Rayner, D. M., & Szabo, A. G. (1978) *Can. J. Chem. 56*, 743–745.
Ricci, R. W. (1974) *Photochem. Photobiol. 12*, 67–75.
Ricci, R. W. & Nesta, J. M. (1976) *J. Phys. Chem. 80*, 974–980.
Robbins, R. J., Fleming, G. R., Beddard, G. S., Robinson, G. W., Thistlewaite, P. J., & Woolfe, G. J. (1980) *J. Am. Chem. Soc. 102*, 6271–6279.
Ross, J. B. A., Rousslang, K. W., & Brand, L. (1981) *Biochemistry 20*, 4361–4369.
Ross, J. B. A., Laws, W. R., Buku, A., Sutherland, J. C., & Wyssbrod, H. R. (1986) *Biochemistry 25*, 607–612.
Saito, I., Sugiyama, H., Yamamoto, A., Muramatsu, S., & Matsuura, T. (1984) *J. Am. Chem. Soc. 106*, 4286–4287.
Saito, I., Muramatsu, S., Sugiyama, H., Yamamoto, A., & Matsuura, T. (1985) *Tetrahedron Lett. 26*, 5891–5894.
Santus, R., Bazin, M., & Aubailly, H. (1980) *Rev. Chem. Intermed. 3*, 231–283.
Shizuka, H., Serizawa, M., Shimo, T., Saito, I., & Matsuura, T. (1988) *J. Am. Chem. Soc. 110*, 1930–1934.
Song, P. S., & Kurtin, W. E. (1969) *Photochem. Photobiol. 9*, 175–178.
Strickland, E. H., & Billups, C. (1973) *Biopolymers 12*, 1989–1995.
Szabo, A. G. (1989) in *The Enzyme Catalysis Process: Energetics, Mechanism, and Dynamics* (Cooper, A., Houben, J. L., & Chien, L. C., Eds.) pp 123–139, Plenum Press, New York.
Szabo, A. G., & Rayner, D. M. (1980) *J. Am. Chem. Soc. 102*, 554–563.
Valeur, B., & Weber, G. (1977) *Photochem. Photobiol. 25*, 441–444.
Walker, M. S., Bednar, T. W., Lumry, R., & Humphries, F. (1971) *Photochem. Photobiol. 14*, 147–161.
Ware, W. R., & Nemzek, T. L. (1975) *J. Chem. Phys. 62*, 477–489.
Werner, T. C., & Forster, L. S. (1979) *Photochem. Photobiol. 29*, 905–914.
Wiget, P., & Luisi, L. (1978) *Biopolymers 17*, 167–180.

Physicochemical Properties of Cloned Nucleocapsid Protein from HIV. Interactions with Metal Ions†

Daniel W. Fitzgerald‡ and Joseph E. Coleman*
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510
Received December 18, 1990; Revised Manuscript Received February 19, 1991

ABSTRACT: The nucleocapsid (NC) protein (p15) of the human immunodeficiency virus (HIV) has been cloned and overproduced (under the control of a phage T7 promoter) in soluble form in an *Escherichia coli* host. The soluble NC protein is a fusion protein containing 15 amino acids from the T7 gene 10 and 7 amino acids from the HIV p24 protein at the N-terminus to make a protein of 171 amino acids. The plasmid containing the fusion gene is designated p15DF. A homogeneous product has been isolated from the induced cells and, when isolated under aerobic conditions, contains 0.3–0.5 mol of Zn/mol of protein and has only 2 titratable SH groups. Reduction and refolding in the presence of Zn(II) yields a protein containing 2.0 mol of Zn/mol of protein and 6 titratable SH groups. On the other hand, if the cells are sonicated in 2 mM $CdCl_2$ and purified at pH 5.0, an unoxidized protein containing 2 mol of Cd/mol of protein is obtained. The Cd(II) ions can be exchanged with Zn(II), Co(II), or ^{113}Cd(II). The $Co(II)_2$ NC protein shows d–d electronic transitions at 695 nm [ϵ = 675 M^{-1} cm^{-1} per Co(II)] and 640 nm [ϵ = 825 M^{-1} cm^{-1} per Co(II)] compatible with regular tetrahedral geometry around both Co(II) ions. The $Co(II)_2$ and $Cd(II)_2$ NC proteins show intense charge-transfer bands in the near-UV, at 355 nm (ϵ = ~4000 M^{-1} cm^{-1}) and 310 nm (ϵ = ~8000 M^{-1} cm^{-1}) for the Co(II) protein and 255 nm (ϵ = ~10^4 M^{-1} cm^{-1}) for the $Cd(II)_2$ NC protein, compatible with $-S^-$ coordination. ^{113}Cd NMR of the $^{113}Cd(II)_2$ NC protein shows two ^{113}Cd NMR signals at 659 and 640 ppm, respectively, each integrating to ~1 Cd(II) ion. The downfield chemical shifts suggest coordination of each ^{113}Cd(II) ion to 3 sulfur donor atoms. The spectroscopic data fully support the prediction that the NC protein binds metal ions to each of the tandem repeats of the -Cys-X_2-Cys-X_4-His-X_4-Cys- sequence contained in the N-terminal half of the molecule. ^{113}Cd NMR shows, however, that the sites are not identical. Isolation of the NC protein under standard aerobic conditions results in oxidation of the sulfhydryl groups and loss of the coordinated Zn(II) ions, while preparation of the NC protein as the Cd(II) derivative at low pH protects the sulfhydryl groups from oxidation.

The nucleocapsid protein (NC) of the human immunodeficiency virus (HIV), also known as p15, is a protein of 149 amino acids and is coded for by the 3′-end of the GAG gene of the AIDS virus, HIV (Ratner et al., 1985; Guyader et al., 1987). The GAG gene product, a polyprotein, is transported to the cell membrane and incorporated into the budding viral particle where it is subsequently cleaved by the viral-encoded protease into three major polypeptides, the NC protein being released from the C-terminal portion of the polyprotein. There appears to be further cleavage of a portion of the NC protein into two smaller products, p7 and p9 (Veronese et al., 1987; Mervis et al., 1988; Henderson et al., 1988), but the functional

† This work was supported by NIH Grants GM21919 and DK09070. The 500-MHz NMR was supported by NIH Grant RR03475, NSF Grant DMB-8610557, and ACS Grant RD259.
* Corresponding author.
‡ Supported in part by NIH Cancer Education Grant R25-CA47883.

