0.6% BME at 56°C for 3 hours. Actually, the incubation with BME is responsible for aggregation of Tat (see discussion). However, the 0.5 M elute from the S-Sepharose column of the Cd-Tat protein does have a slight extra band on top of the Tat protein band (see Lane 12 of Figure 5). In order to get rid of this slight extra band, it is necessary to incubate with BME. Even though it may be possible to get Cd-Tat at this step, the later incubation (strong acid and guanidine hydrochloride) condition of cyanogen bromide will probably strip the protein of Cd(II).

II. Refolding the Tat protein in the Presence and Absence of Cd(II) and Zn(II)

Attempts to refold the two peaks (presumabably two different sizes of Tat protein) by using Methods 1 & 2 under Procedure II.1. above resulted in much aggregation. The native gel electrophoresis running refolded sample from Method 1 has only one band (Figure 5B, Lane 1), which is similar to the band in the same sample after incubation with BME for 1 hour (Figure 5B, Lane 2). The resemblance appears to indicate that both bands represent a monomer since BME reduces aggregates to monomers. However, the sample also has a large Pk 1 on the size column, indicating that a large aggregate is present. The apparent discrepancy between results from the size column and the native gel may be explained if a large aggregate in the sample remained on the top of the gel, being too large to gain entry. Figure 4B shows an increase in Pk 1, which may be a large molecular aggregate. Note that the three peaks in Figure 4B are equivalent to the three peaks (Pk 1, 2 & 3) of Figure 4A. Even through the refolding can result in one size of protein, upon concentration the Tat would shift from a smaller to a larger size. In order to prevent this from happening, the refolding process has included Cd(II) and Zn(II). The Cd(II) and Zn(II) were introduced to the 0.5 M elute of S-Sepharose of Tat protein after it was denatured and reduced by guanidine hydrochloride and DTE at basic pH. Then the metals were introduced at acidic pH, and the elute was dialyzed in the presence of cystamine/cysteamine at basic pH. The results indicate that refolding of Tat in the presence of a reducing agent and Cd(II) yields one Tat (one peak, Figure 6A) in the Superose 12 HR 10/30 gel filtration column. Refolding of Tat in the presence of a reducing agent and Zn(II) yields two sizes of Tat or two different conformations of Tat as judged by the size column (see Figure 6B). This result may reflect the fact that the binding affinity of Zn(II) is not strong enough to complete the binding as the cadmium was able to do. The binding affinities of the metal ions are in the order of Cd(II), Zn(II), Co(II). In order to make the Zn(II)-Tat, it may be necessary to make the apo-Tat from the Cd-Tat and then introduce the excess Zn(II) to the apo-Tat. In other words, in order to prepare an

unoxidized Tat protein, it may be necessary to prepare the Cd-Tat first and then replace the cadmium with zinc. Work is underway to study this possibility. The Cd(II) was also introduced to the 1 M elute of S-Sepharose of Tat protein. This isolate of Tat was mostly large aggregates as evidenced by the fact that the fraction before the refolding is mostly Pk 1. After 72 hours of dialysis, the refolded peak is more broad than the 0.5 M fraction after refolding (Superose 12 graph not shown). This may be due to the fact that the 1 M elute, according to the amino acid sequencing data, is not as pure a fraction as compared with the 0.5 M Tat protein fraction. UV spectroscopy has also been run on the Cd-Tat protein. There is an increase in absorption in the 255 nm region after refolding with Cd(II). This is due to the development of S→Cd(II) charge-transfer band (see Figure 8) (6, 18).

III. Physicochemical Property Studies of the Tat Protein

1. Circular Dichroism of the Cd-Tat Protein - The Cd(II)-Tat prepared from the above, after 72 hour dialysis in arginine, cadmium chloride, cystamine/cysteamine is dialyzed into 10 mM sodium phosphate, 40 μM cadmium chloride. The concentration of the protein is about 400 μg/ml. The result (Figure 9) shows an α-helix of 9.6%, 0% turn, 54.9% β-sheets and 35.5% random. Preliminary results from CD measurement of Tat before refolding with Cd(II) seem to indicate that there was an increase in α-helix and a decrease in random coil upon refolding with Cd(II). However, it is peculiar that a protein like Tat, which contains ten proline residues, has 0% turns. According to Fitzgerald et al. (18) regular CD computer programs cannot be used to predict a protein secondary structure having many multiple turns and coils. Further work is needed to study the secondary structure of the Tat protein.

2. Binding of Tar-RNA to Tat protein by fluorescence spectroscopy - The results are shown on Figure 7. There are one trptophan and two tyrosines residues in the Tat protein. The tryptophan fluorescence spectrum (λ_{excit}=295 nm) of Figure 7B was quenched by about 50% and the quenching is more extensive when excitated at 280 nm (Figure 7A). The results for the experiment of stoichemical binding of Tar-RNA to this fraction of Tat (most likely a monomer) is shown in Figure 7C. It began to reach saturation when 0.8 ml of Tar-RNA was added. The ratio of Tar-RNA to Tat is calculated to be 1.04, which is approximately a 1:1 ratio.

