Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4047–4051, June 1989
Biochemistry

A retroviral Cys-Xaa$_2$-Cys-Xaa$_4$-His-Xaa$_4$-Cys peptide binds metal ions: Spectroscopic studies and a proposed three-dimensional structure

LORA M. GREEN AND JEREMY M. BERG

Department of Chemistry, The Johns Hopkins University, 34th and Charles Streets, Baltimore, MD 21218

Communicated by Richard H. Holm, March 9, 1989 (received for review December 20, 1988)

ABSTRACT **Retroviral *gag* gene-encoded core nucleic acid binding proteins contain either one or two sequences of the form Cys-Xaa$_2$-Cys-Xaa$_4$-His-Xaa$_4$-Cys. Previously, one of us has proposed that these sequences form metal-binding domains in analogy with the "zinc finger" domains first observed in transcription factor IIIA. We report that an 18-amino acid peptide derived from the core nucleic acid binding protein from Rauscher murine leukemia virus binds metal ions such as Co^{2+} and Zn^{2+}. The absorption spectrum of the peptide–Co^{2+} complex is highly suggestive of tetrahedral coordination involving three cysteinates and one histidine. Titration experiments indicate that the dissociation constant for the peptide–Co^{2+} complex is 1.0 μM and that Zn^{2+} binds more tightly than Co^{2+}. A detailed three-dimensional structure for this domain based on conserved substructures in other crystallographically characterized metalloproteins and on a detailed analysis of the Cys-Xaa$_2$-Cys-Xaa$_4$-His-Xaa$_4$-Cys sequences from retroviruses and other related sources is proposed.**

In 1985, two groups observed the occurrence of nine tandem sequences of the form Cys-Xaa$_4$-Cys-Xaa$_{12}$-His-Xaa$_3$-His in the deduced amino acid sequence of *Xenopus* transcription factor IIIA (TFIIIA) (1, 2). Based on the presence of zinc in a purified TFIIIA–5S RNA complex (1, 3), it was proposed that each of these sequences forms a metal-binding domain—that is, a relatively discrete structural unit stabilized by the tetrahedral coordination of a zinc ion to the invariant cysteine and histidine residues. These domains were termed "zinc fingers" (1). Subsequently, numerous other deduced protein sequences have been found that contain quite similar sequences that match the template described above (4–6). Where it is known, the function of these proteins is to act as specific nucleic acid binding proteins. Studies with several proteins have revealed that zinc is required for this activity (3, 7–9). The hypothesis that these sequences do indeed form metal-binding domains has been amply supported by a wide variety of methods including limited proteolysis studies of the TFIIIA–5S RNA complex (1), extended x-ray absorption fine structure spectroscopic studies of the zinc sites in the TFIIIA–5S RNA complex (10), studies of the structure of the TFIIIA gene (11), hydroxyl radical footprinting studies of a series of shortened versions of TFIIIA on a 5S RNA gene (12), and studies of single domain peptides (13, 14).

Shortly after the discovery of the zinc finger motif, one of us developed a systematic search procedure for identifying potential metal binding domains in protein sequences (15). Several classes of proteins that had been implicated in nucleic acid binding or gene regulatory processes were identified. These include the bacteriophage gene 32 protein and the adenovirus E1A large protein. Each of these proteins has subsequently been shown to contain a stoichiometric amount of zinc that appears to be bound via the proposed sequence (16, 17).

One of the most striking sequence motifs identified by the search method has the form Cys-Xaa$_2$-Cys-Xaa$_4$-His-Xaa$_4$-Cys. Hereafter, this motif is referred to as the CCHC box. One or two such sequences occur in the *gag*-encoded small nucleic acid binding proteins of retroviruses. Indeed, the presence of this conserved motif had been previously noted (18), although its potential to form a metal ion-based domain had not been discussed. Furthermore, sequences of this form have also been discovered in systems other than retroviruses such as the *Drosophila* transposable element *copia* (19) and cauliflower mosiac virus (20) that appear to share the property that they undergo a reverse transcription step at some point in their life cycles (21). The importance of the conserved cysteine and histidine residues for viral replication has been directly demonstrated by site-directed mutagenesis in two systems (22, 23). Results obtained by using a radioactive zinc blotting technique indicated that these proteins have an affinity for zinc under certain conditions (24). We report herein that an 18-amino acid sequence Asp-Gln-Cys-Ala-Tyr-Cys-Lys-Glu-Lys-Gly-His-Trp-Ala-Lys-Asp-Cys-Pro-Lys derived from the sequence of the nucleic acid binding protein from Rauscher murine leukemia virus (18) binds Co^{2+} to produce a complex that has an absorption spectrum highly suggestive of tetrahedral S_3N coordination. Titration experiments reveal that the dissociation constant for this complex is 1.0 μM at pH 7.0 and that Zn^{2+} readily displaces Co^{2+} from the peptide. This result provides strong evidence that the sequences in the proteins do indeed form metal-binding domains. In addition, we propose a detailed three-dimensional structure of these domains that is based on conserved substructures from crystallographically characterized metalloproteins and is consistent with an analysis of the properties of the CCHC box sequences.

