Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC

TERRI L. SOUTH AND MICHAEL F. SUMMERS

Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland 21228

(RECEIVED August 7, 1992; REVISED MANUSCRIPT RECEIVED September 11, 1992)

Abstract

The nucleic acid interactive properties of a synthetic peptide with sequence of the N-terminal CCHC zinc finger (CCHC = Cys-X₂-Cys-X₄-His-X₄-Cys; X = variable amino acid) of the human immunodeficiency virus (HIV) nucleocapsid protein, Zn(HIV1-F1), have been studied by ¹H NMR spectroscopy. Titration of Zn(HIV1-F1) with oligodeoxyribonucleic acids containing different nucleotide sequences reveals, for the first time, sequence-dependent binding that requires the presence of at least one guanosine residue for tight complex formation. The dynamics of complex formation are sensitive to the nature of the residues adjacent to guanosine, with residues on the 3' side of guanosine having the largest influence. An oligodeoxyribonucleotide with sequence corresponding to a portion of the HIV-1 psi-packaging signal, d(ACGCC), forms a relatively tight complex with Zn(HIV1-F1) ($K_d = 5 \times 10^{-6}$ M). Two-dimensional nuclear Overhauser effect (NOESY) data indicate that the bound nucleic acid exists predominantly in a single-stranded, A-helical conformation, and the presence of more than a dozen intermolecular NOE cross peaks enabled three-dimensional modeling of the complex. The nucleic acid binds within a hydrophobic cleft on the peptide surface. This hydrophobic cleft is defined by the side chains of residues Val¹, Phe⁴, Ile¹², and Ala¹³. Backbone amide protons of Phe⁴ and Ala¹³ and the backbone carbonyl oxygen of Lys² that lie within this cleft appear to form hydrogen bonds with the guanosine O6 and NH atoms, respectively. In addition, the positively charged side chain of Arg¹⁴ is ideally positioned for electrostatic interactions with the phosphodiester backbone of the nucleic acid. The structural findings provide a rationalization for the general conservation of these hydrophobic and basic residues in CCHC zinc fingers, and are consistent with site-directed mutagenesis results that implicate these residues as direct participants in viral genome recognition.

Keywords: human immunodeficiency virus; NMR; nucleocapsid protein; viral genome recognition; zinc finger

All retroviruses, including human immunodeficiency virus (HIV), encode a gag precursor polyprotein that functions in the recognition of viral RNA and in the assembly of virus particles (Bolognesi et al., 1978; Dickson et al., 1985). Subsequent to assembly and budding, the gag polyproteins are processed by the viral protease to give smaller proteins, including the nucleocapsid (NC) protein. Without exception, all retroviral gag and NC proteins contain one or two copies of a conserved amino acid sequence termed a CCHC array, Cys-X₂-Cys-X₄-His-X₄-Cys, where X = variable amino acid (Henderson et al., 1981), and

numerous lines of experimental evidence support the proposal (Berg, 1986) that the arrays function by coordinating zinc (Green & Berg, 1989, 1990; Roberts et al., 1989; South et al., 1989, 1990a,b, 1991; Summers et al., 1990, 1992; Fitzgerald & Coleman, 1991; Omichinski et al., 1991). In particular, recent zinc-edge extended X-ray absorption fine structure (EXAFS) measurements on intact retroviruses reveal that tightly associated viral zinc, which is present in quantities sufficient to populate the arrays (Bess et al., 1992), is coordinated to the CCHC zinc fingers of the NC proteins in mature virions (Chance et al., 1992; Summers et al., 1992). The three-dimensional structures of the HIV1-NC protein and its constituent zinc finger domains have been determined to high atomic resolution by NMR methods (Summers et al., 1990, 1992;

Reprint requests to: Michael F. Summers, Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland 21228.

Omichinski et al., 1991; South et al., 1991; Morellet et al., 1992).

Although NC zinc fingers have received extensive spectroscopic and structural characterization, a functional assay for this class of nucleic acid binding proteins is currently unavailable. Unlike for DNA-binding proteins, where the DNA binding sequences are generally well characterized, little is known about the sequence or tertiary structures of the nucleic acid targets of the gag and NC proteins. A major problem is that the CCHC zinc finger arrays apparently bind to different targets during the course of the retrovirus life cycle. Thus, site-directed mutagenesis experiments reveal that these arrays are essential for viral genome packaging (Gorelick et al., 1988, 1990; Meric & Goff, 1989; Aldovini & Young, 1990), and further indicate that sequence specificity is conferred by the CCHC zinc fingers (Meric & Goff, 1989). Proper RNA packaging also requires an essential Psi-packaging signal that may serve as the primary site for sequence- (or structure-) specific protein-RNA binding. This Psi sequence is comprised of ca. 44 nucleotides and is located between the splice donor and the gag initiation codon near the 5'-end of the viral genome (Lever et al., 1989; Aldovini & Young, 1990), and deletion of as few as 21 nucleotides within the Psi sequence is sufficient to abolish RNA packaging (Lever et al., 1989; Aldovini & Young, 1990).

After processing, ca. 2,500–3,000 copies of the NC protein contribute to a histone-like NC-RNA ribonucleoprotein complex (Smith & Bailey, 1979; Chen et al., 1980; Henderson et al., 1981; Copeland et al., 1984; Meric et al., 1984; Karpel et al., 1987; Meric & Goff, 1989; Aldovini & Young, 1990; Dupraz et al., 1990; Gorelick et al., 1990). Thus, at this stage of the retrovirus life cycle, the CCHC arrays appear to facilitate nonspecific binding to nucleic acid. The NC proteins are involved in other nucleic acid interactive processes that occur during infectivity, and appear to facilitate viral RNA dimerization, tRNA primer hybridization, and nucleic acid unwinding during reverse transcription (Meric & Spahr, 1986; Prats et al., 1988, 1991; Meric & Goff, 1989; Bieth et al., 1990; Cornille et al., 1990).

Given the absence of an established functional assay, nucleic acid interactive studies of these proteins have generally involved the use of fluorescent nucleic acids, homopolymers, or RNA fragments as probes (Karpel et al., 1987; Luban & Goff, 1991; Delahunty et al., 1992; Kahn & Giedroc, 1992; Summers et al., 1992). Recent studies of the HIV-1 NC protein that involved detection of the intrinsic fluorescence of the Trp residue of the C-terminal CCHC zinc finger indicate that nucleic acid binding properties of the NC protein are dependent on nucleic acid sequence (Summers et al., 1992). To gain additional insights into the nature of this apparent sequence-specific binding, we have carried out high-resolution NMR studies of oligonucleotide binding to a peptide with sequence

of the N-terminal CCHC zinc finger of the HIV-1 NC protein, Zn(HIV1-F1). The peptide exhibits preferential binding to oligodeoxyribonucleic acids that contain at least one guanosine, and binding appears to be potentiated by the nature of the nucleotides adjacent to the guanosine, with the 3' nucleotide having the largest influence. Tight binding is abolished by removal of zinc from the peptide with a high excess of ethylenediamine tetracetate (EDTA), demonstrating that zinc coordination is essential for sequence-specific nucleotide binding. A five-residue oligonucleotide with sequence corresponding to a segment of the Psi-packaging element, dACGCC, binds Zn(HIV1-F1) with a dissociation constant of ca. 5×10^{-6} M, and nuclear Overhauser effect (NOE) NMR studies enabled three-dimensional modeling of the Zn(HIV1-F1):d(ACGCC) complex. Hydrogen bonding and electrostatic interactions are identified that appear to serve as the determinants of sequence-specific binding. The dependence of nucleic acid binding on zinc coordination and on nucleotide sequence has important implications for the rational design of a new class of HIV antivirals that target retroviral CCHC zinc fingers (Rice et al., 1992).