0006-2960/91/0430-5195$02.50/0

significance of this additional cleavage is not clear. The nucleocapsid protein is a single-stranded nucleic acid binding protein and is incorporated into the virus particle in an estimated 2000–3000 copies (Meric et al., 1984).

There are several functions reported for the retroviral NC proteins. The NC nonspecifically coats the single-stranded RNA genome (Meric et al., 1984; Fleissner & Tress, 1973). Mutations in the NC protein have shown that the protein is essential for the specific packaging of viral RNA, since mutant NC proteins can lead to viral particles which either do not contain the viral RNA or contain cellular RNA (Meric & Goff, 1989; Gorelick et al., 1988; Aldovini & Young, 1990; Gorelick et al., 1990). Many investigators have postulated that the NC proteins bind specifically to a region near the 5′-end of the retroviral genome designated the ψ sequence. This sequence located between a 5′ splice donor site and the ATG translational start signal for the GAG gene is necessary for the correct packaging of viral RNA (Mann et al., 1983; Lever et al., 1989). Formation of a complex between the RNA genome and the NC protein at this site has been postulated to have a role in holding the two copies of the viral genome together (Meric & Spahr, 1986; Bieth et al., 1990). Synthesis of the double-stranded DNA copy of the viral RNA, following endocytosis of the viral particle, is catalyzed by the reverse transcriptase (RT) and appears to involve transcription from a genome still contained in a macromolecular aggregate consisting of the viral RNA and various GAG gene products. It has therefore been postulated that the NC protein also has a role in assisting the replication process carried out by the reverse transcriptase (Sykora & Moelling, 1981; Prats et al., 1988).

The amino acid sequence of the NC protein shows the N-terminal half of the protein from HIV as well as NC proteins from a number of other retroviruses to contain tandem repeats of the sequence -C-X_2-C-X_4-H-X_4-C- (Ratner et al., 1985; Guyader et al., 1987). Point mutations of the CCHC amino acid residues of these sequences result in defective packaging of viral RNA (Gorelick et al., 1988; Meric & Goff, 1989; Aldovini & Young, 1990; Gorelick et al., 1990), suggesting that the CCHC residues are essential to the proper functioning of the NC proteins. This sequence is similar to the sequence of a Zn(II)-binding domain first found in the single-stranded DNA binding protein, gene 32 from phage T4,where the metal binding sequence is -C-X_3-H-X_5-C-X_2-C- (Giedroc et al., 1986). It has been proposed by a number of investigators that the corresponding sequences in the retroviral NC proteins also bind Zn(II) ions (Giedroc et al., 1986; Berg, 1986; Schiff et al., 1988). A corollary of this observation is the hypothesis that the metal binding is essential to function and it is the metal binding that is interrupted by point mutations in the CCHC residues. This proposal remains controversial because the NC proteins isolated by the standard techniques do not seem to contain stoichiometric amounts of Zn(II) (Jentoft et al., 1988), although studies in vitro show that it is possible under controlled conditions to bind metal ions to these proteins or to peptides containing the proposed metal binding sequence from the NC proteins (Roberts et al., 1989; South et al., 1989, 1990). The present paper reports the cloning and overproduction in an *Escherichia coli* host of the NC protein from HIV and the metal binding properties of this protein.

Materials and Methods

Cloning and Overexpression of the NC Protein from HIV. A 554 base pair fragment encoding the nucleocapsid protein was isolated from λpBH10 as follows. The plasmid known as λpBH10 (Ratner et al., 1985), a gift of Dr. Robert Gallo, was digested with *Eco*RI (all restriction enzymes were from New England Biolabs). A 4-kb *Eco*RI fragment was isolated on an 8% TBE acrylamide gel and electroeluted from the gel. The isolated fragment was inserted into the *Eco*RI site of pUC8 with T4 DNA ligase. This new plasmid, pEco A, was digested with *Hpa*II and *Dra*I. A 722 base pair fragment encoding the C-terminus of the GAG gene, corresponding to base pairs 1178–1900 of λpBH10, was isolated from the digest. The 5′-overhanging end of this fragment was filled in by Klenow DNA polymerase. A *Bam*HI linker was ligated to both blunt ends of this fragment. The resultant 746 bp fragment was digested with *Bam*HI and inserted into the *Bam*HI site of the pIBI24 plasmid (International Biotechnologies, Inc.). The resultant plasmid, pDG5907, was digested with *Bst*XI which cleaves 80 base pairs downstream of the GAG gene translational stop codon (position 1732 of λpBH10). A *Sal*I linker was introduced at this site. A *Bam*HI–*Sal*I fragment (base pairs 1178–1900 of λBH10) was then isolated and inserted between the *Bam*HI and *Sal*I sites of pIBI24. This plasmid with a 544 bp insert encoding the HIV NC protein is designated pDG71412.

The NC protein gene was inserted downstream of the T7 promoter in the overexpression plasmid pAR3039 (Studier & Moffatt, 1986), using a *Bam*HI–*Hin*dIII fragment from pDG71412 (containing the entire structural gene for the NC protein) and inserting it between the *Bam*HI and *Hin*dIII sites of pAR3039. This places the coding sequence for the NC protein 42 base pairs downstream of an ATG start codon provided by pAR3039. A fusion protein is expressed which contains 15 amino acids from the T7 gene 10, encoded by pAR3039, plus seven amino acids from HIV p24 fused to the N-terminus of the NC protein. The entire product is 171 amino acids in length, and the plasmid containing this fusion gene is designated p15DF and was used to overexpress the NC fusion protein employed in the present investigation.