MATERIALS AND METHODS

The peptide was synthesized on a Milligen model 9050 Pepsynthesizer using *N*-fluorenylmethoxycarbonyl amino acid pentafluorophenyl esters (from Milligen). Once the peptide synthesis was complete, the resin was washed several times with dichloromethane and dried. Cleavage of the peptide from the resin and removal of side-chain protecting groups was effected by treatment with trifluoroacetic acid with 2% phenol and 2% ethanedithiol as scavengers. The peptide was purified by reverse-phase high performance liquid chromatography on a Vydac C4 column using a gradient of acetonitrile/0.1% trifluoroacetic acid in 0.1% trifluoroacetic acid/water (0–22%). The largest peak was collected and the solvent was removed with a Savant Speed Vac concentrator. The peptide was reduced by treatment with 0.33 M dithiothreitol for 2 hr at 45°C. The reduced peptide was purified as described above. All manipulations of the reduced peptide were per-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: TFIIIA, transcription factor IIIA.

formed under an atmosphere of purified dinitrogen. Amino acid analysis was performed on a Waters Picotag amino acid analysis system. Free thiol concentrations were determined using 5,5'-dithiobis(2-nitrobenzoic acid) (25). Peptide concentrations were determined using an extinction coefficient of 7000 $M^{-1}\cdot cm^{-1}$ at 280 nm.

Metal-binding studies were performed in 20 mM Hepes/50 mM NaCl, pH 7.00, buffer. Absorption spectra were recorded on a Hewlett–Packard 8451A diode array spectrophotometer. Cobalt dichloride hydrate (99.999%) and zinc dichloride (99.999%) were purchased from Aldrich. Metal ions were added in the same buffer as described above. Spectral subtractions and determinations of the relative concentration of peptide–Co^{2+} complex were performed with software supplied with the spectrophotometer. The dissociation constant for the peptide–Co^{2+} complex was determined from spectrophotometric data using locally written software. All computer model building and graphics were done on an Apple Macintosh II computer using Chem3D (Cambridge Scientific Computing, Cambridge, MA).

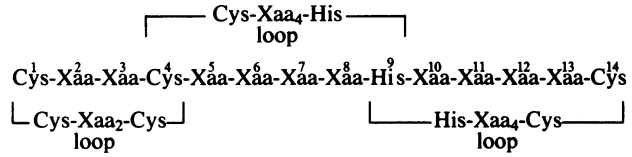
RESULTS

The peptide Asp-Gln-Cys-Ala-Tyr-Cys-Lys-Glu-Lys-Gly-His-Trp-Ala-Lys-Asp-Cys-Pro-Lys has been synthesized by solid-phase methods. Amino acid analysis of the purified peptide was consistent with the expected composition. The peptide in its reduced form is quite sensitive to air oxidation as judged by thiol content determination. However, the peptide can be prepared in a reduced form with 3.0 free thiols per peptide by treatment with dithiothreitol and purification under conditions that minimize exposure to air.

The reduced peptide binds metal ions. Treatment with solutions of Co^{2+} at pH 7.00 produced a chromophore with absorbances at 314 (ε = 3950 $M^{-1}\cdot cm^{-1}$), 350 (shoulder), 650 (ε = 520 $M^{-1}\cdot cm^{-1}$), and 696 (ε = 470 $M^{-1}\cdot cm^{-1}$) nm. The dissociation constant for Co^{2+} binding was determined by titrating a solution of the peptide with Co^{2+} and monitoring the absorption spectrum as shown in Fig. 1. The titration data were fit using a least-squares procedure to yield a dissociation constant of K_d^{Co} = 1.0 μM at pH 7.0. The spectrum of the Co^{2+} complex is largely bleached by the addition of one equivalent of Zn^{2+} per peptide, indicating that Zn^{2+} effectively competes with Co^{2+} for the peptide metal-binding site. Experiments with oxidized preparations of the peptide revealed no detectable interactions with Co^{2+}.

The CCHC box sequences from retroviral and other sources are shown in Table 1. Inspection of these sequences reveals that, in addition to the invariant cysteine and histidine residues, several other features are commonly observed that presumably have structural implications. These provide additional constraints for use in model building.