Results and discussion

Sequence-dependent binding of Zn(HIV1-F1) to nucleic acids

Initial studies of Zn(HIV1-F1)-nucleic acid interactions were carried out with oligodeoxyribonucleic acids due to their availability and favorable stability towards degradation relative to oligoribonucleic acids. It was shown earlier that Zn(HIV1-F1) binds to d(TTTGGTTT), resulting in large chemical shift changes for the Phe⁴ and Ile¹² side-chain proton signals, and also that the peptide destabilizes the duplex form of (d(GGAATTCC))₂ (South et al., 1990b). As an extension of these studies, we have carried out similar NMR experiments with the somewhat simpler oligonucleotide, d(TTTGTTT). Figure 1 shows portions of the ¹H NMR spectra obtained for d(TTTGTTT) (bottom), for Zn(HIV1-F1) (middle), and for a 2:1 mixture of these species (0.98 mM Zn(HIV1-F1), 1.96 mM d(TTTGTTT)). Several prominent changes occur upon mixing, including very large upfield shift changes for the Ile-γCH₃ and δCH₃ proton signals and concomitant upfield shifting (and broadening) of the Phe-aromatic proton signals. In addition, the guanosine-H8 proton signal is broadened and shifted to a somewhat lower field. Similar findings were reported previously in preliminary NMR studies with d(TTTGGTTT) (South et al., 1990b; Delahunty et al., 1992).

The chemical shift changes observed upon addition of d(TTTGTTT) to Zn(HIV1-F1), which occur only for a few residues of the peptide, suggest that peptide:nucleic acid binding occurs via interactions with the Phe⁴ and Ile¹² hydrophobic surface residues. To test our early hy-

Fig. 1. Selected regions of the ^1H NMR spectra obtained in D_2O solutions for the oligodeoxyribonucleotide, d(TTTGTTT) (bottom, 0.4 mM; the signals at 0.8–1.0 ppm are due to impurities), for the CCHC zinc finger peptide, Zn(HIV1-F1) (middle, 0.98 mM), and for a mixture of Zn(HIV1-F1) and d(TTTGTTT) (top, 0.98 mM peptide and 1.96 mM nucleic acid). Complex formation between Zn(HIV1-F1) and d(TTTGTTT) results in large upfield chemical shifts for the peptide Phe⁴ and Ile¹² proton signals, and in broadening and shifting of several of the base and deoxyribose proton signals of the nucleic acid. Similar spectral changes are observed for other oligodeoxyribonucleic acids that contain at least one guanosine residue (see text).

pothesis that binding may be dominated by hydrophobic interactions (South et al., 1990b) we carried out similar NMR studies with a series of oligonucleotides, d(TTTN TTT); N = dT, dA, dC, and dU. As shown in Figure 2, the Phe⁴ aromatic proton signals are generally broadened upon addition of nucleic acid, indicating that all sequences interact with Zn(HIV1-F1) to some extent. However, none of these oligonucleotides produce the dramatic broadening and large chemical shift changes for the Phe⁴ and Ile¹² proton signals that were observed for the sequence with N = dG (Fig. 1). These data clearly demonstrate that Zn(HIV1-F1) binds to d(TTTGTTT) in a manner that is unique compared to its interactions with other d(TTTNTTT) sequences.

Potential interactions with other oligodeoxyribonucleotide sequences that lack a guanosine residue have been carried out, including studies with oligonucleotides d(TTTAATT) and d(TTAAATT). As shown in Figure 3,

titration of Zn(HIV1-F1) with d(TTAAATT) does not lead to significant chemical shift changes for the Phe⁴ and Ile¹² proton signals of the peptide, nor to significant changes in the shifts of the purine, pyrimidine, or deoxyribose proton signals. (No Watson–Crick imino proton signals could be detected for these solutions, indicating that the oligonucleotides do not exist as stable duplexes.) Without exception, the only nucleic acids studied thus far that are capable of producing significant (>0.2 ppm) chemical shift changes of the Phe⁴ aromatic and Ile¹² methyl proton signals are those oligonucleotides that contain at least one guanosine residue.

Additional NMR studies were performed with the oligonucleotides d(TTAGTTT) and d(TTTGATT) to determine if Zn(HIV1-F1):nucleic acid binding is sensitive to the nature of the residues adjacent to the central guanosine (Gua). The upfield regions of the ^1H NMR spectra obtained upon titration of Zn(HIV1-F1) with these oli-

Fig. 2. Selected portions of the ^1H NMR spectra obtained in D_2O solution for $\text{Zn}(\text{HIV1-F1})$ (bottom) and upon addition of one equivalent of oligodeoxyribonucleotides with sequences: $\text{d}(\text{TTTNTTT})$, $\text{N} = \text{dA, dC, dT, dU}$. Although some signal broadening is observed for the Phe^4 aromatic protons, which probably reflects weak, nonspecific interactions, none of these oligonucleotides induces the large chemical shift changes for the peptide Phe and Ile protons that is observed upon addition of $\text{d}(\text{TTTGTTT})$ to $\text{Zn}(\text{HIV1-F1})$ (see Fig. 1). These data indicate that $\text{Zn}(\text{HIV1-F1})$ binds oligonucleotides that contain guanosine in a manner that is unique compared to oligonucleotide sequences that lack a guanosine residue.

gonucleotides and with $\text{d}(\text{TTTGTTT})$ are shown in Figure 4. The very broad $\text{Ile-}\gamma\text{CH}_3$ signal observed for a solution containing 0.98 mM $\text{Zn}(\text{HIV1-F1})$ and 0.49 mM $\text{d}(\text{TTTGTTT})$ (left, Fig. 4) reflects a $\text{Zn}(\text{HIV1-F1}) \leftrightarrow \text{Zn}(\text{HIV1-F1})\text{:d}(\text{TTTGTTT})$ exchange rate of $\text{ca. } 2 \times 10^3 \text{ s}^{-1}$ (based on a two-site-exchange lineshape simulation). This signal shifts to higher field upon titration with additional nucleic acid, until a total nucleic acid concentration of $\text{ca. } 2 \text{ mM}$ is obtained. Significant additional shifting or narrowing does not occur at higher nucleic acid concentrations. Analysis of the ^1H chemical shift-versus-concentration data yields an estimated dissociation constant for the complex of $1.1 \times 10^{-5} \text{ M}$.

Addition of $\text{d}(\text{TTAGTTT})$ leads to similar spectral changes, although the width at half-height of the $\text{Ile-}\gamma\text{CH}_3$ signal does not narrow to the same extent as observed for $\text{d}(\text{TTTGTTT})$ (Fig. 4). In contrast, addition of $\text{d}(\text{TTTGATT})$ results in substantial broadening of the Ile-

γCH_3 signal, which remains broad even after titration to a $\text{d}(\text{TTTGATT})\text{:Zn}(\text{HIV1-F1})$ ratio of 2:1. This line broadening probably results from dynamic structural averaging; therefore, these data do not allow a quantitative evaluation of the relative binding affinities of the different sequences, and additional thermodynamic and kinetic studies are in order. However, the data show clearly that zinc finger peptide–nucleic acid binding is sensitive to the nature of the residues adjacent to the high-affinity guanosine, with the highest sensitivity apparent for the residue at the 3'-adjacent position.