Discussion

Difficulties in preparing the Tat protein can be related to its origin as a fusion protein since Rop-Tat is an aggregate. Even after cyanogen bromide treatment in 6 M guanidine and 0.25 M HCl for 24 hours, the Tat protein under non-reducing conditions (without BME) stays mostly on top of the separating 15% gel (gel result not shown). After cyanogen bromide cleavage, the sample solution should contain a mixture of Rop and Tat proteins. However, the 0.25 M elute of the Q-Sepharose contains very little of the free Rop protein. After incubating the protein sample with 0.6% BME at 56°C for 3 hours, the 0.25 M elute of the Q-Sepharose increases greatly from a minor to a major peak (Figure 3A). This result indicates that the Rop is attached to Tat via disulfide bond(s). Attempts to reduce the incubation time to under 3 hours resulted in increased contamination of the protein band just atop of the Tat band and

slightly below the Rop-Tat fusion protein. This may be due to the fact that the 2 cysteine residues from the Rop participate in a disulfide linkage. Future work is necessary to replace the cysteine residue of the Rop with another amino acid in order to prevent the problem faced here. Even if not cross-linked to the Rop protein, Tat is capable of self-association with its own 7 cysteine residues, and it seems likely that incubation with BME would still be required to obtain free Tat. Karn *et al.* (19) use an incubation in 6 M guanidine hydrochloride buffer containing 0.1 M DTT at 70°C for 5 min. in preparing Tat protein. Raising the temperature from 56°C to 70°C can shorten the incubation time. It was found that incubating the sample with 0.6% BME at 70°C for 10 min. is equivalent to results from incubation at 56°C for 30 minutes.

After the BME incubation, the sample must immediately be applied onto the Q and S Sepharoses. If the incubation mixture is directly applied onto a Sephacryl S-200 size column, the majority of the protein appears to be Pk 1 (see Figure 4A), which is a large protein aggregate. This result seems to indicate that the ion-exchange matrix refolds the protein and prevents it from aggregation (20). The high tendency of Tat protein to multimerize may be important for Tat function in higher eukaryotes (21).

Researchers have debated whether the functionally active Tat protein is a monomer or a metal-linked dimer (6, 17). When Tat is over-expressed from *E. coli*, mammalian cells, or wheat germ, all are in a monomer form (7, 22, 23, 24). Tat behaved anomalously on the SDS-gel electrophoresis as if it had a molecular weight a few thousand daltons larger than its true molecular weight (7, 23,25). In our case, the Tat, from the 85 amino acid residues (9.6 kDa minus the first methionine residue) has a molecular weight half way between 21 and 14 kDa in a 15% SDS gel. The Pk 3 of Figure 4A, should presumably be a monomer, but on the Superose 12 size column it had a molecular weight larger than 2.5 kDa when compared with the protein standard running in a 0.5 M arginine pH 7.5 column buffer (Figure 4A). Using the mass spectrometer, this fraction had a molecular weight of 9,650 Da (mass spectrum not shown), indicating that Tat behaved as if it had a higher molecular weight than it was shown to have both by SDS gel electrophoresis and by size column. Churcher *et al.* (26) reported a Tat isolate which had a molecular weight of about 20 kDa, also from the Superose 12 size column. This larger-than-actual hydrodynamic size may be due to Tat's rodlike shape rather than a spherically shaped structure. Another possibility is that the Tat protein belongs to a class of all β-globular proteins (β-II)(27). This type of protein had a CD negative band around 200 nm and may have CD bands in the near UV region. Indeed, the Cd(II)-Tat's CD spectrum has both of these features (see Figure 9). A protein is classified as all-β protein β-II when it has predominantly β-sheets and little or no α-helix. The negative band around 200 nm resembles the spectrum of the random coil of an unfolded protein. In order to find out whether the protein is a β-II protein, X-ray diffraction data is needed.

The Pk 3 of the Sephacryl S-200 column can be dialyzed into a 0.05 M tris buffer (pH 8). However, it was noted that if one ran either this Pk 3 or the 0.5 M elute of the S-Sepharose fraction without urea on the size column, *i.e.*, using a 0.05 M tris pH 8 buffer, the protein would trail throughout the column. Similar protein behavior is observed while running the

protein sample without urea on the S-Sepharose column. Consequently, the Tat sample has to be eluted with 1 M NaCl instead of the 0.5 M NaCl when 6 M urea was present. This is possibly due to strong interaction between the protein and the column matrix and not due to a solubility problem because this Tat fraction can be dissolved into the 0.05 M sodium acetate pH 5 buffer up to 2 mg/ml.

The Cd-Tat of Figure 6 shows a larger molecular weight compared with the molecular weight before refolding with Cd(II). Further work is needed to see whether this is a metal-linked dimer. Another possibility is that this change is due to conformational changes of Tat upon intramolecular linkage with Cd(II), and thereby causing a positional change in the size column. In a mass spectroscopy experiment run on Cd-Tat, a 9,650 Da mass was observed. The calculated molecular weight of the Tat protein lacking the initial methionine residue and the Cd(II) metal is 9,653 Da. Although this result is consistent with the molecular weight of a Tat monomer, it is not known whether cadmium was released during the mass spectrum analysis, and therefore that an original dimer was destroyed in the process.