For the purpose of model development, the CCHC sequence will be divided into overlapping fragments as shown below:



The first region involves six amino acids and includes the first two invariant cysteine residues. As noted previously (52), two classes of crystallographically characterized proteins have metal-chelating Cys-Xaa_2-Cys regions that are not part of larger arrays of closely spaced cysteine residues. These are *Escherichia coli* aspartate transcarbamoylase (53) and the rubredoxins (54–56). The structures of the three unique Cys-Xaa-Xaa-Cys-Xaa-Xaa loops are shown in Fig. 2. In-

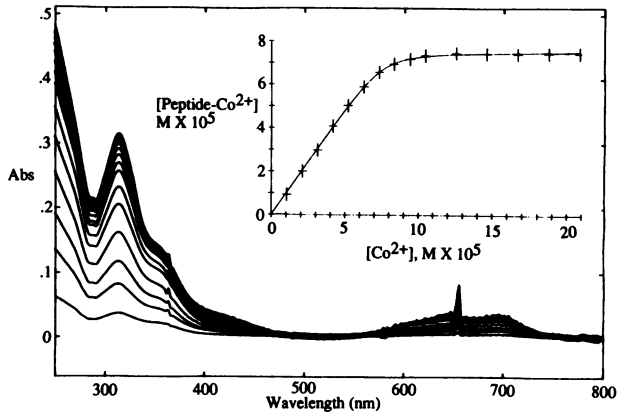


FIG. 1. Titration of the CCHC box peptide with Co^{2+}. A solution of the reduced peptide in 20 mM Hepes/50 mM NaCl, pH 7.00, buffer was treated with aliquots of $CoCl_2$ in the same buffer and the absorbance of the solution was monitored. The spectra have been corrected for the absorbance due to the free peptide and for dilution. (*Inset*) Plot of the concentration of the peptide–Co^{2+} as a function of added Co^{2+} concentration. Experimental points are shown (+). The curve represents a fit to the data using a dissociation constant of 1.0 μM.

spection reveals that the structures of these three regions are quite similar. The structures are characterized by the presence of peptide NH to cysteinyl sulfur hydrogen bonds (57). The paths of the backbones of the two loops from the rubredoxins are nearly identical to one another. Each has a glycine residue following the second cysteine. This allows a local conformation that permits a hydrogen bond involving the NH group of the glycine and the C=O group of the first cysteine residue while maintaining the NH···S hydrogen bond involving residue X^6aa and the second cysteine residue. The region from ATCase differs from these in that it has a nonglycine residue following the second cysteine. No NH···O=C hydrogen bond is present but the position and orientation of residue X^6aa is nearly the same as those in the rubredoxin structures. Of the sequences shown in Table 1, 28/46 have a glycine residue in position X^6aa.

Regions corresponding to the Cys-Xaa_4-His loop have been crystallographically observed in three copper-binding proteins (58–60). In each case, the sequence has the form Cys-Xaa-Xaa-Xaa-Gly-His. The structures of these regions appear to be highly similar; a representative structure is shown in Fig. 3. The structure has the thiolate sulfur atom of the cysteine and the δ- rather than the ε-nitrogen coordinated to the metal ion. An NH···S hydrogen bond involving the NH group of the residue that corresponds to X^6aa is present in each case. The region corresponding to X^6aa-X^7aa-Gly^9-His^{10} forms a type II β-turn, a conformation requiring the presence of glycine in the third position. Inspection of Table 1 reveals that glycine is almost invariantly present in this position in the CCHC box sequences.

The proposed model for the CCHC box–metal complex was derived with the assumptions that the structural features for the Cys-Xaa_2-Cys and Cys-Xaa_4-His loops discussed above are present and that the metal coordination site is tetrahedral. The structure of the Cys-Xaa_2-Cys-Xaa_2 region from residues 4–8 of rubredoxin (54) was used as a starting point. The region Xaa-Xaa-Gly-His from *A. denitrificans* azurin (58) was added by superimposing the first residue onto the last residue of the Cys-Xaa_2-Cys loop structure. The torsional angles at the residue corresponding to residue X^6aa and the side-chain torsional angles of the histidine residue were adjusted to move the δ-nitrogen into one of the two remaining positions around the tetrahedral metal ion with a metal–nitrogen distance of 2.0 Å. Only one of the two possible positions was accessible. This defines the absolute

Table 1. Sequences of the form Xaa_2-Cys-Xaa_2-Cys-Xaa_4-His-Xaa_4-Cys-Xaa_2 from retroviral and other sources