Dependence on zinc

The above NMR experiments reveal two general types of binding to $\text{Zn}(\text{HIV1-F1})$: (1) relatively strong, regiospecific interactions involving the hydrophobic surface side chains of Phe^4 and Ile^{12} that occur with Gua-containing

Fig. 3. Portions of the ^1H NMR spectra obtained for Zn(HIV1-F1) (bottom), for oligonucleotide d(TTAAATT) (second from bottom), and for Zn(HIV1-F1) in the presence of increasing amounts of d(TTAAATT). The lack of significant chemical shift changes for the peptide and nucleic acid proton signals indicates that binding does not occur or is weak at best.

oligonucleotides, and (2) weaker (possibly nonspecific) interactions that occur for oligonucleotides that do not contain a Gua. To determine if the sequence-dependent binding is sensitive to zinc coordination, NMR studies were performed with the oligonucleotide, d(TTGTT) (Fig. 5). As observed for studies with d(TTGTTT), addition of one equivalent of d(TTGTT) to Zn(HIV1-F1) (1.0 mM) leads to upfield shifts for the Phe⁴ aromatic and Ile¹² methyl proton signals. The Gua-H8 proton signal is also shifted downfield and broadened slightly, and at least one of the thymine (Thy)-H6 proton signals undergoes a small downfield chemical shift (Fig. 5). Chemical shift changes are also observed for specific deoxyribose-H1' signals (Fig. 5). Addition of a 50-fold molar excess of ethylenediamine tetraacetate (EDTA) to this solution results in downfield shifts of the histidine-H2 and -H4 protons, and the chemical shift values of these protons after addition of EDTA are characteristic of histidine in an unligated, Zn-free state (South et al., 1989). Other peptide signals revert to chemical shift values characteristic for a random coil-type conformation. Concomitant with

removal of zinc from HIV1-F1 by EDTA, the Phe⁴ and Ile-CH₃ proton signals revert to their normal chemical shifts, as do the Thy-H6 and deoxyribose H1' proton signals (Fig. 5, top). These data demonstrate that zinc coordination is required for the observed sequence-specific nucleic acid interactive properties of Zn(HIV1-F1).

Interactions with Psi-site analog, d(ACGCC)

Correct packaging of the viral RNA depends on an essential 44-residue Psi-packaging sequence that is located between the splice donor and the gag initiation codon near the 5'-end of the viral genome, and studies of nucleic acid deletion mutant retroviruses have identified segments of the psi-packaging signal that are essential for specific packaging of the viral genome (Lever et al., 1989; Aldovini & Young, 1990). One of these essential segments is 21 bases in length and contains three Gua: 5'-UACGCC AAAAATTTTGACTAG (Aldovini & Young, 1990). The Gua residue located near the 5' end of this segment contains adjacent pyrimidine residues only, making it a po-

Fig. 4. Upfield regions of the ^1H NMR spectra obtained for Zn(HIV1-F1) (bottom, 0.98 mM), for oligodeoxyribonucleotides d(TTTGTTT) (left), d(TTAGTTT) (center), and d(TTTGATT) (right), and for Zn(HIV1-F1) (0.98 mM) in the presence of increasing amounts of these oligonucleotides. Upfield shifts of the Ile 12 - γ -methyl group protons are observed upon titration with each oligonucleotide, indicating that each of these nucleic acids is capable of high-affinity binding. However, the signal observed upon addition of d(TTTGATT) is significantly broader than observed for d(TTAGTTT) and d(TTTGTTT), indicating that the dynamic properties of the complex formed are sensitive to the nature of the residues adjacent to the central guanosine.

tentially attractive candidate for zinc finger binding and structural studies. Figure 6 shows the ^1H NMR spectra obtained upon addition of the corresponding DNA analog, d(ACGCC), to Zn(HIV1-F1). Preliminary analysis of titration data obtained at a peptide concentration of 0.9 mM indicate a dissociation constant of ca. 5×10^{-6} M for d(ACGCC). The data in Figure 6 were obtained with a peptide concentration of 9.8 mM, and at this high concentration, the equilibrium is shifted (due to the quadratic nature of the equilibrium relationship) such that $>98\%$ of the nucleic acid is bound to peptide when equimolar amounts of peptide and nucleic acid are present. Thus, as shown in Figure 6, the Ile and Phe side chain protons are significantly broadened upon addition of 0.5 equivalents of d(ACGCC) but narrow considerably and move to higher-field chemical shifts upon addition of one equivalent of d(ACGCC).

NMR studies of the Zn(HIV1-F1):d(ACGCC) complex

Two-dimensional (2D) nuclear Overhauser effect (NOESY) and homonuclear Hartmann-Hahn (HOHAHA) spectra were obtained for the Zn(HIV1-F1):d(ACGCC) complex in D_2O and in 90% $\text{H}_2\text{O}/10\%$ D_2O solutions. In addition, 2D rotating frame Overhauser effect (ROESY, $\tau_m = 120$ ms) and double quantum-filtered homonuclear correlated (2QF-COSY) spectra were obtained for the complex in D_2O and in 90% $\text{H}_2\text{O}/10\%$ D_2O solutions, respectively. (Literature citations for these pulse sequences are given in the Materials and methods section.) The NOESY data were collected with a 300-ms mixing period because earlier studies of the isolated peptide indicated a maximum NOE for most protons at ca. 350 ms (Summers et al., 1990). The NMR data were processed and analyzed

Fig. 5. Selected regions of the ^1H NMR spectra obtained for $\text{Zn}(\text{HIV1-F1})$ (bottom, 0.98 mM), oligonucleotide d(TTGTT) (second from bottom), and for a solution containing $\text{Zn}(\text{HIV1-F1})$ and d(TTGTT) (second from top). Broad, upfield-shifted signals associated with the Phe^4 and Ile^{12} protons are characteristic of $\text{Zn}(\text{HIV1-F1})$:nucleic acid complex formation. The top spectrum was obtained after addition of EDTA (50-fold molar excess). The peptide signals shift to values characteristic of a zinc-free, random coil structure, indicating that EDTA has effectively removed zinc from the peptide. Concomitant with removal of zinc, the nucleic acid signals revert to chemical shifts observed in the absence of $\text{Zn}(\text{HIV1-F1})$, indicating that zinc coordination is essential for high-affinity guanosine-specific binding.

in the usual manner (Wüthrich, 1986), affording complete assignment of the peptide and nucleotide ^1H NMR spectra (Table 1). The chemical shifts of the peptide resonances are consistent with shifts observed previously for the uncomplexed peptide (Summers et al., 1990), with the exception that signals for several residues are shifted to higher field (shift differences exceeding 0.2 ppm are denoted by bold characters in Table 1).

The downfield region of the NOESY spectrum obtained in 90% $\text{H}_2\text{O}/10\%$ D_2O solution is shown in Figure 7. Intra-peptide NOE cross-peak patterns and intensities were essentially identical to those observed for the free peptide, indicating that the peptide does not undergo a significant structural change upon binding to d(ACGCC). The only notable exceptions include the presence of additional NOEs from the Ala^{13} backbone amide, which appears to be in slower chemical exchange with solvent protons compared to the free peptide. As discussed below, the struc-

tural data suggest that this proton participates in a hydrogen bond in the complex (but not in the free peptide). Interestingly, moderate cross relaxation is observed between the Ala^{13} and Phe^4 backbone amide protons (Fig. 7). This NOE is significant because, as discussed below, the structural data indicate that both of these protons are hydrogen bonded to a common oxygen atom (O6) on Gua.

Portions of the 2D NOESY spectra obtained for the $\text{Zn}(\text{HIV1-F1})$:d(ACGCC) complex that contain important intranucleic acid NOE cross peaks are shown in Figure 8. The purine-H8 and pyrimidine-H6 protons exhibited NOE cross peaks to the $\text{H}2''$ and $\text{H}2'$ protons of the 5' neighboring residue. Intense intraresidue signals were observed only for the H6/H8 to $\text{H}2'$ cross peaks in the 2D ROESY spectrum, indicating that the nucleotides exist in an anti conformation. In the upper portion of Figure 8, weak cross peaks involving the cytidine (Cyt)- $\text{H}5$ - $\text{H}2',2''$

Fig. 6. Selected regions of the ^1H NMR spectra of Zn(HIV1-F1) (bottom, 9.8 mM), d(ACGCC) (second from bottom), and Zn(HIV1-F1) (9.8 mM) in the presence of 0.5 (second from top) and 1.0 (top) equivalents of d(ACGCC). At these high stoichiometric concentrations, $>98\%$ of the nucleic acid is bound to peptide ($K_d = \text{ca. } 5 \times 10^{-6} \text{ M}$).