Binding of purified Tat protein to Tar-RNA as assayed by fluorescence was calculated to be a 1:1 ratio as reported by other investigators (28).

The abbreviations used are: HIV-1, human immunodeficiency virus 1; Tar, trans-acting-responsive RNA element; LTR, long terminal repeat; DTT, dithiothreitol; DTE, dithioerythritol; BME, β-mercaptoethanol; CD, circular dichroism; UV, ultraviolet; EDTA, ethylenediaminetetraacetic acid; PMSF, phenylmethanesulfonyl fluoride; FPLC, fast protein liquid chromatography; SDS, sodium dodecyl sulfate; Cd(II), cadmium chloride; Zn(II), Zinc chloride; PAGE, polyacrylamide gel electrophoresis.

REFERENCES

1. Cullen, B. R., (1992) Micrbiol. Rev. 56, 375-394
2. Dayton, A.L., Sodroski, J.G., Rosin, C.A., Goh, W.C., and Haseltine, W.A. (1986) Cell 44, 941-947
3. Fisher, A.G, Aldovini, A., Debouk, C., Gallo, R.C. and Wong-Staal, F. (1986) Nature (London) 320, 367-371
4. Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., Skinner, M.A., and Valerio, R. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 6925-6929
5. Ruben, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W. A., and Rosen, C. A. (1989) J. Virol. 63, 1-8
6. Frankel, A.D., Chen, L., Cotter, R.J., and Pabo, C.O. (1988) Proc. Natl. Acad. Sci. USA 85, 6297-6300
7. Rice, A.P., and Chan, F., (1991) Virology 185, 451-454
8. Sim, I.S., (1990) Annals Ny.Acad. Sci. 616, 64-71
9. Giza, P. E., and Huang, R.C.C. (1989) Gene 78, 73-84
10. Lee, T.C. and Christie, G.E. (1990) J. Biol. Chem. 265, 7472-7477
11. Goldenberg, D.P., (1989) in Prot. Stru.a practical approach (Creighton, T.E. ed) pp.225-250, IRL Press, Oxford, England
12. Hames, B.D. (1990) in Gel Electrophoresis of Proteins (Hames, B.D., and Rickwood, D. ed) pp. 34, IRL Press, Oxford, England
13. Yang, J.T., Wu, C.-S.C., and Martinez, H.M. (1986) in Methods in Enzymology (Hirs, C.H.W., and Timasheff, S.n., ed), 130, 208-269, Academic Press, San Diego
14. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L.S., Kuperman, R., Delling, U., and Sonenberg, N.(1991) J. Virol. 65, 5196-5202
15. Garcia,J.A., Harrich,D., Soultanakis, E., Wu,F., Zmitsuyasu, R., and Gaynor, R.B., (1989) EMBO J. 8, 765-778
16. McKenna, M., Anderson, D., Cascio, D., and Eisenberg, D., (1994) Acta Cryst. D50, 527-534
17. Frankel, A.D., Bredt, D.S., and Pabo, C.O. (1988) Science 240, 70-73
18. Fitzgerald, D.W., and Coleman, J.E. (1991) Biochemistry 30, 5195-5201
19. Karn, J., Churcher, M.J., Rittner, K., Kelley, A., Butler, P.J.G., Mann, D.A., and Gait, M.J. (1995) in HIV vol. 2 (Karn, J. ed)pp. 152, IRL Press, Oxford, England
20. Jaenicke, R. and Rudolph R. (1989) in Protein Structure a practical approach (Creighton, T.E. ed)pp. 200, IRL Press, Oxford, England
21. Bogerd, H.P., Fridell, R.A., Blair, W.S., and Cullen, B.R. (1993) J. Virol. 67, 5030-5034
22. Rice, A.P., and Carlotti, F. (1990) J. Virol. 64, 6018-6026
23. Waszak, G.A., Hasler, J.M., McQuade, T.J., Tarpley, W.G., and Deibel Jr,M.R. (1992) Prot. Expre. & Purifi. 3, 126-133
24. Rhim, H. Chan, F., Echetebu, C.O., and Rice, A.P. (1993) Prot. Expre. & Purifi. 4, 24-31
25. Müller, W.E.G., Okamoto, T., Reuter, P., Ugarkovic, D. (1990) J. Biol. Chem. 265, 3803-3808
26. Churcher, M.J., Lamont, C., Hamy, F., Dingwall, C., Green, S.M., Lowe, A.D., Butler, P.J.G., Gait, M.J., and Karn, J. (1993) J. Mol. Biol. 230, 90-110
27. Wu, J., Yang, J.T., and Wu C.C. (1992) Anal. Biochem. 200, 359-364
28. Dingwall,C., Ernberg, I., Gait, M. J., Green, S.M., Heaphy, S., Karn, J.,Lowe, A.D., Singh, M., and Skinner, M.A. (1990) EMBO J. 9, 4145-4153

```
                ↓                                10  ↓
met thr lys gln glu lys thr ala leu asn met ala arg phe ile arg ser gln thr leu

                                                 30
thr leu leu glu lys leu asn glu leu asp ala asp glu gln ala asp ile cys glu ser

                                                 50          53 ↓
leu his asp his ala asp glu leu tyr arg ser cys met glu pro val asp pro arg leu
                                                             1           5
                                                 70
glu pro trp lys his pro gly ser gln pro lys thr ala cys thr asn cys tyr cys lys
   10                15                20                25
                                                 90
lys cys cys phe his cys gln val cys phe ile thr lys ala leu gly ile ser tyr gly
   30                35                40                45
                                                 110
arg lys lys arg arg gln arg arg arg pro pro gln gly ser gln thr his gln val ser
   50                55                60                65
                                       130                  138
leu ser lys gln pro thr ser gln ser arg gly asp pro thr gly pro lys glu
   70                75                80                85  86
```