The plasmid p15DF was transformed into *E. coli* strain BL21(DE3) which has a T7 RNA polymerase gene downstream of a lac promoter on the bacterial chromosome (Studier & Moffatt, 1986). Cells were grown in M9 medium, including 2% glucose and 1% bactotryptone, to an OD_{600} of 0.8 and induced with IPTG (Pharmacia), 0.3 mg/mL. After 4 h, the final OD_{600} was 1.5–2.0 and the cells were harvested by centrifugation. A 10-L growth yields ~50 g of cells.

Purification of the NC Protein from HIV. Cells were sonicated in 50 mM Tris-HCl, pH 8.0, 100 mM KCl, 5 mM BME, 100 μM $ZnCl_2$, and 0.03 mg/mL PMSF for 5 min (1 g of cells/2 mL of buffer). The cell extract was spun for 10 min at 10K rpm in a Sorvall GSA rotor. Poly(ethylenimine) (Aldrich) was added to the supernatant to 1% (v/v) and the DNA allowed to precipitate for 10 min on ice. The supernatant was then separated at 10K rpm in a Sorvall GSA rotor and loaded onto an SP-Trisacryl M column (IBF Biotechnics, Columbia, MD) (20 mm diameter column, 1 cm length/g of cells). The HIV NC protein binds tightly to this column and was eluted with a salt gradient of 100–500 mM NaCl in 50 mM Tris-HCl, 5% glycerol, 2 mM BME, and 100 μM $ZnCl_2$, pH 8.0. Fractions containing the NC protein (detected by gel electrophoresis) were pooled, concentrated by an Amicon ultrafiltrator (PM10 membrane), and loaded onto a Sephadex G-75 size exclusion column. The protein was eluted with 50 mM Tris-HCl, pH 8.0, 100 mM KCl, 5 mM BME, and 100 μM $ZnCl_2$. The protein was greater then 95% pure after this step as judged by SDS–PAGE, was concentrated to 1 mg/mL, and was stored under nitrogen. All further manipulations were

carried out under a nitrogen atomosphere.

Isolation of a Cd(II) NC Protein. A modification of the above protocol made it possible to isolate an unoxidized Cd(II) derivative of the NC protein. The cells were sonicated for 3 min in 50 mM potassium acetate, 100 mM KCl, 5 mM BME, 2 mM $CdCl_2$, pH 5.0, and 0.03 mg/mL PMSF. Two milliliters of buffer was used per gram of cells. Poly(ethylenimine) was added directly to the sonicated cells, and a single centrifugation at 15K rpm was used to remove cell debris and precipitated material. The salt gradient used for the SP-Trisacryl column was modified to a 200–800 mM KCl gradient in 10 mM potassium acetate, 2 mM BME, and 100 μM $CdCl_2$, pH 5.0 (buffer A). Fractions containing NC protein were pooled, concentrated, and loaded onto a Sephadex G-75 size exclusion column. The protein was washed from the column with buffer A plus 100 mM KCl and was greater than 95% pure as judged by SDS–PAGE. This protein contained ~2 mol of Cd(II) ions/mol of protein as assayed by atomic absorption (see below).

Preparation of the Zn(II), Cd(II), and Co(II) Derivatives of the NC Protein of HIV. Apo NC protein was prepared by dialyzing the concentrated Cd(II) protein, purified at pH 5.0, against 10 mM potassium acetate, 100 mM KCl, 2 mM BME, and 10 mM EDTA, pH 5.0, for 24 h. Dialysis buffer was changed twice at 8-h intervals. Apoprotein was then dialyzed extensively against 10 mM potassium acetate, 100 mM KCl, and 2 mM BME, pH 5.0, for 12 h to remove the EDTA. All buffers were made metal free by passing them over a Chelex 100 (Bio-Rad) column and degassed thoroughly before use.

To make the Co(II) protein, the apoprotein was dialyzed against metal-free 10 mM Tris-HCl, 100 mM NaCl, and 5% glycerol, pH 5.0 (TNG buffer), to remove BME. Cobalt sulfate, 0.5 mM, was added to 0.1 mM NC protein, and the pH was raised to 7.5 by the addition of 30 mM Tris-HCl, pH 7.5. Samples used for the determination of the absorption spectra of the Co(II) protein were ~10^{-4} M in the presence of equimolar free Co(II) to ensure maximum metal binding (~2 mol/mol) and thus maximal absorption (see Results). Ultrafiltration and dilution, while freeing the protein from excess Co(II), always results in some loss of Co(II) from one of the sites, reflecting the fact that binding affinities of the metal ions are in the order Cd(II) > Zn(II) > Co(II) (see Table I below). Visible and UV spectra were taken on a Cary 118 spectrophotometer.

The Zn(II) protein was made by adding a 5-fold molar excess $ZnSO_4$ to the apoprotein at pH 5.0. The protein was then dialyzed against 10 mM Tris-HCl, 100 mM NaCl, 5% glycerol, 2 mM BME, and 50 μM $ZnCl_2$, pH 7.5, for 8 h to raise the pH. Excess Zn(II) was removed by dialyzing against TNG buffer and 2 mM BME for 12 or 24 h.

The Cd(II) protein was usually obtained from the purification procedure carried out in the presence of excess Cd(II) as described above. The Cd(II) derivative can be regenerated by the addition of a 5-fold excess of $CdSO_4$ to the apoprotein at pH 5.0. The protein was then dialyzed against 20 mM Tris-HCl and 50 μM $CdSO_4$, pH 7.5, for 8 h. Excess Cd(II) was removed by dialyzing against 20 mM Tris-HCl, pH 7.5, for 24 h. Formation of the Cd(II) protein can be confirmed by monitoring the development of S → Cd(II) charge-transfer bands at 255 nm.

Protein concentrations were determined by amino acid analysis carried out by the Yale Protein Chemistry Facility. From the optical density of the solutions at 280 nm an extinction coefficient, ϵ_{280}, of 1.78×10^4 M^{-1} cm^{-1} was calculated for the cloned p15.