(preceding residue) are observed, but these signals are not observed in the 2D ROESY spectrum, indicating that they result from spin diffusion. In contrast, the moderately intense interresidue Cyt-H5-H3' (preceding) cross peaks observed in the NOESY spectrum (bottom portion) were also observed in the ROESY spectrum (Supplementary material, Diskette Appendix). Interresidue cross peaks from base H6/H8 protons to H1' protons of the preceding residue were not observed in either the NOESY (Fig. 8) or ROESY spectra (Supplementary material). These data are collectively consistent with a predominantly A-helical nucleic acid conformation, based on conformational interpretations using interproton distances reported for idealized A-, B-, and Z-type DNA helices (Wüthrich, 1986). For example, sequential base H8/H6-deoxyribose H3' distances in Z, B, and A helices are $>5 \text{ \AA}$ (3.4 $\text{\AA}$ for H6), 4.6 $\text{\AA}$, and ca. 3.1 $\text{\AA}$, respectively, and corresponding signals of moderate intensity were observed in the 2D NOESY and ROESY spectra. The sequential base H5-deoxyribose H3' distances in idealized Z, B, and A helices are 4.2 $\text{\AA}$, $>5 \text{ \AA}$, and 3.3 $\text{\AA}$, and corresponding signals

of moderate intensity were observed for these proton pairs.

More than 20 intermolecular NOE cross peaks were observed in the 2D NOESY data. Many of these cross peaks are visible in the regions of the 2D NMR data displayed in Figure 8, including: Cyt⁴-H1' to Val¹- γCH_3 , Cyt⁴-H1' to Val¹- γCH_3 , Ile¹²- δCH_3 to Gua³-H8, Ala¹³-CH₃ to Gua³-H8, Ala¹³-CH₃ to Cyt²-H5, Ala¹³-CH₃ to Cyt²-H1', Phe⁴-aromatic to Gua³-H8, Phe⁴-aromatic to Cyt²-H6. Except for signals associated with the Phe⁴ aromatic protons, which undergo very fast T1 ρ relaxation, these intermolecular dipolar cross peaks were also observed in the 2D ROESY spectrum. Other weaker intermolecular cross peaks were observed in the 2D NOESY spectrum but not in the ROESY spectrum, and because it is possible that these latter NOE cross peaks result from spin diffusion, these signals were not used to generate distance restraints for structure refinement (see below). Thus, for structure refinement, a minimal set of NOE distance restraints was employed using only those NOEs for which direct dipolar cross relaxation could be confirmed in the ROESY spectrum.

Table 1. ^1H NMR chemical shift assignments for Zn(HIV1-F1) and d-ACGCC^a

Peptide chemical shifts							
Residue	NH	CaH	CbH	Others			
V(1)	^b	3.83 (3.69)	2.14 (2.01)	γCH_3	0.61 , 0.77 (0.82 , 0.88)		
K(2)	8.42 (8.17)	4.51 (4.41)	1.64 (1.43)	γCH_2	1.25, 1.13 (1.20)		
				δCH_2	1.54 (1.50)		
				ϵCH_2	3.04 (2.51)		
C(3)	8.42 (8.27)	4.19 (4.10)	1.82 (1.78)				
			2.84 (2.79)				
F(4)	9.21 (8.67)	4.74 (4.50)	3.18 (3.15)	H2-6	6.73-7.35 (7.32-7.38)		
			2.61				
N(5)	10.03 (9.58)	4.87 (4.75)	2.80 (2.80)	Amide	7.04, 8.33 (7.07, 8.21)		
			3.29 (3.13)				
C(6)	8.85 (8.83)	5.02 (4.88)	2.65 (2.53)				
			3.31 (3.20)				
G(7)	7.83 (7.93)	3.90 (3.77)					
		4.20 (4.09)					
K(8)	8.34 (8.17)	4.45 (4.37)	1.80 (1.80)	γCH_2	1.68 (1.47, 1.62)		
			1.93 (1.89)	δCH_2	1.54 (1.76)		
				ϵCH_2	3.10 (3.01)		
E(9)	8.34 (8.38)	4.67 (4.50)	1.94 (1.87)	γCH_2	2.25 (2.25, 2.28)		
			1.99				
G(10)	8.48 (8.43)	3.82 (3.72)					
		4.50 (4.40)					
H(11)	7.14 (7.25)	4.74 (4.74)	3.20 (3.15)	H2	7.41 (7.42)		
			3.10	H4	6.89 (6.87)		
I(12)	7.50 (7.84)	4.47 (4.69)	1.68 (2.10)	γCH_2	1.09, 0.93 (1.21)		
				γCH_3	-0.04 (0.87)		
				δCH_3	0.58 (0.90)		
A(13)	9.09 (8.72)	4.69 (4.75)	1.68 (1.40)				
R(14)	8.96 (8.22)	4.25 (4.12)	1.72 (1.60)	γCH_2	1.65, 1.59 (1.60)		
			1.99 (1.90)	δCH_2	3.27 (3.15)		
N(15)	8.01 (8.01)	5.07 (5.05)	2.56 (2.59)	Amide	6.96, 7.87 (7.00, 7.91)		
			3.10 (3.11)				
C(16)	7.72 (7.65)	3.78 (3.77)	3.09 (3.02)				
			3.33 (3.33)				
R(17)	8.34 (8.38)	4.58 (4.55)	1.62 (1.56)	γCH_2	1.36 (1.56)		
			2.04 (1.99)	δCH_2	3.19 (3.19)		
A(18)	8.41 (8.39)	4.09 (4.07)	1.37 (1.32)				
Oligonucleotide chemical shifts							
Nucleotide	H8, H6	H1'	H2', 2''	H3'	H4'	H5', 5''	Others
Ado ¹	8.17	6.27	2.54, 2.63	4.88	4.30	3.81	H2 8.17
Cyt ²	7.81	6.27	2.22, 2.54	4.99	4.37	4.13	H5 5.96
Gua ³	8.15	6.23	2.55, 2.55	5.04	4.50	4.19	
Cyt ⁴	7.92	6.18	2.15, 2.46	4.77	4.30	4.18	H5 6.10
Cyt ⁵	7.94	6.23	2.31, 2.45	4.55	4.15	4.07	H5 6.12

^a Chemical shifts in ppm relative to internal H₂O (4.725 ppm). Signals in parentheses are for free peptide. Bold indicates chemical shift differences >0.2 ppm.

^b Signal for terminal NH₃⁺ not observed. There was a typographical error in the original assignment of H(11) amide. This value has been corrected from 8.30 to 7.25 ppm.

Structural modeling

Zn(HIV1-F1):d(ACGCC) structures were generated with DSPACE as follows. First, distance restraints for the peptide only were used to generate and refine the peptide structure independent of the nucleic acid. Intra-peptide

proton–proton distance restraints of 2.0–2.5 Å, 2.0–3.5 Å, and 2.0–4.5 Å were assigned to qualitatively observed strong, medium, and weak cross-peak intensities, respectively. As in previous structural studies, distance restraints were used to enforce Zn–S and Zn–N^e distances of 2.3 and 2.0 Å, respectively, and additional restraints

Fig. 7. Downfield region of the 300-ms NOESY spectrum obtained for the Zn(HIV1-F1):d(AC GCC) complex (9.8 mM, 90% H₂O/10% D₂O, pH 6.5, T = 30 °C). Amide-to-amide cross-peak patterns and intensities are very similar to those observed for the isolated peptide (Summers et al., 1990), which indicates that the peptide has not undergone a significant structural change upon binding nucleic acid. NOE cross peaks involving the Ala¹³ and Phe⁴ backbone amide protons were not observed in previous studies, possibly due to the enhanced exchange rate of the Ala¹³ amide proton in the isolated peptide. The presence of this cross peak is consistent with our modeling data that indicate that these protons are hydrogen bonded to a common oxygen (O6) on the bound guanosine residue.

were employed to maintain proper sp^2 and sp^3 hybridization geometries of the coordinated N and S atoms, respectively. A total of 10 restraints for the NH-S and NH-O hydrogen bonds identified in earlier studies were included during structure refinement. In addition, we have identified in the present studies an additional hydrogen bond between the Asn¹⁵-NH^e side chain proton and the backbone carbonyl oxygen of Ile¹², and restraints for this hydrogen bond were included in the refinement.