Source	Sequence	Ref.
RSV	GL C YT C GSPG H YQAQ C PK	26
	ER C QL C NGMG H NAKQ C RK	
ASV	GL C YT C GSPG H YQAQ C PK	27
	ER C EL C NGMG H NAKQ C RK	
BaEV	DQ C AY C KERG H WTKD C PK	28
FeLV	DQ C AY C KEKG H WVRD C PK	29
MoLV	DQ C AY C KEKG H WAKD C PK	30
MoSV	DQ C TY C EEQG H WAKD C PK	31
BLV	GP C YR C LKEG H WARD C PT	32
	GP C PI C KDPS H WKRD C PT	
RaMLV	DQ C AY C KEKG H WAKD C PK	18
AKVMuLV	DQ C AY C KEKG H WAKD C PK	33
HTLV-I	QP C FR C GKAG H WSRD C TQ	34
	GP C PL C QDPT H WKRD C PR	
HTLV-II	QP C FR C GKVG H WSRD C TQ	35
	GP C PL C QDPS H WKRD C PQ	
HIV (HTLV-III)	VK C FN C GKEG H TARN C RA	36
	KG C WK C GKEG H QMKD C TE	
HIV (LAV)	VK C FN C GKEG H IARN C RA	37
	KG C WK C GKEG H QMKD C TE	
HIV-2	FK C WN C GKEG H SARQ C RA	38
	QG C WK C GKPG H IMTN C PD	
ARV-2	VK C FN C GKEG H IAKN C RA	39
	KG C WR C GREG H QMKD C TE	
SRV-1	GC C FK C GRKG H FAKN C HE	40
	GL C PR C KRGK H WANE C KS	
SIV_{MAC}	IK C WN C GKEG H SARQ C RA	41
	QG C WK C GLMD H VMAK C PN	
SIV_{AGM}	LR C YN C GLFG H MQRQ C PE	42
	TK C LK C GKLG H LAKD C RG	
HERV	GK C YN C GQIG H LKKN C PV	43
	DL C PR C KKGK H WASQ C RS	
MPMV	GC C FK C GKKG H FAKN C HE	44
	GL C PR C KRGK H WANE C KS	
Visna	QK C YN C GKPG H LARQ C RQ	45
	II C HH C GKRG H MQKD C RQ	
EIAV	QT C YN C GKPG H LSSQ C RA	46
	KV C FK C KQPG H FSKQ C RS	
IAP	KA C FN C GRMG H LKKD C QA	47
	KL C YR C GKGY H RASE C R–	
Copia	VK C HH C GREG H IKKD C FH	19
G	PQ C FR C QGFG H TQRY C FL	48
F	VQ C TN C QEYG H TRSY C TL	49
I	LR C KK C LRFG H PTPI C KS	50
CaMV	CR C WI C NIEG H YANE C PN	20
CERV	CR C WV C WIEG H YANE C PN	51

Standard one-letter abbreviations are used. RSV, Rous sarcoma virus; ASV, avian sarcoma virus; BaEV, baboon endogenous retrovirus; FeLV, feline leukemia virus; MoLV, Moloney murine leukemia virus; MoSV, Moloney murine sarcoma virus; BLV, bovine leukemia virus; RaMLV, Rauscher murine leukemia virus; AKVMuLV, AKV murine leukemia virus; HTLV-I, human T-cell leukemia virus I; HTLV-II, human T-cell leukemia virus II; HTLV-III, human T-cell leukemia virus III; HIV, human immunodeficiency virus; LAV, lymphadenopathy-associated virus; HIV-2, human immunodeficiency virus 2; ARV-2, AIDS-associated retrovirus; SRV-1, molecular clone of simian acquired immunodeficiency syndrome; SIV_{MAC}, simian immunodeficiency virus from macaque; SIV_{AGM}, simian immunodeficiency virus from African green monkey; HERV, human endogenous retrovirus; MPMV, Mason–Pfizer monkey virus; Visna, visna lentivirus; EIAV, equine infectious anemia virus; IAP, Syrian hamster intracisternal A particle; *copia*, *Drosophila copia* element; G, *Drosophila* G element; F, *Drosophila* F element; I, *Drosophila* I factor; CaMV, cauliflower mosaic virus; CERV, carnation etched ring virus.

configuration around the metal center. The structure of the His-Xaa_4-Cys loop is less clear since there are no crystallographically characterized examples of such a structure. The region was built assuming reasonable angles along the polypeptide chain placing the cysteinyl sulfur in the final tetrahedral coordination site around the metal with a metal–sulfur distance of 2.3 Å. A β-bend involving residues His^{9}-Xaa^{10}-Xaa^{11}-Xaa^{12} was included in analogy with the conformation following the coordinated histidine in the azurins (58, 59). An alternative structure includes a β-bend involving residues Xaa^{11}-Xaa^{12}-Xaa^{13}-Cys^{14}, analogous to the structure of a metal-chelating Cys-Xaa_4-Cys loop from aspartate transcarbamoylase (53). The structure is shown in Fig. 4.