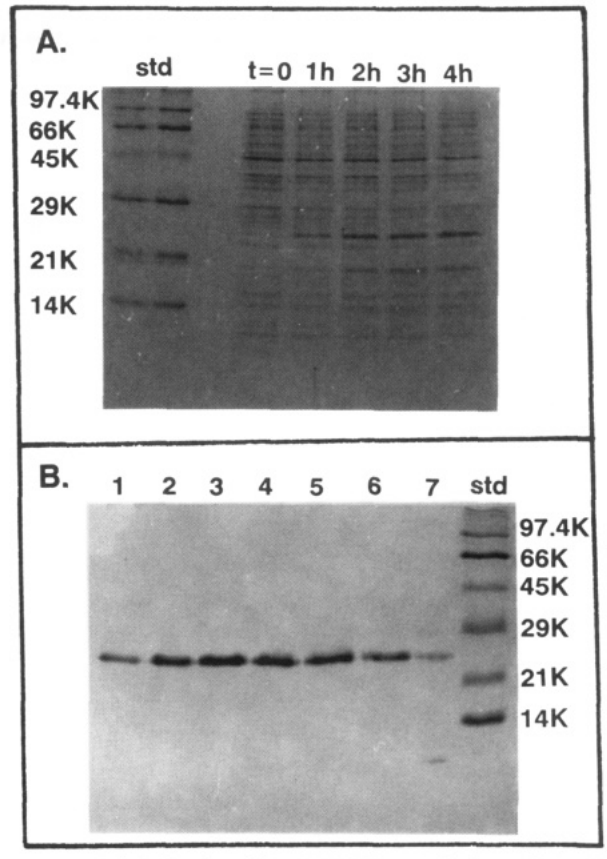

FIGURE 1: (A) Induction of the HIV NC protein from the p15DF plasmid as followed by SDS–PAGE. Gels of the total cell extract are shown at 0, 1, 2, 3, and 5 h after inducing with IPTG. The NC protein migrates at 23 kDa. (B) SDS–PAGE of the fractions (1–7) containing the NC protein as eluted from a Sephadex G-75 column as described under Materials and Methods.

Metal ion analyses for Zn, Cd, and Co were by flame atomic absorption on an Instrumentation Laboratories Model 157 atomic absorption spectrometer.

^{113}Cd NMR was performed on a Bruker AM-500 spectrometer with a 10-mm broad-band probe. The spectra were acquired at 25 °C with a 45° pulse angle and a delay time of 1 s. The spectral width was 15.2 kHz. The chemical shift is plotted relative to that of 0.1 M $^{113}Cd(ClO_4)_2$, $\delta = 0$ ppm.

Results

Cloning and Purification of the NC Protein from HIV. A number of plasmids were constructed in the attempt to overexpress the NC protein from HIV; however, only the p15DF construct yielded soluble protein sufficient for biochemical and spectroscopic studies. Other constructs produced protein that was either insoluble or completely proteolyzed intracellularly. When *E. coli*, transformed with p15DF, are induced with IPTG, the NC protein becomes one of the major cellular proteins (Figure 1A). The protein migrates at a position corresponding to 23 kDa on SDS–PAGE under our conditions. Greater than 90% of the induced protein is found in the supernatant after sonication of the whole cells. The cloned NC protein is subject to proteolysis by *E. coli* proteases, and purification must be carried out rapidly to avoid extensive degradation. The major early proteolytic fragments can be separated from the uncleaved protein by a Sephadex G-75 column (Figure 1B). Purification at pH 8.0 yielded protein that was greater than 95% pure. However, this protein when isolated without the addition of metal ions to the buffers contained only 0.3–0.4 mol of Zn(II)/mol of protein (Table I). The protein isolated in this manner was analyzed for the presence of free

Table I: Metal Content of the Nucleocapsid Protein from HIV

method of preparation	mol/mol Zn	Co	Cd	free S^- titrated
standard pH 8.0 purification as described under Materials and Methods	0.3			2
protein boiled in DTT, refolded in the presence of Zn(II)	1.7			6
pH 5.0/excess Cd/G-75 gel exclusion[a]			2.0	
Co protein (after ultrafiltration)[b]		1.75		
Zn protein				
12-h dialysis[c]	2.0			
second 12-h sample	1.85			
Zn protein				
24-h dialysis[c]	1.65			
second 24-h sample	1.60			
Cd^{113} protein, 24-h dialysis[d]			2.2	

[a] Buffer was 10 mM potassium acetate and 2 mM β-mercaptoethanol, pH 5.0. [b] Metal-free buffer for ultrafiltration consisted of 30 mM Tris-HCl, 100 mM NaCl, and 5% glycerol, pH 7.5. Both the Co(II) and Zn(II) proteins were made from the Cd(II) protein originally isolated at pH 5.0 to avoid oxidation of the sulfhydryls as described under Materials and Methods. [c] Dialysis was against metal-free buffer consisting of 10 mM Tris-HCl, 100 mM NaCl, 5% glycerol, and 2 mM β-mercaptoethanol, pH 7.5. [d] Dialysis was against metal-free buffer consisting of 20 mM Tris-HCl, pH 7.5.

sulfhydryls to determine if oxidation of the -SH groups during purification could account for the loss of Zn(II) binding. Reaction of the protein with DTNB showed that there were only 2 free sulfhydryls. The protein contains 6 Cys residues, all contained in the CCHC putative ligand sequences. If the protein was boiled in the presence of 10 mM DTT, dialyzed extensively under nitrogen to remove DTT, 6 free sulfhydryls were titrated with DTNB, suggesting that, without precautions, the NC protein is subject to oxidation of its metal binding sulfhydryl groups. In fact, this regenerated protein can bind stoichiometric amounts of Zn(II) (Table I).

Isolation of a Cd(II) NC Protein To Prevent the Oxidation of Sulfhydryl Groups during Purification. A purification scheme was devised to protect the cysteine sulfhydryl groups from oxidation based upon the purification of metallothionein in which Cd(II) is added to the extracted protein at low pH at a very early stage in the isolation. The NC protein was purified in the presence of excess Cd(II) at pH 5.0 beginning with the addition of 2 mM Cd(II) to the sonication buffer. The binding of Cd(II) to the NC protein can be followed by measuring the S → Cd(II) charge-transfer band at 255 nm (Figure 2). Purified in this way, the Cd(II) protein is more than 95% pure and contains 2 mol of Cd/mol of protein (Table I). The Cd(II) can be removed by dialyzing against 10 mM EDTA to produce an apoprotein that can be used to make other metallo derivatives. Zn(II) or a spectroscopically active divalent cation like Co(II) can then be added to the apoprotein to produce the desired metalloprotein. Table I shows the metal stoichiometry of metalloproteins reconstituted in this way.