Upper distance restraints of 5.5 Å were employed for all intermolecular NOEs. No intrapeptide distance restraints were employed for the side-chain protons of residues that exhibit NOEs to nucleic acid protons, i.e., for Val¹, Phe⁴, and Ile¹². These conservative restraints were employed in order to allow maximum conformational freedom of the nucleic acid coordinates relative to the peptide coordinates. A complete listing of the intrapeptide distance restraints has been deposited in the Supplementary materials.

Atoms for the peptide were embedded into random positions in 3-space, and variable-velocity simulated annealing and conjugate gradient minimization were performed to refine the peptide geometry. Details of the procedures used, and a complete listing of the macros developed for automated refinement with DSPACE, have been published (Blake et al., 1991, 1992). Final refinement was per-

formed with equal weighting to all types of distance restraints (including restraints for primary and NOE-derived distances), and afforded a structure with a total penalty value (penalty = squared sum of the deviations between the defined distances and the actual distances in the refined model) of 0.022 Å². The backbone folding of this structure was essentially identical to the folding determined previously for the same peptide using different approaches (Summers et al., 1990; Blake et al., 1991).

In preliminary studies, we were unable to adequately define a single-stranded helical nucleic acid geometry on the basis of the sequential interresidue interproton distance restraints alone. Although high-resolution structures have been determined for nucleic acid duplexes (or higher-order structures) from NMR data, where hydrogen-bond, symmetry, and van der Waals restraints involving residues of the complementary strand restrict conformational variability (Nerdal et al., 1988; Koning et al., 1991; Nikonowicz et al., 1991; Schmitz et al., 1991), no structures have been determined for single-stranded nucleic acids in solution on the basis of NMR data (to our knowledge).

Due to the inherent difficulties associated with defining a single-stranded nucleic acid conformation on the basis of NOE restraints, we have employed an approach that involves modeling with a fixed, idealized nucleic acid geometry. As discussed above, the NOE data indicate that

Fig. 8. Portions of the 2D NOESY data obtained for the Zn(HIV1-F1):d(ACGCC) complex in D₂O solution (other solution conditions are as described in the Fig. 7 legend). In the lower portion of the spectrum, interresidue aromatic-to-H3' NOE cross peaks of moderate intensity are observed, but no aromatic-to-H1' cross peaks are observed. These and other interresidue NOE data are consistent with an A-helical-type conformation for d(ACGCC) when bound to Zn(HIV1-F1). The upper plot shows several of the intermolecular NOE cross peaks, including: Gua³-H8 to Ile¹²-δCH₃ and to Ala¹³-CH₃, Cyt⁴-H5,6 to Ile¹²-CH₃, Cyt⁴-H1' to Val¹-γ,γ'CH₃, Cyt²-H5 to Ala¹³-CH₃, and broad signals from Phe⁴-aromatic to Cyt²-H6, Cyt²-H1', and Cyt²-H2'.

d(ACGCC) exists predominantly as an A-type single-stranded helix when bound to Zn(HIV1-F1). As such, coordinates for d(ACGCC) were fixed to an idealized single-stranded A-helix using the Quanta software package, and these coordinates were subsequently converted to DSPACE coordinates with an in-house program (pdb2xyz). Simulated annealing and conjugate gradient minimization were then performed with DSPACE in such a manner that the nucleic acid atoms were held in fixed positions and only the atoms of the peptide were allowed to move. Additional structures were generated by performing a 5 Å translation of the Zn(HIV1-F1) coordinates to separate the peptide from the nucleic acid, followed by variable velocity simulated annealing and conjugate gradient minimization.

It must be emphasized that, with the above modeling approach, we are only testing to see if an A-helical, single-stranded d(ACGCC) model is capable of satisfying the NOE restraints by “docking” with the Zn(HIV1-F1) coordinates. Although conformation space of the peptide is adequately sampled, no statistical sampling of conformational space is performed for the nucleic acid. It would therefore be improper to consider this as a “nucleic acid structure determination” in the usual sense. On the other hand, as indicated above, the NMR data are fully consistent with an A-helical nucleic acid conformation. As such,

we believe that this approach, which represents the most conservative modeling approach that we can take at the present time, is appropriate. As an additional control experiment, structural modeling was carried out in an identical manner using an idealized B-helical d(ACGCC) conformation.

Structure refinement and identification of intermolecular hydrogen bonds

A total of 80 structures generated by the above approach using an idealized A-helical nucleic acid conformation were refined to total penalties in the range 0.028–12.2 Å². A total of 20 of these structures exhibited penalty values of 0.04 Å² or less. Structures with penalty values greater than 0.04 Å² contained severe peptide:nucleic acid van der Waals contacts, improper peptide phi/psi angles, and/or improper side-chain conformations, and these structures were not utilized for further analyses. By comparison, the penalties obtained upon refinement with a B-helical nucleic acid conformation were in the range 0.9–4.2 Å², and this indicates that the B-helical nucleic acid is unable to dock with the peptide without introducing severe violations of the loose intermolecular NOE distance restraints.

Fig. 9. Superposition of 15 DG/SA structures generated with conservative intermolecular distance restraints (1.8–5.5 Å) but no intermolecular hydrogen bond restraints. The figure was generated by superposing the backbone C, C α , N atoms of peptide residues Cys³–Cys¹⁶. Potential hydrogen bonds between Gua³-O6 and the Phe⁴ and Ala¹³ backbone amides and between the Lys²-O and Gua³-NH atoms are denoted by broken green lines (see text).

The backbone C, C α , and N atoms of residues Cys³–Cys¹⁶ of the 20 low-penalty Zn(HIV1-F1):d(ACGCC) (A-helical) structures were superposed, and the resulting coordinates for 15 structures are displayed in Figure 9 and Kinemage 1. Good convergence was observed for these peptide backbone atoms, with pairwise root mean square

(RMS) deviations in the range 0.15–0.33 Å. Most important, the relative position of the nucleic acid is well defined, even though the atoms of the nucleic acid were not used to perform the superposition (Fig. 9). Clearly, the application of very conservative distance restraints is sufficient to define the position of the nucleic acid relative to the peptide (Fig. 9). This is due to the topology of the intermolecular NOEs, which span from Val¹ through Ala¹³ of the peptide and from residue Cyt² through Cyt⁴ of the nucleic acid (Fig. 10, green lines). As noted above, distance restraints from the Phe⁴ aromatic protons were not employed to generate the structures due to very rapid T₂ relaxation, which may result from multiple interconverting conformations of the Phe⁴ phenyl group. However, green lines connecting the Phe⁴ aromatic protons with protons on Cyt² are included in Figure 10 to indicate that the observed NOEs involving Phe⁴ are consistent with the distance geometry models. We cannot, at present, rule out the possibility that the Phe side chain may actually be transiently intercalating between the Cyt²–Gua³ and/or the adenosine¹ (Ado¹)–Cyt² bases.

The above structures provide insights into what may be the molecular-level determinants of sequence-specific binding. As shown in Figure 9, the O6-carbonyl oxygen of Gua³ is placed in relatively close proximity to the backbone amide protons of Phe⁴ (2.2–2.8 Å) and Ala¹³ (3.2–3.9 Å), and the HN1-imine proton of Gua³ is proximal to the backbone carbonyl of Lys² (2.9–3.5 Å). The facts that (1) chemical exchange with solvent of the Ala¹³ NH proton is retarded in the complex, (2) the Ala¹³ and Phe⁴ NH protons exhibit cross relaxation, and (3) both of these protons exhibit significant downfield shifts upon complex formation, are all consistent with the suggestion that these protons are hydrogen bonded to the Gua³-O6 exocyclic oxygen. Unfortunately, attempts to observe the

Fig. 10. Stereo drawing of a representative Zn(HIV1-F1):d(ACGCC) structure (peptide backbone in blue and nucleic acid in red) showing the topology of the more prominent intermolecular NOEs observed (green).

NMR signal for the Gua³-HN1 proton have failed, presumably due to rapid complex dissociation (k ca. 10^3 s⁻¹; see above) and exchange with other labile protons, including solvent.