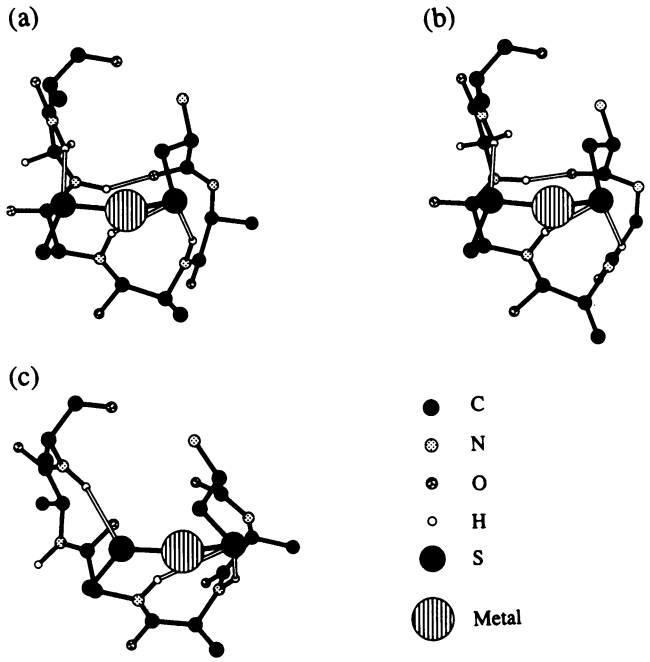


FIG. 2. The structures of crystallographically characterized Cys-Xaa-Xaa-Cys-Xaa-Xaa loops. Only the β carbons of each noncysteine side chain are shown and most hydrogen atoms have been omitted for clarity. The shading scheme shown is used for this and succeeding figures. Covalent bonds are shown as solid bonds and hydrogen bonds are shown as open bonds. (*a*) The sequence Cys^{6}-Thr-Val-Cys-Gly-Tyr from rubredoxin (55). (*b*) The sequence Cys^{39}-Pro-Leu-Cys-Gly-Val from rubredoxin (55). (*c*) The sequence Cys^{138}-Lys-Tyr-Cys-Glu-Lys from the regulatory chain of aspartate transcarbamoylase (53).

DISCUSSION

The CCHC box peptide binds the metal ions Co^{2+} and Zn^{2+}. The spectrum of the Co^{2+} complex strongly suggests a

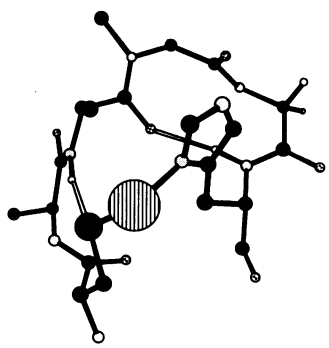


FIG. 3. The structure of a crystallographically characterized Cys-Xaa-Xaa-Xaa-Gly-His loop. Only the β carbons of the noncoordinated side chains are shown. This is taken from *A. denitrificans* azurin (58). Similar structures have been observed in two other copper-binding proteins (59, 60).

tetrahedral metal-binding site and is completely consistent with S_3N coordination. An absorption envelope due to *d–d* transitions is present from 550 to 750 nm. The large extinction coefficients for these bands of $\approx 500\ M^{-1}\cdot cm^{-1}$ are most consistent with a tetrahedral site (62). This conclusion is supported by recent studies of four- and five-coordinate complexes of Co^{2+} with nitrogen and sulfur ligands (63, 64). The spectrum also includes charge transfer bands indicative of metal–cysteinate coordination. A comparison of the absorption spectrum of the Co^{2+} complex of the CCHC box peptide and that of a single zinc finger peptide (13) related to TFIIIA is shown in Fig. 5. The spectra are similar but significant differences are apparent. In particular, the *d–d* absorption envelope of the Co^{2+} complex of the CCHC box peptide is red-shifted relative to that of the zinc finger peptide. In addition, the charge transfer bands appear at somewhat lower energies. Importantly, these differences are similar to the differences between the spectra of $Co(SR)_3(N\text{-MeIm})^-$ and $Co(SR)_2(N\text{-MeIm})_2$ where SR represents a substituted benzenethiolate and *N*-MeIm is 1-methylimidazole (64).

The similarity between the CCHC box peptide and the single zinc finger peptide extends to the strength of metal ion binding. The CCHC box peptide binds Co^{2+} with a dissociation constant of 1.0 μM. Under the same conditions, the dissociation constant for the Co^{2+} complex of the TFIIIA-derived peptide is 3.8 μM (65). In addition, in each case Zn^{2+} binds more strongly than does Co^{2+} (13, 65). These observations indicate that the retroviral domains have sufficient affinity for metal ions to bind them *in vivo* under appropriate conditions.