Absorption Spectroscopy of the Co(II) NC Protein. The visible optical absorption spectra of cobalt-substituted metalloproteins, due to the d–d electronic transitions of the metal ion, can be used to determine in a qualitative way the geometry at the metal binding sites, while the UV absorption spectra can detect the presence of sulfur atoms as donors via the typical S → Co(II) charge-transfer bands. Addition of Co(II) in 20 mM Tris-HCl at pH 7.5 to the apo NC protein produces a blue-green protein. Figure 3 shows the visible and near-UV spectra of this cobalt-substituted NC protein. This derivative has a d–d absorption envelope in the visible range with maxima at 695 nm (ϵ = 675 M^{-1} cm^{-1}) and 640 nm (ϵ = 825 M^{-1} cm^{-1}), both expressed per single Co(II) ion. These are among the highest extinction coefficients observed for Co(II)-substituted Zn(II) proteins and are similar to those observed for regular tetrahedral Co(II) in model complexes rather than distorted tetrahedral or 5-coordinate complexes of Co(II). The energies of these absorption bands are similar to those observed previously for the Co(II) derivatives of the T4 gene 32 protein (Giedroc et al., 1986) and the Co(II) derivative of the NC protein from MuLV known as p10 (Roberts et al., 1989). There are two overlapping S → Co(II) charge-transfer bands at 355 and 310 nm (ϵ = ~4000 and ~8000 M^{-1} cm^{-1}, respectively), indicating that sulfurs are involved in Co(II) ligation. Metal analysis shows that NC protein from HIV binds 1.75 mol of Co(II)/mol of protein when the free Co(II) is

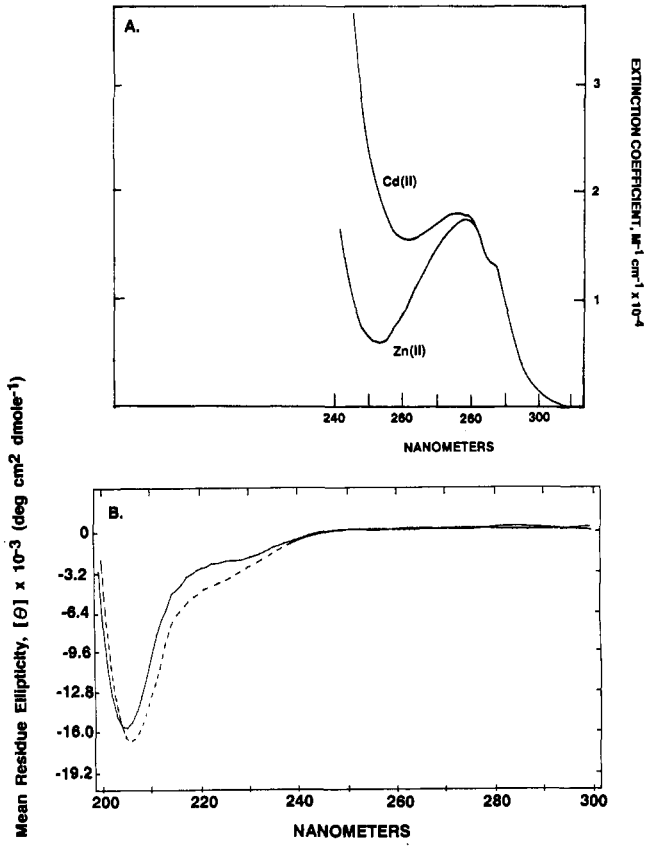

FIGURE 2: (A) Near-UV absorption spectrum of $Zn(II)_2$ and $Cd(II)_2$ NC protein. A difference spectrum shows the S → Cd(II) charge-transfer band to have a maximum near 255 nm. (B) Circular dichroism of the $Zn(II)_2$ NC protein (—) and the apo (reduced) NC protein (---). Conditions: 10 mM Tris-HCl, 100 mM NaCl, pH 7.5, and 10 μM protein.

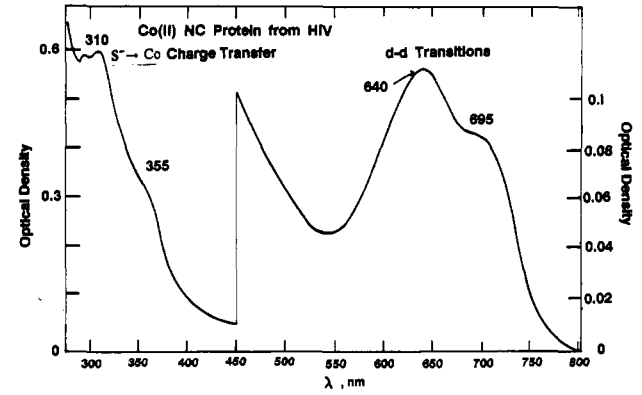

FIGURE 3: Visible and near-UV absorption spectra of $Co(II)_2$ NC protein. Conditions: 10 mM Tris-HCl, 100 mM NaCl, 5% glycerol, pH 7.5, and 7.2 × 10^{-5} M protein.

