of its conformation space upon binding. The possible implications of this observation for the development of synthetic vaccines will be outlined elsewhere.* This aspect of recognition is often difficult to probe using natural amino acid replacements due to the difficulty of maintaining the steric and electrostatic features of the natural substrate for the receptor. In utilising the non-natural analogues of *S*-proline, the overall charge of the peptide analogues was maintained as were the number of hydrogen bonding sites available for interaction with the receptor. However, as judged by NMR methods, these non-natural amino acids modified the conformational mobilities of analogues. Their use as replacements for *S*-proline in other biologically important substrates, such as bradykinin, may allow some determination of the relative importance of dynamic properties in systems involving the interaction of flexible proline-containing peptides with cellular receptors.

Acknowledgements—We thank Dr. M. Hann and Dr. D. Reynolds (Glaxo) for the gift of [^{3}H]acetic anhydride. This work was supported by the SERC (U.K.) under the Molecular Recognition Initiative. Additional funding was provided by the Royal Society of Chemistry (Hickinbottom Fellowship to J.A.R.), the Nuffield Foundation (N.G.J.R.), and Glaxo Group Research (U.K.).

† *Department of Chemistry*
The University
Southampton SO9 5NH, U.K.
‖ *Tenovus Research Laboratory*
General Hospital
Tremona Road
Southampton SO9 4XY, U.K.
¶ *Organisch-Chemisches Institut*
der Universität Zürich
CH-8057 Zürich
Winterthurerstrasse 190
Switzerland

NIGEL G. J. RICHARDS†‡
MARK G. HINDS†§
DAVID M. BRENNAND‖
MARTIN J. GLENNIE‖
JOHN M. WELSH¶
JOHN A. ROBINSON‡¶

* Brennand DM *et al.*, manuscript submitted for publication.

‡ To whom correspondence should be addressed.

§ Current address: Organisch-Chemisches Institut, der Universität Zürich, CH-8057 Zürich, Winterthurerstrasse 190, Switzerland.

REFERENCES

1. Takahashi N, Hayano T and Suzuki M, Peptidyl-prolyl *cis-trans* isomerase is the cyclosporin A-binding protein cyclophilin. *Nature* **337**: 473–475, 1989.
2. Fischer G, Wittmann-Liebold B, Lang K, Kiefhaber T and Schmid FX, Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature* **337**: 476–478, 1989.
3. Dyson HJ, Lerner RA and Wright PE, The physical basis for induction of protein reactive antipeptide antibodies. *Annu Rev Biophys Biophys Chem* **17**: 305–324, 1988.
4. Dyson HJ, Rance M, Houghton RA, Lerner RA and Wright PE, Folding of immunogenic peptide fragments of proteins in water solution. *J Mol Biol* **201**: 161–200, 1988.
5. Derome AE, *Modern NMR Techniques in Chemistry Research*. Pergamon Press, Oxford, 1987.
6. Kopple KD, Ohnishi O and Go N, Conformations of cyclic peptides III: Cyclopentaglycyltyrosyl and related compounds. *J Am Chem Soc* **91**: 4264–4272, 1969.
7. Bothner-By AA, Stephens RL, Lee JT, Warren CD and Jeanloz RW, Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame. *J Am Chem Soc* **106**: 811–812, 1984.
8. Still WC and Lipton MA, The multiple minimum problem in molecular modelling. Tree searching internal coordinate conformational space. *J Comput Chem* **9**: 343–355, 1988.
9. Weiner SJ, Kollman PA, Nguyen DT and Case DA, An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* **7**: 230–236, 1986.
10. Abraham DJ and Leo AJ, Extension of the fragment method to calculate amino acid zwitterion and side-chain partition coefficients. *Proteins, Structure, Function Genetics* **2**: 130–152, 1987.
11. Colman PM, Laver WG, Varghese JN, Baker AT, Tulloch PA, Air GM and Webster RG, Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. *Nature* **326**: 358–363, 1987.
12. Burgen ASV, Roberts GCK and Feeney J, Binding of flexible ligands to macromolecules. *Nature* **253**: 753–755, 1975.
13. Burkett U and Allinger NL, *Molecular Mechanics* (ACS Monograph 177), Chap. 3. American Chemical Society, Washington DC, 1982.
14. Wilmot CM and Thornton JM, Analysis and prediction of the different types of β-turn in proteins. *J Mol Biol* **203**: 221–232, 1988.