Figure 1. Fusion protein, Rop-Tat. Residues 1 to 52 define the Rop protein, and residues 53 to 138 define the Tat protein. The bottom numbers indicate where the Tat protein begins. Cyanogen bromide cleavage sites are indicated by arrows.

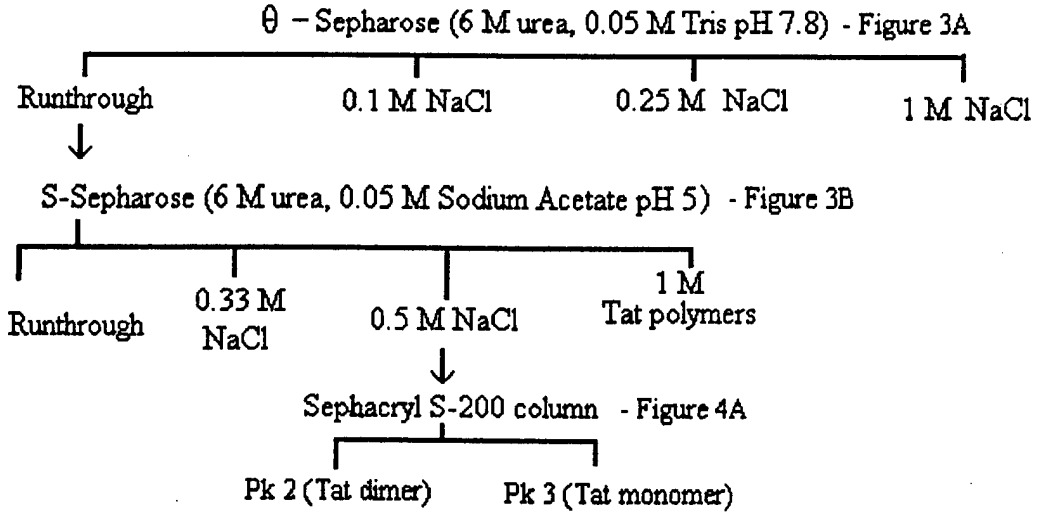


Figure 2 - Flow diagram of purification procedures

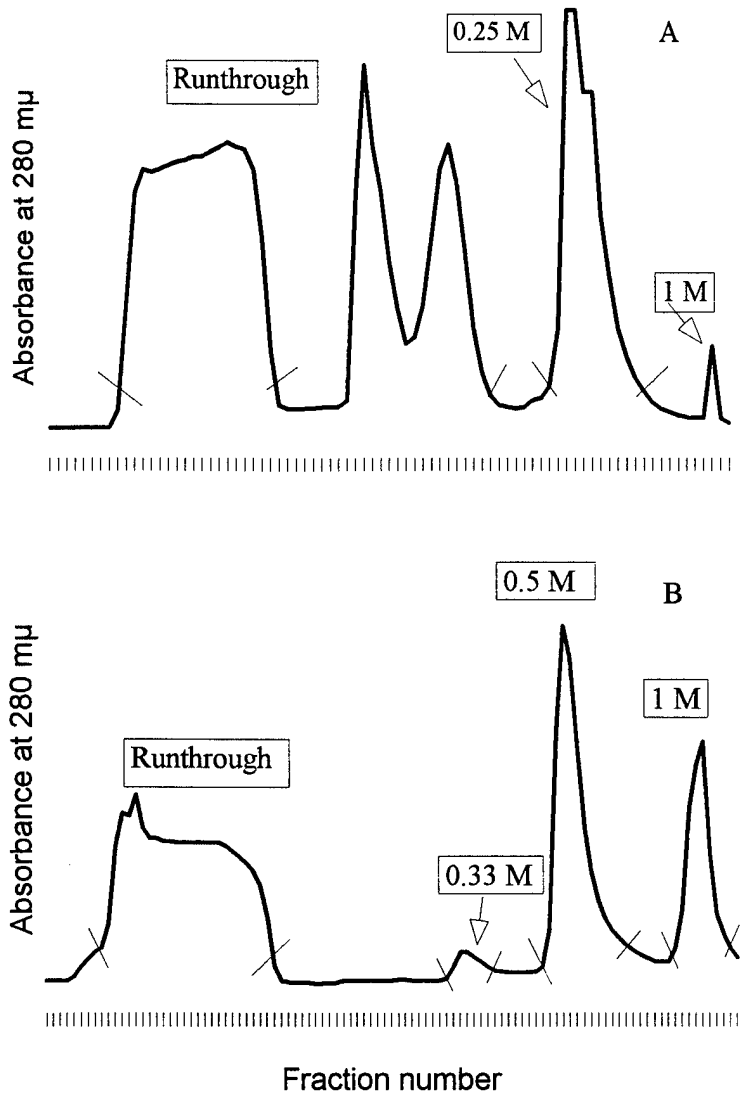


Figure 3.