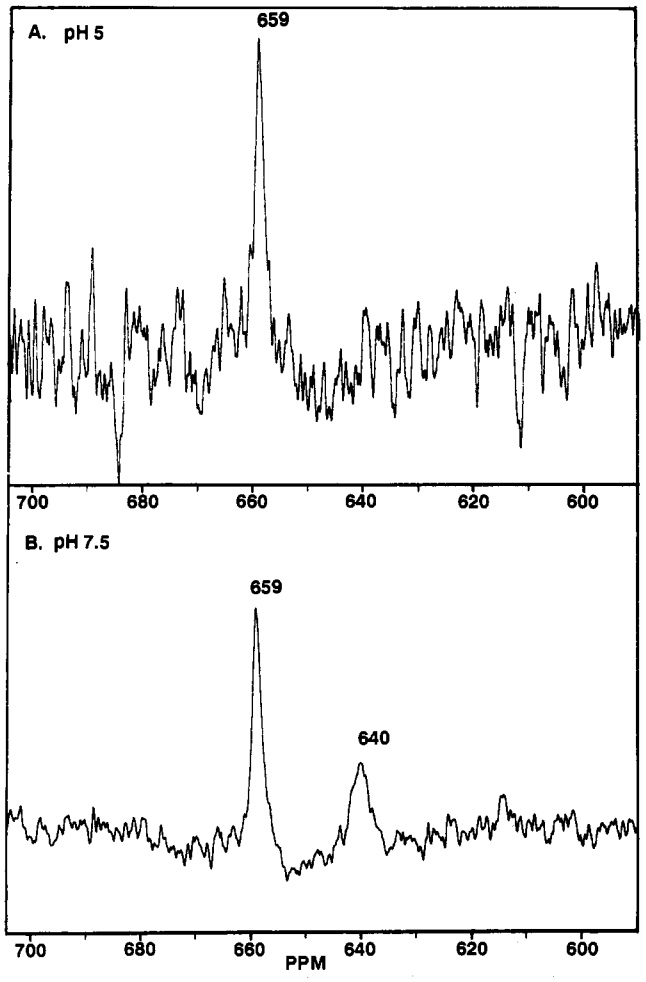

FIGURE 4: ^{113}Cd NMR spectra of ^{113}Cd derivatives of the HIV NC protein. (A) Apo NC protein (0.5 mM) in the presence of 5 equiv of $^{113}Cd(II)$, pH 5.0, 10 mM potassium acetate, 100 mM KCl, and 2 mM β-mercaptoethanol (17 000 scans). (B) $^{113}Cd_2$ NC protein, pH 7.5, 20 mM Tris, and 0.5 mM protein (78 500 scans).

removed by ultrafiltration and the sample diluted for atomic absorption (Table I). Dialysis against $\sim 10^{-5}$ M free cobalt and determination of protein-bound Co(II) by difference between the inside vs the outside of the dialysis bag yield 2.0 ± 0.1 mol/mol of protein.

^{113}Cd NMR of the Cd(II) NC Protein from HIV. ^{113}Cd NMR can be used to determine the number of sulfur donors at the metal binding sites of metalloproteins, since sulfur ligation induces large downfield shifts in ^{113}Cd resonances. These downfield shifts are roughly proportional to the number of sulfur ligands present. When 5 equiv of $^{113}Cd(II)$ is added to the apoprotein at pH 5.0, a single ^{113}Cd resonance is observed at 659 ppm (Figure 4A). When the pH is raised to 7.5 and the excess Cd(II) is removed, two ^{113}Cd signals are observed, one at 659 ppm and a second at 640 ppm (Figure 5B). Each of these resonances is within the chemical shift range expected for ligation of the $^{113}Cd(II)$ ion to 3 $-S^-$ ligands. Thus all 6 sulfhydryl groups of the NC protein must be involved in $^{113}Cd(II)$ coordination. Similar chemical shifts are seen for T4 gene 32 (637 ppm) and for the $^{113}Cd(II)$ derivative of the NC protein form MuLV (648 ppm) (Giedroc et al., 1989; Roberts et al., 1989). These ^{113}Cd chemical shifts are also similar to that observed for the $^{113}Cd(II)$ complex of an 18-residue synthetic peptide containing the sequence of one of the HIV CCHC regions (653 ppm) (South et al., 1989). While the amplitude of the peak at 659 ppm is ~ 3 times that of the peak at 640 ppm, the line width of the peak at 640 ppm is at least twice that of the peak at 659 ppm. Collecting the ^{113}Cd NMR spectra at longer delay times established that the spectra in Figure 4B are fully relaxed, and integration of the signals at 659 and 640 ppm shows them to correspond to 1.0 and 0.80 mol of Cd nuclei/mol of protein.

Circular Dichroism of the NC Protein from HIV. Circular dichroism was used to determine if metal binding induces a change in the structure of the peptide backbone. The CD spectra of the apoprotein and the Zn(II) protein are generally similar in shape, with a shoulder between 220 and 240 nm and a maximum negative ellipticity at 206 nm (Figure 2B). These spectra cannot be easily fit by a combination of the CD spectra observed for the known secondary structures of poly(L-lysine), α-helix, β-sheet, or "random coil" (Greenfield & Fasman, 1969). Nor is the CD spectrum of the NC protein identical with that for the "random" form of poly(L-lysine). When Zn(II) is added to the apoprotein, there is an increase in ellipticity between 203 and 240 nm. The difference in ellipticity is maximum near 220 nm where the positive difference in mean residue ellipticity is $+1.5 \times 10^3$ deg cm^2 dmol^{-1}.

Discussion

A number of overexpression systems were designed to overproduce the nucleocapsid protein from HIV. By varying the N- and C-termini of the gene product, we were able to maximize overproduction, maintain solubility, and limit proteolysis. The fusion protein with 22 extra amino acids on the N-terminus yielded the best results (Figure 1A). Millimolar concentrations of the NC protein can be produced from p15DF suitable for NMR and other spectroscopic experiments. This greatly facilitates the study of the HIV nucleocapsid protein which has been only available to date from viral-infected cultures (Henderson et al., 1988; Barat et al., 1989; South et al., 1990).