A proposed structure for the complex of a CCHC peptide and a tetrahedral metal ion has been developed. The structure is based on substructures that have been observed experimentally in other metalloproteins. Similar methods have been used to develop a structure proposal for the zinc finger domains from TFIIIA and related proteins (52). Experimental results on a single zinc finger peptide indicated that at least some aspects of this proposal are correct (14). The CCHC box peptide complex structure is based on structures for sequences of the form Cys-Xaa-Xaa-Cys-Xaa-Xaa and Cys-Xaa-Xaa-Xaa-Gly-His. Overlapping these fragments accounts for 9 of the 14 residues in the CCHC box. The structures of the fragments allow prediction of the absolute configuration of the metal ion as S with priorities set $\overset{1}{Cys} > \overset{4}{Cys} > \overset{9}{His} > \overset{14}{Cys}$. This absolute configuration is forced by the "U-shaped" conformation of the Cys-Xaa-Xaa-Cys-Xaa-Xaa region. The structural proposal is supported by further analysis of the CCHC box sequences. First, proline

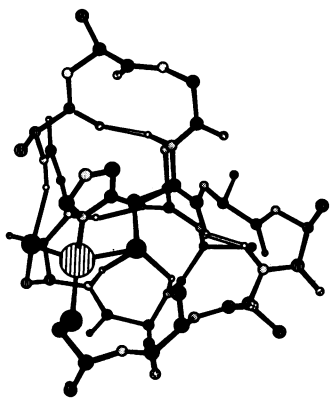


FIG. 4. A proposed three-dimensional structure for the metal complex of a Cys-Xaa_2-Cys-Xaa_4-His-Xaa_4-Cys peptide. A ball and stick view of the complex is shown. Only β carbons are shown for noncoordinated side chains.

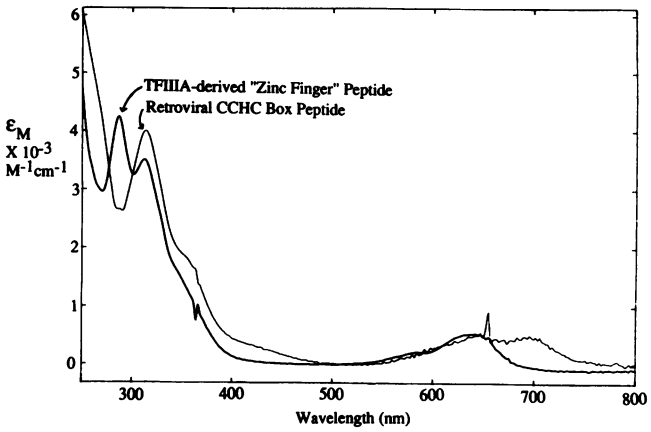


FIG. 5. A comparison of the absorption spectra of the Co^{2+} complexes of a retroviral CCHC box peptide and of a single zinc finger peptide related to TFIIIA. The coordination sphere of the tetrahedral Co^{2+} consists of three thiolates and one imidazole in the CCHC box peptide and two thiolates and two imidazoles in the zinc finger peptide.

is observed only in positions $\overset{2}{Xaa}$, $\overset{7}{Xaa}$, and $\overset{10}{Xaa}$. In the proposed structure, none of these residues is involved in a hydrogen bond involving its NH group, an interaction that would obviously be precluded by the presence of proline. Second, there is a clear tendency for there to be a large hydrophobic residue in either position $\overset{2}{Xaa}$ or $\overset{3}{Xaa}$ but not both. In the structure proposed for the $\overset{1}{Cys}$-$\overset{2}{Xaa}$-$\overset{3}{Xaa}$-$\overset{4}{Cys}$ region, the side chains for residues $\overset{2}{Xaa}$ and $\overset{3}{Xaa}$ are oriented in nearly the same direction. Thus, a large side chain from either position could occupy nearly the same position in space.

It has recently been claimed that retroviral CCHC box regions are not zinc binding fingers on the basis of experimental studies of avian myeloblastosis virus (22). The purified nucleocapsid protein did not contain stoichiometric amounts of zinc or other metal ions. Furthermore, addition of zinc to the protein did not significantly affect the affinity of the protein for poly(ethenoadenylic acid) or its circular dichroism spectrum. Finally, studies of both this virus (22) and Moloney murine leukemia virus (23) have revealed that, in their purified forms, virions do not contain nearly enough zinc to fill the potential metal-binding sites in the nucleocapsid proteins. It seems that the apparent inconsistency between these observations and those described here can be accounted for by consideration of two points. First, based on our experience with this and other cysteine-rich metal-binding peptides, air oxidation of the cysteine residues occurs readily and results in a loss of metal-binding activity. Although bound metal ions do protect the cysteine residues to some extent, oxidation still occurs. Thus, the purified protein from avian myeloblastosis virus may be in some partially oxidized form that is incapable of binding metal ions. The potential presence of a disulfide bond in this protein (which contains two CCHC boxes) was noted (22). Clearly, further studies of these proteins in which the oxidation states of the cysteine residues are more clearly defined will be important in clarifying this point. Second, the assay that has been generally used for these proteins is a nonspecific nucleic acid binding assay. Several observations suggest that this assay may not be a good indicator of the full biological activity of these proteins. The proteins are generally very basic so that binding to polyanions such as nucleic acids is not unexpected. Alkylation (66, 67) or oxidation to cysteic acid (67) of the cysteine residues in certain of these proteins did not dramatically affect the behavior of the proteins in nonspecific nucleic acid binding assays, suggesting that the form of the cysteine residues is not important for this activity. In con-