Panel A. Q-Sepharose Column Chromatography (size 2.5 x 12 cm) Profile of Cyanogen Bromide Cleavage Products of Rop-Tat. The column was pre-equilibrated with 6 M urea, 0.05 tris at pH 7.8. The column was eluted stepwise with 0.1 M, 0.25 M, 1 M NaCl in the column buffer.

Panel B. S-Sepharose (2.5 x 12 cm) of Runthrough of Q-Sepharose of Figure 2. The column was pre-equilibrated with 6 M urea, 0.05 Sodium Acetate at pH 5. The column was eluted with 0.1 M, 0.33 M, 0.5 M and 1 M NaCl in the column buffer.

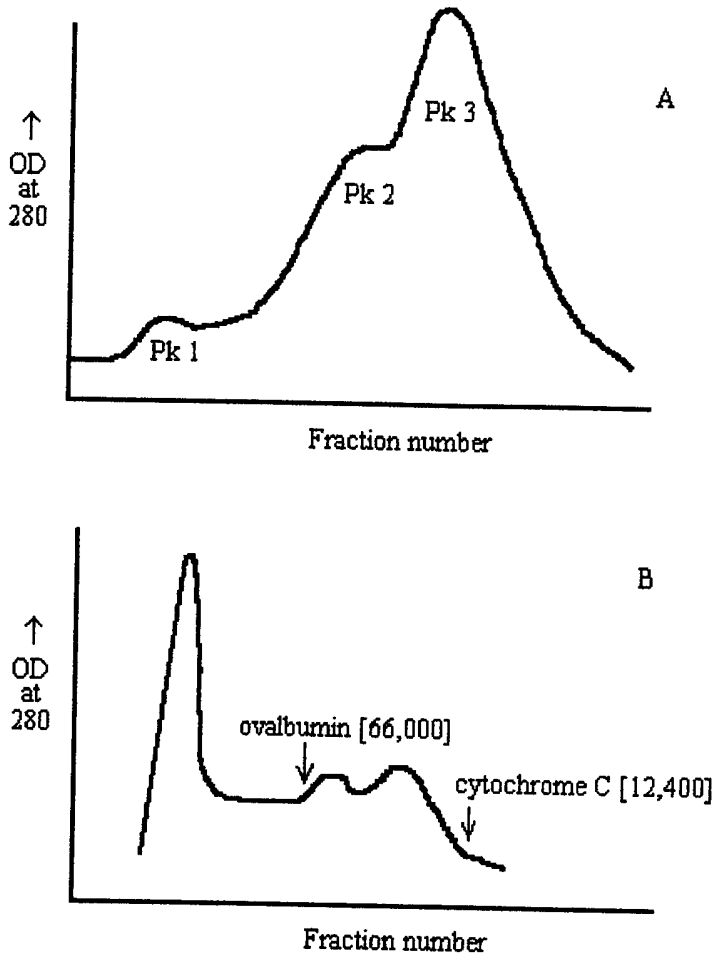


Figure 4.

Panel A - Sephacryl S-200 column (2 x 42 cm)
column buffer either 2 M urea, 0.05 M sodium acetate, 1 M NaCl at pH 5.0, or in 0.5 M arginine at pH 7.5

Panel B - Sephacryl S-200 HR (1.5 x 45 cm)
column buffer 2 M urea, 0.05 M sodium acetate, 1 M NaCl at pH 5.0
arrows indicate molecular standards; flow rate for both columns = 0.5 ml/min