Predictions based upon the primary sequences (Giedroc et al., 1986; Berg, 1986), metal binding to short peptides containing the putative metal binding amino acid sequence (South et al., 1989; Green & Berg, 1989), and metal binding to the complete synthetic NC protein from MuLV (Roberts et al., 1989) all suggested that retroviral nucleocapsid proteins would bind stoichiometric amounts of zinc under reducing conditions. The isolation of the expected metallo forms of the native NC proteins, however, has remained elusive. A number of groups including our own either found no zinc or were unable to show satisfactory metal ion stoichiometry (Jentoft et al., 1988). This may have been due to the oxidation of the cysteine sulfhydryls during purification at pH 8.0, a phenomenon observed in our laboratory during early attempts to carry out proton NMR studies on p10 isolated from MuLV. Oxidation of the sulfhydryls clearly occurs in the cloned NC protein from HIV (Table I) and can be avoided by purifying the Cd(II)-substituted protein at pH 5.0. If the pH is kept well below the pK_a of the cysteine sulfhydryls, but high enough for Cd(II) to occupy the two metal sites, the Cd(II) can then easily be removed by EDTA and Zn(II) restored to the metal sites at a higher pH under an inert atmosphere.

The stoichiometry of metal binding to the purified metalloprotein (Table I), as well as the two ^{113}Cd resonances observed for the $^{113}Cd(II)$-substituted protein (Figure 4), proves that the HIV nucleocapsid protein binds two metal ions. The d–d absorption bands in the spectrum of the Co(II)-substituted protein show that these metal ions are coordinated in an approximately tetrahedral symmetry. The S → Cd(II) and S → Co(II) charge-transfer bands observed in the respective derivatives suggest one or more sulfur ligands. The chemical shifts of the ^{113}Cd signals for both sites are consistent with

coordination of each Cd(II) ion to three $-S^-$ donors. These spectroscopic observations strongly suggest that the CCHC ligand assignment for both metal binding sites in the NC protein is correct.

While the primary sequences of the two metal binding domains are similar, the ^{113}Cd NMR indicates not only that there are small conformational differences between the two sites but also that the affinities of the two sites for Cd(II) are probably not identical. In the ^{113}Cd NMR spectrum at pH 5.0 in the presence of excess Cd(II), the absence of the signal at 640 ppm is most likely due to a chemical exchange broadening judging from previous experience with the ^{113}Cd NMR of proteins (Coleman et al., 1979). The chemical exchange process could represent either an exchange of the central metal ion at a rate corresponding to intermediate chemical exchange or a conformational modulation of the protein. Since the line width of the signal at 659 ppm does not seem to be affected by this exchange, central metal ion exchange may be the more likely process. The appearance of a well-defined signal from the second site at 640 ppm at pH 7.5 in the absence of excess Cd(II) fits well with central metal ion exchange as the modulating process. Even at pH 7.5 the modulation still significantly broadens the signal (Figure 4).

A conformational and exchange difference between the two sites is particularly interesting in light of recent evidence that the two metal-binding domains in the NC protein from RSV, p12, may play different functional roles in the life cycle of the virus (Meric & Goff, 1989). While one might expect the separated CCHC ligand arrangements to have identical stabilities and ^{113}Cd NMR chemical shifts, it is possible that the folding induced by the two-metal form of the protein makes the sites inequivalent. Some sort of interaction between the two closely spaced metal ion domains when both are occupied could give rise to a negative cooperativity of metal ion binding. The slow loss of one of the Zn(II) ions on dialysis (Table I) also suggests that one of the CCHC sites binds Zn(II) less tightly. Except for a conserved 4-residue spacer between the second C ligand and the H, the amino acid residues forming the spacers between the ligand residues are all different in the two tandem sequences.

The CD spectrum of the apoprotein and the Zn(II) protein indicates that metal binding probably does not cause large changes in the folding of the peptide backbone. A number of studies have shown that metal ions can induce folding of small peptides of less than 30 residues containing the ligand arrangement found in the NC protein, peptides that would not be expected to fold in the absence of the metal ion because of their short length (South et al., 1989). On the other hand, a protein of 171 residues like the NC protein construct used in the present work would be expected to have a folded structure. The CD spectrum of the HIV nucleocapsid protein in either its apo or metallo form shows that neither protein contains any large amount of α-helix or β-sheet structure (Figure 2B). If the predominant fixed structure is composed of turns and coils of some kind, then the CD will not be particularly useful in defining this structure, and the magnitude of additional folding induced by the metal ions is probably not easy to judge on the basis of the CD changes. The Zn(II) derivative of p15 has an additional positive contribution to ellipticity centered in the region of 220 nm (Figure 4). While this could reflect a change in the conformation of the polypeptide backbone upon Zn(II) binding, it could also reflect a contribution from the $-S^- \rightarrow$ Zn(II) charge-transfer bands, expected to be located in this region of the spectrum. A similar ellipticity band has been observed in the Zn(II) derivative of p10 from MuLV (Roberts et al., 1989).

The present data create a standard for the physicochemical properties of the cloned NC protein from HIV and show that in the reduced form this protein is capable of forming tetrahedral Zn(II) and Cd(II) complexes with its two CCHC ligand donor sets. These data do not in themselves provide information on the ability of either the metallo form or the nonoxidized metal-free form of the NC protein to function in any of the roles postulated for the NC protein and reviewed in the introduction. To do so requires the design of assay systems that will mimic as far as possible both the enzyme reactions and the multipolypeptide interactions that are probably present at the various stages of the viral life cycle.

Relatively early studies of the reverse transcriptase activity of avian myoblastosis virus (AMV) as assayed with disrupted viral particles detected an enhancement effect of the AMV NC protein on the homologous reverse transcriptase activity (Sykora & Moelling, 1981). More recent studies suggest that the NC protein is required for annealing of the primer, $tRNA^{Lys}$, to the viral RNA template before reverse transcription can begin (Barat et al., 1989). The HIV NC protein also induces the formation of HIV RNA dimers and even higher oligomers as detected on RNA gels (Barat et al., 1989). This might be related to the postulated RNA packaging function. It is the latter function, failing when the CCHC ligands are altered by deletion of substitution mutations, that has led to the strongest suggestions that the metal binding must be essential for function (Meric & Spahr, 1986; Karpel et al., 1987; Gorelick et al., 1988; Jentoft et al., 1988). If the metal binding imparts a specific function to the NC protein, it may not relate to the ability of the NC proteins to bind nonspecifically to single-stranded RNA. Binding of p10 from MuLV to poly(ethenoadenylic acid) is not metal ion dependent, and the -SH groups can be blocked or oxidized without altering its affinity for single-stranded RNA (Karpel et al., 1987; Roberts et al., 1989). Preliminary fluorescence quenching titations of poly(ethenoadenylic acid) with the cloned HIV NC protein indicate that the metal ion stoichiometry appears to have little effect on the tight binding of the NC protein to this homopolymer, $K_d < 10^{-9}$ M. The CD spectra of p15 from HIV suggest that metal ion binding does not cause major changes in protein structure and thus the general fold of the apoprotein may be rather similar to that of the metalloprotein. A metal ion requirement could, however, be essential for one of the postulated specific functions of the NC protein as discussed above. The cloning and isolation of a homogeneous HIV NC protein in which the oxidation of the sulfhydryls and the metal ion stoichiometry can be manipulated will allow testing of the metal ion function in some of the assays for the more specific functions of the HIV NC protein.