To determine the consequences of enforcing intermolecular hydrogen-bond restraints on the total penalties, additional structures were generated using the protocol described above, where hydrogen-bond distance restraints of 1.8–2.7 Å were employed for the Ala¹³NH–Gua³O6, Phe⁴NH–Gua³O6, and Lys²O–Gua³N1H proton pairs. A total of 15 structures refined to total penalties of 0.04 Å² or less, and a superposition of these structures (generated using the backbone atoms of Cys³–Cys¹⁶, as in Fig. 9), are shown in Figure 11 and Kinemage 2. The addition of these hydrogen-bond restraints had a minimal effect on the structure, and served only to move the nucleic acid closer to the peptide by a few tenths of an angstrom.

Description of the structure and biological implications

As discussed above, the conformation of the peptide, generated with NOE-derived distance restraints to the idealized A-helical d(ACGCC) model, is identical to the conformation determined previously in the absence of nucleic acid restraints. The final structures, obtained by “docking” an idealized A-helical nucleic acid to the peptide with simulated annealing and conjugate gradient minimization, exhibit low penalties, reflecting the fact that these structures do not contain significant violations of the primary (van der Waals radii, bond angles, bond lengths) or

Fig. 11. Superposition of 15 DG/SA structures generated with addition of three intermolecular hydrogen-bond restraints (see text and Fig. 9). The figure was generated by superposing the backbone atoms of peptide residues Cys³–Cys¹⁶. The convergence of the nucleic acid coordinates relative to the peptide has improved slightly relative to structures generated without the H-bond restraints (see Fig. 9). Hydrophobic residues that make intermolecular NOE contacts are colored green, and the side chain of Arg¹⁴, which is in close proximity to the phosphodiester backbone, is colored blue.

NOE-derived distance restraints. Most important, the nucleic acid coordinates converged to a unique orientation relative to the peptide, demonstrating that the application

Fig. 12. Space-filling representation of Zn(HIV1-F1) showing the hydrophobic cleft formed by conservatively substituted Val¹, Phe⁴, Ile¹², and Ala¹³ residues (the nucleic acid has been removed for clarity). Within this cleft, the backbone amides of Phe⁴ and Ala¹², and the carbonyl oxygen of Lys², are poised to form hydrogen bonds to Gua³. Side chains are colored according to the Shapely model coloring scheme (blue, basic; green, hydrophobic; red, acidic). The figure was generated with RASTER3D software.

Fig. 13. Space-filling representation of the Zn(HIV1-F1):d(ACGCC) complex (nucleic acid in red) showing the orientation of nucleic acid relative to the hydrophobic cleft and the proximity of the Arg¹⁴ side chain to the phosphodiester backbone. The figure was generated with RASTER3D software.

13, for the requirement of a basic residue at position 14, for the high frequency of Gly residues at positions 7 and 10, and of course, for the absolute requirement of Cys and His residues at positions 3, 6, 11, and 16. The conservation of these residues is even greater when considering the N-terminal zinc finger domains of NC proteins from the 23 known strains of HIV and simian immunodeficiency virus (SIV) (see Scheme 1). Our structural findings are also consistent with the results of site-directed mutagenesis experiments that show that mutations of the conservatively substituted hydrophobic residues within the CCHC arrays alter the RNA recognition and packaging properties of the retroviral gag protein (Meric & Goff, 1989). Differential packaging of wild-type and viral-like RNAs may thus be influenced by the "shape" of the hydrophobic cleft on the CCHC zinc finger surface.

Little is currently known about the mechanism by which the NC protein facilitates infectivity, and DNA binding roles for this protein cannot be ruled out. In this regard, a nucleic acid binding protein from humans that contains seven sequential copies of the CCHC zinc finger motif is capable of binding with sequence specificity to both single-stranded DNA and single-stranded RNA (Rajavashisth et al., 1989). The CCHC zinc finger may thus represent a generic motif for binding single-stranded nucleic acids, analogous to the classical CCHH zinc finger motif found widely in duplex-DNA binding proteins (Summers et al., 1991). At present, however, the only known biological ligand for the HIV-1 gag and NC proteins is RNA.

Although DNA is known to exist in several right- (and left-) handed helical forms, RNA is found predominantly in A-like helical conformations (Saenger, 1983). Most importantly, recent X-ray structural studies of a plant virus capsid particle reveal that the observable RNA with the capsid core exists in a single-stranded, A-helical conformation (Chen et al., 1989). It is therefore significant that the NMR data are consistent with an A-helical d(ACGCC) conformation for the nucleic acid when bound to peptide. This finding supports our use of oligodeoxyribonucleic acids as initial structural probes for the RNA-binding CCHC zinc finger peptide.

Scheme 1. N-terminal CCHC domains of 23 known strains of HIV and SIV NC proteins.

Atomic-level structures have been reported recently for the intact HIV-1 NC protein (Morellet et al., 1992; Summers et al., 1992) and for a synthetic peptide fragment that contains the zinc finger domains of the HIV-1 NC protein (Omichinski et al., 1991). In all of these structures, the conformations of the zinc finger domains are very similar to structures reported previously for isolated synthetic zinc finger peptides (Summers et al., 1990; South et al., 1991). The structure of the virus-isolated protein, determined in our laboratory, is consistent with the findings by Clore, Gronenborn, and coworkers that the N- and C-terminal zinc fingers exist as structurally independent domains on a flexible polypeptide chain (Omichinski et al., 1991; Summers et al., 1992). For these structures, binding of nucleic acids to the intact NC protein could easily occur in the manner described here for the isolated N-terminal zinc finger domain. In contrast, the hydrophobic surfaces of N- and C-terminal zinc fingers of the structure proposed by Roques and coworkers interact with each other, with the aromatic groups of Phe¹⁶ and Trp³⁷ forming what appears to be a mini-hydrophobic core. This structure also appears to be inconsistent with fluorescence measurements that indicate that Trp³⁷ is in a highly polar environment in the native protein (Summers et al., 1992). A large structural reorientation would have to occur in the Roques HIV-1 NC model in order for nucleic acid binding to occur in the same manner as described here for the isolated N-terminal zinc finger.

Summary

We have shown that a peptide with sequence of the N-terminal zinc finger domain of the HIV-1 NC protein is capable of binding selectively to oligodeoxyribonucleotides that contain at least one Gua, and that this type of selective binding is dependent on zinc coordination. The oligonucleotide d(ACGCC), which corresponds to an essential portion of the viral Psi-packaging signal, binds to the zinc finger, and 2D NOESY data indicate that the bound oligonucleotide exists in an A-helical conformation. Binding specificity is apparently conferred by hydrogen bonds with backbone atoms of Lys², Phe⁴, and Ala¹³, and by the shape of the hydrophobic cleft that is defined in part by the side chains of residues Val¹, Phe⁴, Ile¹², and Ala¹³. Additional stability may result from electrostatic interactions between the phosphodiester backbone and the side chain of Arg¹⁴. These findings provide a rationale for the conservative placement of hydrophobic and basic residues at these positions in retroviral CCHC zinc finger sequences. Furthermore, our findings are consistent with mutagenesis results that implicate several of these residues as essential elements for viral genome recognition and packaging.

Materials and methods

NMR data collection and processing

Oligopeptides and deoxyribonucleotides were obtained from Peptide Technologies, Inc. (Washington, D.C.) and from Macromolecular Resources, Inc. (Fort Collins, Colorado). NMR data were collected on a GE Omega-PSG 600 (599.72 MHz, ¹H) NMR spectrometer. Sample conditions for nucleic acid binding studies were as follows: 0.5 μ mol (1.2 mg) of Zn(HIV1-F1) in D₂O solution; pH* 6.2; T = 30 °C; deoxyribonucleotides added in 0.25- μ mol increments. Sample conditions for 2D NMR-based structural studies were as follows: 10 mM (12.5 mg) of Zn(HIV1-F1) in D₂O or 90% H₂O/10% D₂O solution; pH 6.5; T = 30 °C. ¹H chemical shifts were referenced to internal H₂O (4.725 ppm at 30 °C). NMR data were transferred to Silicon Graphics computers for processing with FTNMR and FELIX software (Hare Research, Inc.). All 2D data sets were processed with 3 Hz of Gaussian line broadening in t₂, with a 70° shifted sine-squared window function in the t₁ dimension, and were zero filled to a final matrix size of 2K $\times$ 2K real data points. ¹H NMR signals were assigned by standard procedures (Wüthrich, 1986). Parameters used to collect and process the 2D NMR data are described below.