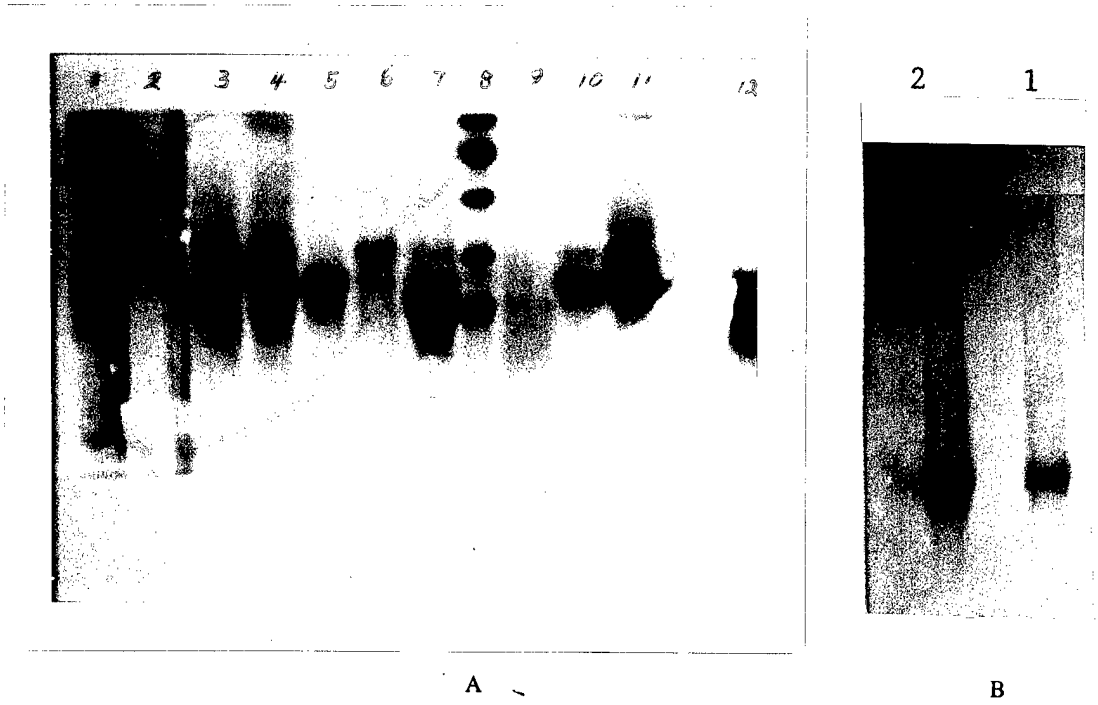
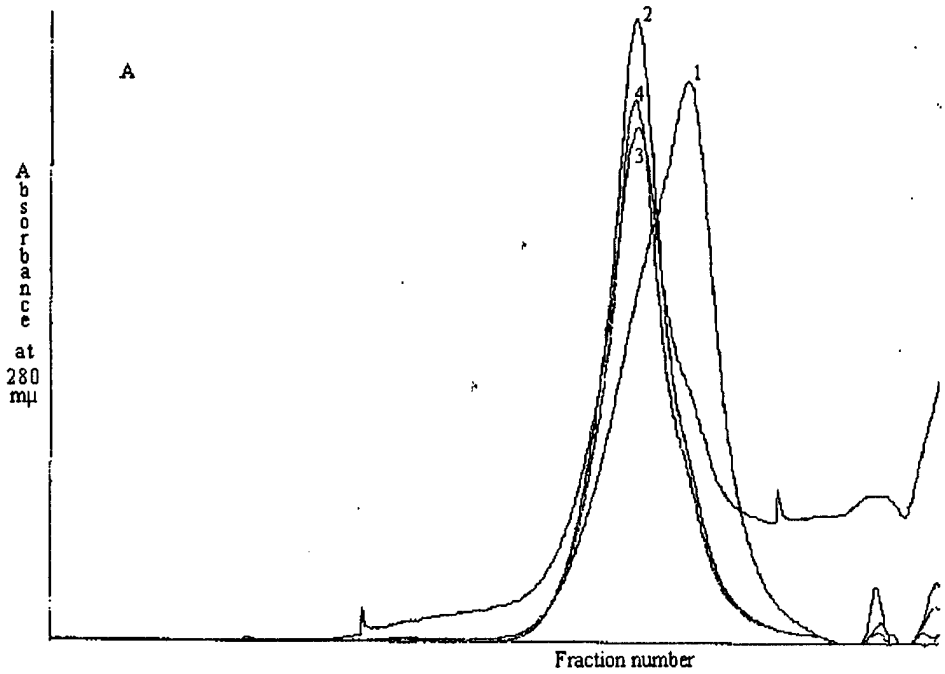