Biochemical Pharmacology, Vol. 40, No. 1, pp. 123–129, 1990.
Printed in Great Britain.

0006-2952/90 $3.00 + 0.00

Zinc fingers and molecular recognition. Structure and nucleic acid binding studies of an HIV zinc finger-like domain

Retroviruses encode for synthesis of a *gag* polyprotein which binds to viral RNA and anchors it to the cell wall for budding [1–3]. All retroviral *gag* proteins [and their nucleic acid binding protein (NABP*) proteolysis products] contain conserved amino acid sequences of the type C-X_2-C-X_4-H-X_4-C (X = variable amino acid residue) [4–6], which have been implicated in zinc binding [7, 8]. Although NABPs in the mature virus lack zinc [8, 9], probably due to the presence of internal disulfides [10], a growing body of evidence provides support for the physiological relevance of the zinc binding at the *gag*-stage of the retroviral life cycle [7, 8, 11–13]. We have determined the structure of an

* Abbreviations: NABP, nucleic acid binding protein; RMSD, route mean square deviation; 2D NOESY, two-dimensional nuclear Overhauser effect spectroscopy; ROESY, rotating frame Overhauser spectroscopy; and DG, distance geometry.

18-residue peptide with sequence of the first zinc finger-like domain from the *gag* protein p55 of HIV. The peptide [Zn(p55F1)† exhibits Type-I and Type-II NH-S tight turns, and is stabilized both by coordination of the three Cys and one His residues to zinc and by extensive internal hydrogen bonding [14]. The folding of residues C(1)-F(2)-N(3)-C(4)-G(5)-K(6) is virtually identical to the folding observed by X-ray crystallography for related residues in the iron domain of rubredoxin [15]; superposition of all main chain and Cys-side chain atoms of residues C(1)-K(6) of Zn(p55F1) onto residues C(6)-Y(11) and C(39)-V(44) of rubredoxin gives RMSDs of 0.46 Å and 0.35 Å respectively [14]. Conservatively substituted residues implicated in RNA binding [8, 11] form a hydrophobic surface patch. Single-stranded oligodeoxyribonucleotide binding studies reported here confirm that these residues are involved in nucleic acid binding *in vitro*. The structural and nucleic acid binding results provide strong evidence for the proposed zinc finger-like motif in retroviral gene recognition.

The amino acid sequence of p55F1 is V-K-C-F-N-C-G-K-E-G-H-I-A-R-N-C-R-A [1, 2, 16]. p55F1 binds one equivalent of zinc via the Cys sulfurs and His-N^{δ} atoms [12]. The ^{1}H NMR spectrum of the zinc adduct was assigned completely using established 2D NMR spectroscopic procedures [17]. Two-dimensional NOESY data (obtained using a modified version of the published pulse sequence; see Refs. 18 and 19) used to determine interproton distance constraints were obtained using aqueous (90% H_2O/10% D_2O and pure D_2O) solutions with mixing times of 0.005, 0.05, 0.1, 0.3, and 0.5 sec. Volume integrals of all resolved and partially overlapping cross- and auto-peaks [14] in the 2D NOESY spectra were calculated using FTNMR*. Cross-peak intensities in the NOESY data obtained with a mixing time of 100 msec were used to make initial interproton distance estimates. Cross-peaks resulting from direct interproton cross-relaxation were differentiated from peaks resulting from relay effects on the basis of signal phases in the 2D ROESY [20] spectra. Estimates of upper and lower interproton distance constraints were used as initial input for the distance geometry (DG) program, DSPACE*. Families of structures were generated using Zn-S(Cys) and Zn-N(His-N^{δ}) distances of 2.3 and 2.0 Å respectively [21]. Initial structures exhibited considerable backbone variations with non-tetrahedral zinc coordination geometries.

These structures were then used to back-calculate theoretical 2D NOESY spectra using the program BKCALC.* A single cross-relaxation rate constant was used, which afforded accurate simulation of the experimental NOE build-up curves for all resolved geminal protons, including the Gly-α protons, the β protons of the three Cys and two Asn residues, as well as for the slowly-exchanging amide protons of Asn. The theoretical spectra generated with BKCALC contained a number of NOE cross-peaks not observed in the experimental spectra, and minimum internuclear distance constraints were added to the constaints list for the appropriate proton pairs. The above cycle was then repeated until the experimental and theoretical 2D NOESY spectra were visually identical. Structure refinements were made by adjusting the interproton distance constraints to obtain a match between the experimental and theoretical NOE build-up curves and auto-peak decay curves.