NOESY

Two-dimensional 600-MHz NOESY data were collected in D₂O solution with a spectral width of 7,017.54; 32 transients for each 512 t₁ increments, with a second set of States-type data collected in a noninterleaved manner for each t₁ increment; 300 ms mixing time. NOESY data were collected for the complex in 90% H₂O/10% D₂O with a pulse sequence employing a Dante pulse train (Hore, 1983) for saturation of the H₂O resonance, followed by a SCUBA sequence (Brown et al., 1988) to allow for dipolar relaxation involving resonances close to the water frequency. In addition, a composite 180° pulse was included in the middle of the mixing interval to invert H₂O magnetization that relaxed during the first half of the mixing time. Relevant parameters are as follows: spectral width = 7,017.54; 1.2 s Dante saturation interval, including 5° pulses separated by 100- μ s intervals; a 40-ms SCUBA delay period; and a 300-ms mixing time, which includes 5-ms homospoil (HS) pulses at the beginning of the mixing period and immediately following the composite 180° pulse.

HOHAHA

HOHAHA (Braunschweiler & Ernst, 1983; Davis & Bax, 1985) spectra were collected with a 90-ms spin lock period that included two 1-ms trim pulses. A total of 16 transients for each of 256 t₁ increments were obtained. HOHAHA data for a sample in 90% H₂O/10% D₂O

were collected with Dante and SCUBA preparation intervals as described above. Other relevant parameters were as described above for NOESY data collection.

ROESY

ROESY (Bothner-by et al., 1984) data were collected in D₂O solution with a spin lock field of 6.25 kHz applied for 120 ms. A total of 32 transients for each of 256 t₁ increments were obtained.

NMR data analysis and signal assignments

Proton NMR signal assignments were made by determining scalar connectivities within amino acid and nucleic acid residues from 2QF-COSY and HOHAHA spectra, and then correlating the signals of adjacent residues on the basis of dipolar connectivities obtained from two-dimensional NOESY data. This general approach has been described in detail (Wüthrich, 1986).

Structure calculations

Structures of the Zn(HIV1-F1):d(ACGCC) complex were calculated with DSPACE™. The program QUANTA™ was used to generate coordinates for d(ACGCC) in idealized A- and B-helical conformations, and these coordinates were converted to DSPACE files with an in-house program (pdb2xyz). Primary restraints, including van der Waals radii, bond distances, and planarity restraints have been published (Blake et al., 1991). Distance restraints were adjusted such that all potential hydrogen-bonding partners (e.g., all NH-O and NH-S(Cys) atom pairs) could approach to internuclear distances equal to 75% of the sum of their van der Waals radii. The nucleic acid conformation was held fixed throughout structure refinement. Multiple structures were generated by performing a 5-Å translation of the nucleic acid coordinates, followed by application of variable-velocity simulated annealing and conjugate gradient minimization. This approach is described in greater detail in the text.

Space-filling representations in Figures 12 and 13 were rendered on a Silicon Graphics computer with RASTER3D (David Bacon, Maryland Center for Advanced Research in Biotechnology, Rockville, Maryland).

Supplementary materials available

A complete list of the distance restraints employed for structural modeling and a portion of the ROESY spectrum showing the relative internucleotide ROE cross-peak intensities used appear on the Diskette Appendix. The distance restraints and the coordinates for the distance geometry structures generated with and without intermolecular hydrogen-bond restraints have been submitted to the Brookhaven Protein Data Bank. The structures are also shown as kinemages.

Acknowledgments

Support from the National Institutes of Health (grants GM42561 and AI30917) is gratefully acknowledged. The psi-packaging sequence of the HIV-1 MN isolate was kindly provided by Dr. Marvin Reitz (NCI). We are especially grateful to Dennis Hare for his unflinching collaborative support.

References

- Aldovini, A. & Young, R.A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **64**, 1920–1926.
- Berg, J.M. (1986). Potential metal-binding domains in nucleic acid binding proteins. *Science* **232**, 485–487.
- Bess, J.W., Jr., Powell, P.J., Issaq, H.J., Schumack, L.J., Grimes, M.K., Henderson, L.E., & Arthur, L.O. (1992). Tightly bound zinc in HIV-1, HTLV-1 and other retroviruses. *J. Virol.* **66**, 840–847.
- Bieth, E., Gabus, C., & Darlix, J.-L. (1990). A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis in vitro. *Nucleic Acids Res.* **18**, 119–127.
- Blake, P.R., Hare, D.R., & Summers, M.F. (1991). Is 2D-NOESY back-calculation necessary for NMR-based structure determination? A systematic study involving a retroviral-type zinc finger peptide. *Techniques Protein Chem.* **11**, 357.
- Blake, P.R., Park, J.-B., Zhou, Z.H., Hare, D.R., Adams, M.W.W., & Summers, M.F. (1992). Solution-state structure by NMR of zinc-substituted rubredoxin from the marine hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Protein Sci.* **1**, 1508–1521.
- Bolognesi, D.P., Montelaro, R.C., Frank, H., & Schafer, W. (1978). Assembly of type C oncornaviruses: A model. *Science* **199**, 183–186.
- Bothner-by, A., Stephens, R.L., Lee, J.T., Warren, C.D., & Jeanloz, R.W. (1984). Structure determination of a tetra-saccharide: Transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811–813.
- Braunschweiler, L. & Ernst, R.R. (1983). Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521–528.
- Brown, S.C., Weber, P.L., & Mueller, L. (1988). Toward complete proton NMR spectra in proteins. *J. Magn. Reson.* **71**, 166–169.
- Chance, M.R., Sagi, I., Wirt, M.D., Frisbie, S.M., Scheuring, E., Chen, E., Bess, J.W., Henderson, L.E., Arthur, L.O., South, T.L., Perez-Alvarado, G., & Summers, M.F. (1992). EXAFS studies of a retrovirus: Equine infectious anemia virus cysteine arrays are coordinated to zinc. *Proc. Natl. Acad. Sci. USA* **89**, 10124–10128.
- Chen, M.-J., Garon, C.F., & Papas, T.S. (1980). Native ribonucleoprotein in an efficient transcriptional complex of avian myeloblastosis virus. *Proc. Natl. Acad. Sci. USA* **77**, 1296–1300.
- Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonosoff, G., & Johnson, J.E. (1989). Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science* **245**, 154–159.
- Copeland, T.D., Morgan, M.A., & Oroszlan, S. (1984). Complete amino acid sequence of the basic nucleic acid binding protein of feline leukemia virus. *Virology* **133**, 137–145.
- Cornille, F., Mely, Y., Ficheux, D., Savignol, I., Gerard, D., Darlix, J.-L., Fournie-Zaluski, M.-C., & Roques, B.P. (1990). Solid phase synthesis of the retroviral nucleocapsid protein NCp10 of Moloney murine leukaemia virus and related “zinc-fingers” in free SH forms. Influence of zinc chelation on structural and biochemical properties. *Int. J. Peptide Protein Res.* **36**, 551–558.
- Davis, D.G. & Bax, A. (1985). Assignment of complex ¹H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. *J. Am. Chem. Soc.* **107**, 2820–2821.
- Delahunty, M.D., South, T.L., Summers, M.F., & Karpel, R.L. (1992). Nucleic acid interactive properties of a peptide corresponding to the N-terminal zinc finger domain of HIV-1 nucleocapsid protein. *Biochemistry* **31**, 6421–6469.
- Dickson, C., Eisenman, R., Fan, J., Hunter, E., & Teich, N. (1985). In *RNA Tumor Viruses*, 2nd Ed., Vols. 1, 2 (Weiss, R., Teich, N.,