Figure 5

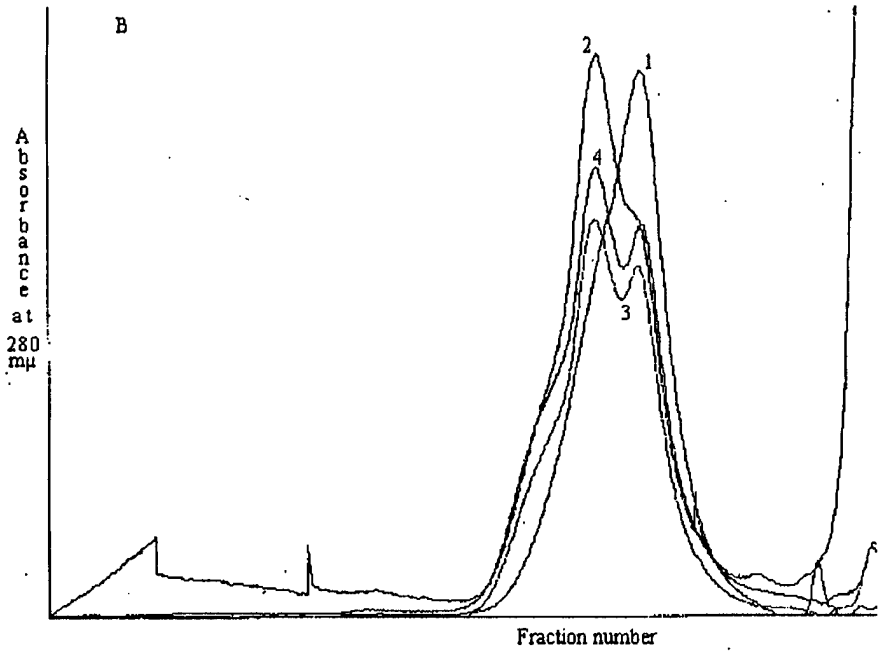
Panel A. SDS gel electrophoresis in 15% gel (reducing conditions)
Lane 1. Total cell lysate before sonication and washing.
Lane 2. Rop-tat after sonication and washing.
Lane 3. After cyanogen bromide cleavage, then incubated with 0.6% BME for 3 hours and before θ-Sepharose
Lane 4. [same as Lane 3]
Lane 5. 0.5 M elute of S-Sepharose
Lane 6. 0.1 M elute of θ-Sepharose
Lane 7. 0.25 M elute of θ-Sepharose
Lane 8. Size markers (see text)
4th marker = trypsin inhibitor 20,100 Da
Bottom marker = ∝-lactalbumin 14,400 Da
Lane 10. Runthrough of S-Sepharose
Lane 11. 1M elute of 0.5 M elute of S-Sepharose
Lane 12. Cd-Tat of 0.5 M elute of S-Sepharose

Panel B. Native gel electrophoresis
Lane 1. Tat from refolding of Method 1
Lane 2. Same sample as Lane 1, incubated in 0.6% BME at 56°C for 1 hr before running the gel.



A. Superose 12 HR 10/30 FPLC filtration in 0.5 M arginine pH 7.5
Flow rate = 0.5 ml/min

Curve 1. Tat protein from 0.5 M elute of S-Sepharose before refolding with cadmium (see Procedure II-3)
2. 24 hr dialysis in arginine, $CdCl_2$, cystamine, cysteamine buffer (see Text)
3. 48 hr dialysis with above buffer
4 72 hr dialysis with above buffer



B. Same as in A except refolded in Zinc Chloride

Figure 6.

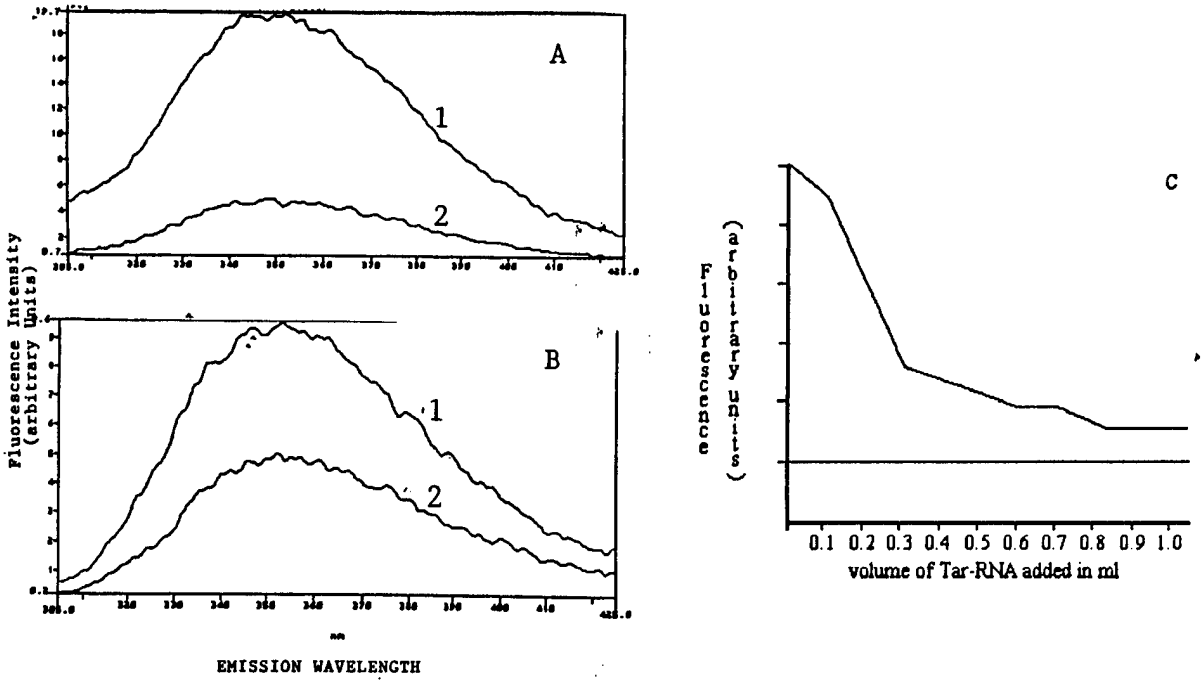


Figure 7. Fluorescence spectra

A - Emission spectra of Tar-RNA and Tat protein at wavelength λ_{excit} = 280 nm
Curve 1 - Tat protein
Curve 2 - Tat + Tar-RNA (1:1 ratio)

B - Same as in A except λ_{excit} = 295 nm

C - Titration curve of Tar-RNA to Tat. The curvette contained 0.8 ml of the Tat solution and aliquots of 0.1 ml of Tar-RNA were added to a total of 1 ml (see Method).