Another aspect of the NC protein biochemistry that could relate to the NC protein function is related to the translational frame shift that occurs when the complete GAG-POL mRNA is translated in the HIV life cycle. The majority of the GAG polyprotein is synthesized by a translational process taking place in the GAG gene reading frame, a translation that terminates just downstream from the end of the NC protein coding region. The C-terminal end of the resultant GAG polyprotein has the amino acid sequence of our cloned p15. During those translational events that result in synthesis of a GAG-POL fusion protein, occurring in about 1 in 20 translational events, the required translational frame shift occurs 7 triplets down from that coding for the last C of the second CCHC sequence of p15. Therefore the C-terminal half of the p15 synthesized by this mechanism has a markedly

different amino acid sequence than the p15 currently under study. Whether the alternate p15 has important separate functions or behaves differently in its metallo form remains an open question. A simple alteration in the sequence of the cloned gene should make isolation of this alternate form of the NC protein possible.

ACKNOWLEDGMENTS

We thank Dr. David Giedroc for much help in the initial cloning of the p15 gene and Dr. Tao Pan for many helpful suggestions.

Registry No. Zn, 7440-66-6; Cd, 7440-43-9; Co, 7440-48-4.

REFERENCES

Aldovini, A., & Young, R. A. (1990) *J. Virol. 64*, 1920–1926.

Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeyre, M. T., Gruninger-Leitch, F., Barre-Sinoussi, F., LeGrice, S. F. J. & Darlix, J. L. (1989) *EMBO J. 8*, 3279–3285.

Berg, J. M. (1986) *Science 232*, 485.

Bieth, E., Gabus, C., & Darlix, J. L. (1990) *Nucleic Acids Res. 18*, 119–127.

Coleman, J. E., Armitage, I. M., Chlebowski, J. R., Otvos, J. D., & Schoot Uiterkamp, A. J. M. (1979) in *Biological Applications of Magnetic Resonance* (Shulman, R. G., Ed) pp 345–395, Academic Press, New York.

Fleissner, E., & Tress, E. (1973) *J. Virol. 12*, 1612–1615.

Giedroc, D. P., Keating, K. M., Williams, K. R., Konigsberg, W. H., & Coleman, J. E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8452–8456.

Giedroc, D. P., Johnson, B. A., Armitage, I. M., & Coleman, J. E. (1989) *Biochemistry 28*, 2410–2418.

Gorelick, R. J., Henderson, L. E., Hanser, J. P., & Rein, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8420–8424.

Gorelick, R. J., Nigida, S. M., Jr., Bess, J. W., Jr., Arthur, L. O., Henderson, L. E., & Rein, A. (1990) *J. Virol. 64*, 3207.

Green, L. M., & Berg, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4047–4051.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4116.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., & Alizon, M. (1987) *Nature (London) 326*, 662–669.

Henderson, L. E., Copeland, T. D., Sowder, R. C., Schultz, A. M., & Oroszlan, S. (1988) *Hum. Retroviruses, Cancer, Aids: Proc. Abbott–UCLA Symp., 1987*, 135–147.

Jentoft, J. E., Smith, L. M., Fu, X., Johnson, M., & Leis, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7094–7098.

Karpel, R. L., Henderson, L. E., & Oroszlan, S. (1987) *J. Biol. Chem. 262*, 4961.

Lever, A., Gottlinger, H., Haseltine, W., & Sodroski, J. (1989) *J. Virol. 63*, 4085–4087.

Mann, R., Mulligan, R. C., & Baltimore, D. (1983) *Cell 33*, 401–407.

Meric, C., & Spahr, P. F. (1986) *J. Virol. 60*, 450–459.

Meric, C., & Goff, S. (1989) *J. Virol. 63*, 1558–1568.

Meric, C., Darlix, J. L., & Spahr, P. F. (1984) *J. Mol. Biol 174*, 531–538.

Mervis, R. J., Ahmad, N., Lillehoj, E. P., Raum, M. G., Salazar, F. H. R., Chan, H. W., & Venkatesan, S. (1988) *J. Virol. 62*, 3993–4002.

Prats, A. C., Sarih, L., Gabus, C., Litvak, S., Keith, G., & Darlix, J. L. (1988) *EMBO J. 7*, 1777–1783.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Staricich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., & Wong-Staal, F. (1985) *Nature (London) 313*, 277–283.

Roberts, W. J., Pan, T., Elliot, J. I., Coleman, J. E., & Williams, K. R. (1989) *Biochemistry 28*, 10043–10047.

Schiff, L. A., Nibert, M. L., & Fields, B. N. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4195–4199.

South, T. L., Kim, B., & Summers, M. F. (1989) *J. Am. Chem. Soc. 111*, 395–396.

South, T. L., Blake, P. R., Sowder, R. C., Arthur, L. E., Henderon, L. E., & Summers, M. F. (1990) *Biochemistry 29*, 7786–7789.

Studier, F. W., & Moffatt, N. (1986) *J. Mol. Biol. 189*, 113–130.

Sykora, K. W., & Moelling, K. (1981) *J. Gen. Virol. 55*, 379–391.

Veronese, F., Rahman, R., Copeland, T. D., Oroszlan, S., Gallo, R. C., & Sarngadharan, M. G. (1987) *AIDS Res. Hum. Retroviruses 3*, 253–263.