trast, changing any of the cysteines individually to serine by site-directed mutagenesis results in loss of infectivity (22, 23). It is difficult to reconcile these results unless these proteins serve a function in addition to nonspecific nucleic acid binding. Indeed, recent results suggest that these proteins may facilitate certain nucleic acid annealing reactions (68). Interestingly, zinc and dithiothreitol were included in the reaction medium, although the results of not including them were not reported. An intriguing hypothesis that appears to be consistent with the extant data is that the proteins serve a dual function. Inside virions, they bind nonspecifically to the viral RNA, neutralizing some of its negative charge. Inside an infected cell, the proteins bind metal ions (most probably zinc) and facilitate either specific RNA binding reactions (23) or certain of the annealing reactions that are necessary for retroviral replication (68).

Note. Since this paper was submitted for review, South *et al.* (61) reported nuclear magnetic resonance studies of the ^{113}Cd complex of a CCHC box peptide derived from HIV. ^{1}H–^{113}Cd heteronuclear spin–echo difference spectroscopic studies indicated that the three cysteines and the histidine are coordinated to the metal ion. Furthermore, the histidine is coordinated through the ε-nitrogen rather than the δ-nitrogen. Incorporation of this modification into the model is possible.

We thank Joyce Lilly and Prof. Chris Anfinson for assistance with the peptide synthesis and Scott Michael for useful discussions concerning the structure proposal. This work was supported by the National Institutes of Health (GM 38230, BRSG SO7 RR 7041), the Camille and Henry Dreyfus Foundation, the Searle Scholar Program/Chicago Community Trust, The National Science Foundation (Presidential Young Investigator Award), and the Exxon Education Foundation.

1. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4,** 1609–1614.
2. Brown, R. S., Sander, C. & Argos, P. (1985) *FEBS Lett.* **186,** 271–274.
3. Hanas, J. S., Hazunda, D. J., Bogenhagen, D. F., Wu, F.-Y. & Wu, C.-W. (1983) *J. Biol. Chem.* **258,** 14120–14125.
4. Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12,** 464–475.
5. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52,** 1–3.
6. Berg, J. M. (1989) *Met. Ions Biol. Syst.* **25,** 235–254.
7. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987) *Cell* **51,** 1079–1091.
8. Nagai, K., Nakaseko, Y., Nasmyth, K. & Rhodes, D. (1988) *Nature (London)* **332,** 284–287.
9. Eisen, A., Taylor, W. E., Blumberg, H. & Young, E. T. (1988) *Mol. Cell. Biol.* **8,** 4552–4556.
10. Diakun, G. P., Fairall, L. & Klug, A. (1986) *Nature (London)* **324,** 698–699.
11. Tso, J. Y., Van den Berg, D. I. & Korn, L. J. (1986) *Nucleic Acids Res.* **14,** 2187–2200.
12. Vrana, K. E., Churchill, M. A., Tullius, T. D. & Brown, D. D. (1988) *Mol. Cell Biol.* **8,** 1684–1696.
13. Frankel, A. D., Berg, J. M. & Pabo, C. O. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4841–4845.
14. Párraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. & Klevit, R. E. (1988) *Science* **241,** 1489–1492.
15. Berg, J. M. (1986) *Science* **232,** 485–487.
16. Giedroc, D. P., Keating, K. M., Williams, K. R., Konigsberg, W. H. & Coleman, J. E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8452–8456.
17. Culp, J. S., Webster, L. C., Friedman, D. J., Smith, C. L., Huang, W.-J., Wu, F. Y.-H., Rosenberg, M. & Ricciardi, R. P. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6450–6454.
18. Henderson, L. E., Copeland, T. D., Sowder, R. C., Smythers, G. & Oroszlan, S. (1981) *J. Biol. Chem.* **256,** 8400–8406.
19. Mount, S. M. & Rubin, G. M. (1985) *Mol. Cell. Biol.* **5,** 1630–1638.
20. Franck, A., Guilley, H., Jonard, G., Richards, K. & Hirth, L. (1980) *Cell* **21,** 285–294.
21. Covey, S. N. (1986) *Nucleic Acids Res.* **14,** 623–633.
22. Jentoft, J. E., Smith, L. M., Xiangdong, F., Johnson, M. & Leis, J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7094–7098.
23. Gorelick, R. J., Henderson, L. E., Hansen, J. P. & Rein, A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8420–8424.
24. Schiff, L. A., Nibert, M. L. & Fields, B. N. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4195–4199.
25. Riddles, P. W., Blakeley, R. L. & Zerner, B. (1983) *Met. Enzymol.* **91,** 49–60.
26. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32,** 853–869.
27. Misono, K. S., Sharief, F. S. & Leis, J. (1980) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **39,** 1611.
28. Tamura, T. (1983) *J. Virol.* **47,** 137–145.
29. Copeland, T. D., Morgan, M. A. & Oroszlan, S. (1984) *Virology* **133,** 137–145.
30. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293,** 543–548.
31. Van Beveren, C., van Straaten, F., Galleshaw, J. A. & Verma, I. M. (1981) *Cell* **27,** 97–108.
32. Copeland, T. D., Morgan, M. A. & Oroszlan, S. (1983) *FEBS Lett.* **156,** 37–40.
33. Herr, W. (1984) *J. Virol.* **49,** 471–478.
34. Copeland, T. D., Oroszlan, S., Kalyanaraman, V. S., Sarngardharan, M. G. & Gallo, R. C. (1983) *FEBS Lett.* **162,** 390–395.
35. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S. Y., Miwa, M. & Sugimura, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3101–3105.
36. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. J., 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,** 277–284.
37. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) *Cell* **40,** 9–17.
38. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987) *Nature (London)* **326,** 662–669.
39. Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D. & Luciw, P. A. (1985) *Science* **227,** 484–492.
40. Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. B., Barr, P. J. & Luciw, P. A. (1986) *Science* **231,** 1567–1572.
41. Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P. & Sonigo, P. (1987) *Nature (London)* **328,** 543–547.
42. Fukasawa, M., Miuri, T., Hasegawa, A., Morikawa, S., Tsujimoto, H., Miki, K., Kitamura, T. & Hayami, M. (1988) *Nature (London)* **333,** 457–461.
43. Ono, M., Yasunaga, T., Miyata, T. & Ushikubo, H. (1986) *J. Virol.* **60,** 589–598.
44. Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, S. (1986) *Cell* **45,** 375–385.
45. Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. & Wain-Hobson, S. (1985) *Cell* **42,** 369–382.
46. Stephens, R. M., Casey, J. W. & Rice, N. R. (1986) *Science* **231,** 589–594.
47. Ono, M., Toh, H., Miyata, T. & Awaya, T. (1985) *J. Virol.* **55,** 387–394.
48. Di Nocera, P. P. (1988) *Nucleic Acids. Res.* **16,** 4041–4053.
49. Di Nocera, P. P. & Casari, G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5843–5847.
50. Fawcett, T., Lister, C. K., Kellett, E. & Finnegan, D. J. (1986) *Cell* **47,** 1007–1015.
51. Hull, R., Sadler, J. & Longstaff, M. (1986) *EMBO J.* **5,** 3083–3090.
52. Berg, J. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 79–82.
53. Honzatko, R. B., Crawford, J. L., Monaco, H. L., Ladner, J. E., Ewards, B. F. P., Evans, D. R., Warren, S. G., Wiley, D. C., Ladner, R. C. & Lipscomb, W. N. (1982) *J. Mol. Biol.* **160,** 219–263.
54. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1980) *J. Mol. Biol.* **138,** 615–633.
55. Adman, E. T., Sieker, L. C., Jensen, L. H., Bruschi, M. & LeGall, J. (1977) *J. Mol. Biol.* **112,** 113–120.
56. Sieker, L. C., Stenkamp, R. E., Jensen, L. H., Prickril, B. & LeGall, J. (1986) *FEBS Lett.* **208,** 73–76.
57. Adman, E. T., Watenpaugh, K. D. & Jensen, L. H. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4854–4858.
58. Baker, E. N. (1988) *J. Mol. Biol.* **203,** 1071–1095.
59. Adman, E. T. & Jensen, L. H. (1981) *Isr. J. Chem.* **21,** 8–12.
60. Guss, J. M., Merritt, E. A., Phizackerly, R. P., Hedman, B., Murata, M., Hodgson, K. O. & Freeman, H. C. (1988) *Science* **241,** 806–812.
61. South, T. L., Kim, B. & Summers, M. F. (1989) *J. Am. Chem. Soc.* **111,** 395–396.
62. Bertini, I. & Luchinat, C. (1984) *Adv. Inorg. Biochem.* **6,** 71–111.
63. Corwin, D. T., Jr., Fikar, R. & Koch, S. A. (1987) *Inorg. Chem.* **26,** 3079–3080.
64. Corwin, D. T., Jr., Gruff, E. S. & Koch, S. A. (1988) *Inorg. Chim. Acta* **155,** 5–6.
65. Berg, J. M. & Merkle, D. L. (1989) *J. Am. Chem. Soc.*, in press.
66. Leis, J. & Jentoft, J. (1983) *J. Virol.* **48,** 361–369.
67. Karpel, R. L., Henderson, L. E. & Oroszlan, S. (1988) *J. Biol. Chem.* **262,** 4961–4967.
68. Prats, A. C., Sarih, L., Gabus, C., Litvak, S., Keith, G. & Darlix, J. L. (1988) *EMBO J.* **7,** 1777–1783.