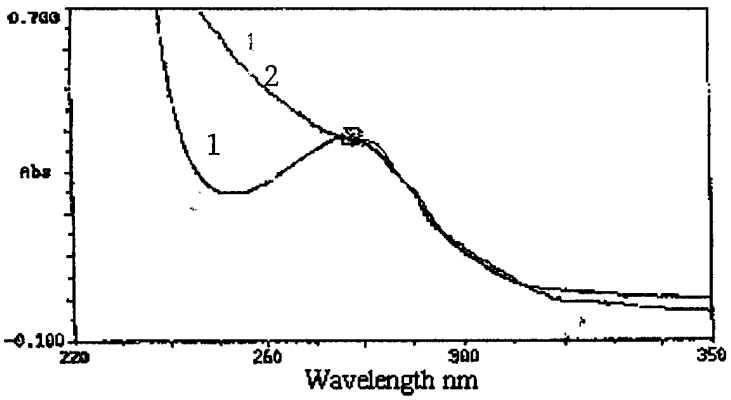


Figure 8. UV absorption spectra

Curve 1: Tat protein before refolding
Curve 2: Tat after refolding in presence of Cadmium chloride [Cd(II)-Tat]

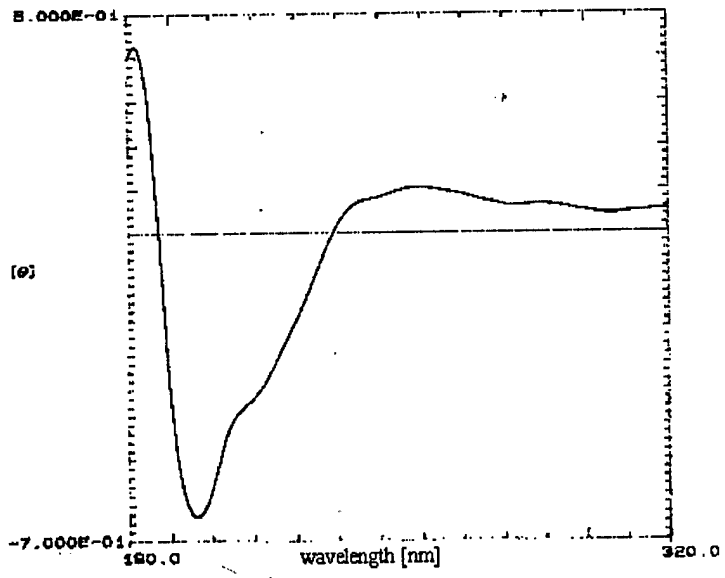


Figure 9. Circular dichroism of Cd(II)-Tat

The protein sample is in 10 mM sodium phosphate, 40 μM $CdCl_2$, pH 7.0

This final report is going to be submitted for publication

Meeting Abstracts

Protein Folding and Design, April 23-26, 1996 at the National Institute of Health, Bethesda, Md.

and to Molecular Biology: The Next Decade, May 20-22, 1996 at the Chinese Academy of Science.

Abstract:

Purification and refolding of recombinant HIV-Tat from bacterial proteins in inclusion bodies of E. coli. Walls, L.C., Nie, W., Giza, P.E., and Huang, R.C., Chemistry Dept., Morgan State University, Baltimore, MD 21239, and Dept. of Biology, Johns Hopkins University, Baltimore, MD 21205, USA

We have earlier reported that high yield (0.5 gm/1) of HIV-Tat protein can be achieved as part of membrane bound inclusion bodies (I.B.) in E. coli by a self-inducing run-away-replication plasmid expression system utilizing the Rop protein (Gene 78:3-84, 1989). Methods and procedures have been developed to dissociate recombinant Rop-Tat protein from membrane bound E. coli proteins and to obtain soluble HIV-Tat protein from I.B. for physical studies. Several cycles of sonication were used to wash off most E. coli contaminant proteins. After ammonium sulfate precipitation, the Rop-Tat was subjected to CNBr cleavage. Isolation of the Tat protein were achieved by using Q and SP-Sepharoses in urea buffer. The eluted fraction from the SP column was further resolved into two different molecular weight fractions by Sephacryl S-200 size column. These two fractions were used for refolding studies. Two different methods of refolding were used: 1) by incubating the Tat in the presence of 8 M urea, o.1 M DTT and then gradually refolding into native buffer; 2) by using an oxidative regeneration system in the presence of oxidized and reduced glutathione. The results of these refolding methods will be discussed.

Personnel:

Research Assistants:
1. Dr. Wenxian Nie, August 1994-August 1995.
2. Mr. Chris Kelley , July 1993-August 4.

Students:
1. Jurga Grigoraviciute
2. Carolyn Myles
3. Taufiki Lee
4. Elham Mohammadi