- Varmus, H., & Coffin, J., Eds.). Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Dupraz, P., Oertle, S., Meric, C., Damay, P., & Spahr, P.-F. (1990). Point mutations in the proximal Cys-His box of Rous sarcoma virus nucleocapsid protein. *J. Virol.* **64**, 4978–4987.
- Fitzgerald, D.W. & Coleman, J.E. (1991). Physicochemical properties of cloned nucleocapsid protein from HIV. Interactions with metal ions. *Biochemistry* **30**, 5195–5201.
- Gorelick, R.J., Henderson, L.E., Hanser, J.P., & Rein, A. (1988). Point mutants of Moloney murine leukemia virus that fail to package viral RNA: Evidence for specific RNA recognition by a “zinc finger-like” protein sequence. *Proc. Natl. Acad. Sci. USA* **85**, 8420–8424.
- Gorelick, R.J., Nigida, S.M., Bess, J.W., Arthur, L.O., Henderson, L.E., & Rein, A. (1990). Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**, 3207–3211.
- Green, L.M. & Berg, J.M. (1989). A retroviral Cys-Xaa2-Cys-Xaa4-His-Xaa4-Cys peptide binds metal ions: Spectroscopic studies and a proposed three-dimensional structure. *Proc. Natl. Acad. Sci. USA* **86**, 4047–4051.
- Green, L.M. & Berg, J.M. (1990). Retroviral nucleocapsid protein-metal ion interactions: Folding and sequence variants. *Proc. Natl. Acad. Sci. USA* **87**, 6403–6407.
- Henderson, L.E., Copeland, T.D., Sowder, R.C., Smythers, G.W., & Oroszlan, S. (1981). Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J. Biol. Chem.* **256**, 8400–8406.
- Hore, P.J. (1983). Solvent suppression in Fourier transform nuclear magnetic resonance. *J. Magn. Reson.* **55**, 283–300.
- Karpel, R.L., Henderson, L.E., & Oroszlan, S. (1987). Interactions of retroviral structural proteins with single-stranded nucleic acids. *J. Biol. Chem.* **262**, 4961–4967.
- Khan, R. & Giedroc, D.P. (1992). Recombinant human immunodeficiency virus type-I nucleocapsid (NC^{p7}) protein unwinds tRNA. *J. Biol. Chem.* **267**, 6689–6695.
- Koning, T.M.G., Boelens, R., van der Marel, G.A., van Boom, J.H., & Kaptein, R. (1991). Structure determination of a DNA octamer in solution by NMR spectroscopy. Effect of fast local motions. *Biochemistry* **30**, 3787–3797.
- Lever, A., Gottlinger, H., Haseltine, W., & Sodroski, J. (1989). Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **63**, 4085–4087.
- Luban, J. & Goff, S.P. (1991). Binding of human immunodeficiency virus type-I (HIV-1) RNA to recombinant HIV-1 gag polyprotein. *J. Virol.* **65**, 3203–3212.
- Meric, C., Darlix, J.-L., & Spahr, P.-F. (1984). It is Rous sarcoma virus protein p12 and not p19 that binds tightly to Rous sarcoma virus RNA. *J. Mol. Biol.* **173**, 531–538.
- Meric, C. & Goff, S.P. (1989). Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J. Virol.* **63**, 1558–1568.
- Meric, C. & Spahr, P.-F. (1986). Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging. *J. Virol.* **60**, 450–459.
- Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J.-L., & Roques, B.P. (1992). Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by ¹H NMR. *EMBO J.* **11**, 3059–3065.
- Nerdal, W., Hare, D.R., & Reid, B.R. (1988). Three-dimensional structure of the wild-type lac Pribnow promoter DNA in solution. *J. Mol. Biol.* **201**, 717–739.
- Nikonowicz, E.P., Meadows, R.P., Fagan, P., & Gornstein, D.G. (1991). NMR structural refinement of a tandem G·A mismatched decamer d(CCAAGATTGG)₂ via the hybrid matrix procedure. *Biochemistry* **30**, 1323–1334.
- Omichinski, J.G., Clore, G.M., Sakaguchi, K., Appella, E., & Gronenborn, A.M. (1991). Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett.* **292**, 25–30.
- Prats, A.-C., Housset, V., Billy, G., Cornille, F., Prats, H., Roques, B., & Darlix, J.-L. (1991). Viral RNA annealing activities by the nucleocapsid protein of Moloney murine leukemia virus are zinc dependent. *Nucleic Acids Res.* **19**, 3533–3541.
- Prats, A.-C., Sarih, L., Gavus, C., Litvak, S., Kieth, G., & Darlix, J.-L. (1988). Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA. *EMBO J.* **7**, 1777–1783.
- Rajavashisth, T.B., Taylor, A.K., Andalibi, A., Svenson, K.L., & Lussis, A.J. (1989). Identification of a zinc finger protein that binds to the sterol regulatory element. *Science* **245**, 640–643.
- Rice, W.G., Schaeffer, C.A., Harten, B., Villinger, F., South, T.L., Summers, M.F., Henderson, L.E., Bess, J.W., Jr., Arthur, L.O., McDougal, J.S., Orloff, S.L., Mendeleyev, J., & Kun, E. (1992). Zinc ejection from HIV-1 zinc fingers and inhibition of infectivity in human lymphocytes by novel aromatic C-nitroso compounds. *Nature*, in press.
- Roberts, W.J., Pan, T., Elliott, J.I., Coleman, J.E., & Williams, K.R. (1989). p10 single-stranded nucleic acid binding protein from murine leukemia virus binds metal ions via the peptide sequence Cys26-X2-Cys29-X4-His34-X4-Cys39. *Biochemistry* **28**, 10043–10047.
- Saenger, W. (1983). In *Principles of Nucleic Acid Structure* (Cantor, C.R., Ed.). Springer-Verlag, New York.
- Schmitz, U., Pearlman, D.A., & James, T.L. (1991). Solution structure of [d(GTATATAC)]₂ via restrained molecular dynamics simulations with nuclear magnetic resonance constraints derived from relaxation matrix analysis of two-dimensional nuclear Overhauser effect measurements. *J. Mol. Biol.* **221**, 271–292.
- Smith, B.J. & Bailey, J.M. (1979). The binding of an avian myeloblastosis virus basic 12,000 dalton protein to nucleic acids. *Nucleic Acids Res.* **7**, 2055–2072.
- South, T.L., Blake, P.R., Hare, D.R., & Summers, M.F. (1991). C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **30**, 6342–6349.
- South, T.L., Blake, P.R., Sowder, R.C., Arthur, L.O., Henderson, L.E., & Summers, M.F. (1990a). The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers. *Biochemistry* **29**, 7786–7789.
- South, T.L., Kim, B., Hare, D.R., & Summers, M.F. (1990b). Zinc fingers and molecular recognition. Structure and nucleic acid binding studies of an HIV zinc finger-like domain. *Biochem. Pharmacol.* **40**, 123–129.
- South, T.L., Kim, B., & Summers, M.F. (1989). ¹¹³Cd NMR studies of a 1:1 Cd adduct with an 18-residue finger peptide from HIV-1 nucleic acid binding protein, p7. *J. Am. Chem. Soc.* **111**, 395–396.
- Summers, M.F. (1991). Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. *J. Cell. Biochem.* **45**, 41–48.
- Summers, M.F., Henderson, L.E., Chance, M.R., Bess, J.W., Jr., South, T.L., Blake, P.R., Sagi, I., Perez-Alvarado, G., Sowder, R.C., III, Hare, D.R., & Arthur, L.O. (1992). Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci.* **1**, 563–574.
- Summers, M.F., South, T.L., Kim, B., & Hare, D.R. (1990). Structure of an HIV zinc finger via a new NMR-based distance geometry approach. *Biochemistry* **29**, 329–340.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. Wiley, New York.