No hydrogen bonding or ligand-Zn-ligand bond angle constraints were utilized at any point in the refinement process. Each of the final structures refined to a total penalty (penalty = squared sum of the covalent and experimental bounds violations) of less than 0.21 Å^2, using 226 interproton distance constraints with bounds for the 72 observable distances typically in the range ± 0.05 Å. Anomalous results were obtained only for the C-terminal Ala, which gave intra-residue NOE build-up curves that were all very strong and could not be fit with the -60 sec^{-1} cross-relaxation rate constant. All other residues gave self-consistent inter- and intra-residue NOE build-up profiles. No inconsistent NOEs were observed between three or more protons that would be indicative of dynamic exchange between multiple structures.

Eight superpositioned final structures are shown in Fig. 1. Pairwise RMS deviations for backbone atoms of all but the first two and last two residues, for all backbone atoms, and for all atoms in the structures are in the ranges 0.07 to 0.25 Å, 0.20 to 0.73 Å, and 1.10 to 1.90 Å respectively. Although not implicitly invoked, all of the final structures exhibit tetrahedral or near-tetrahedral Zn coordination geometries, with ligand(1)-Zn-ligand(2) bond angles in the range 96–117°. To facilitate comparison of our structural results as they pertain to other retroviral *gag* proteins, an internal numbering scheme is utilized here wherein the first cysteine in the zinc domain is designated as C(1).

Residues C(1), F(2), N(3), C(4), G(5), and K(6) fold in a manner nearly identical to the folding observed for related residues that bind iron in the X-ray crystal structure of rubredoxin [15]. The N(3) and C(4) backbone amide protons of Zn(p55F1) are oriented in a manner consistent with hydrogen bonding to the C(1) sulfur in what has been described as a Type-I NH-S tight turn [15]. The N-H^{E} side chain amide proton of N(3) appears to be hydrogen bonded to the C(14) sulfur, consistent with the slow chemical exchange rate observed for this proton. The C(4) α-carbon serves as a corner between the Type-I tight turn and a short, orthogonally-directed Type-II NH-S tight turn, with hydrogen bonding between the amide of K(6) and the C(4) sulfur. The carbonyl oxygen is pointing in a direction consistent with hydrogen bonding to the amide hydrogen of C(1), and the amide hydrogen of G(5) exhibits hydrogen bonding to the C(1) carbonyl oxygen. Superposition of all backbone and Cys side chain atoms of residues C(1) through K(6) of a representative Zn(p55F1) DG structure onto the respective atoms of residues C(6) through Y(11) and residues C(39) through V(44) of rubredoxin gives RMSDs of 0.46 Å and 0.35 Å respectively. This folding places the β protons of the K(6) side chain in the vicinity of the H(9) imidazole-H^{2} proton, and also positions the K(-1) α proton in the vicinity of the E(7) α proton. The G(8) α-carbon provides a corner for a 3_{10} turn, with the pro-R α-H in close proximity to the E(7) carbonyl oxygen. Backbone atoms including the carbonyl of G(8) through the amide of I(10) form a β-like stretch. None of the backbone atoms within this stretch appear to be involved in hydrogen bonding.

The A(11) α-carbon serves as a corner that leads to a Type-II tight turn comprising residues A(11) through C(14). This tight turn is stabilized by apparent hydrogen bonding between the A(11) carbonyl oxygen and the C(14) amide proton, and by coordination of the C(14) sulfur to Zn. The methyl group of A(11) is located directly above the F(2) and N(3) backbone atoms, and the pseudo-plane made by the backbone atoms of the Type-II tight turn is nearly perpendicular to the pseudo-plane made by the backbone atoms of the Type-I NH-S tight turn (see Fig. 2). The backbone amide protons of V(−2), K(−1), and R(+1) were not observed in the 2D NOESY spectra due to rapid exchange, and the conformation of the N- and C-terminal residues of Zn(p55F1) were determined with less accuracy.

Two of the most striking structural features exhibited by Zn(p55F1) are (1) the presence of at least seven internal hydrogen bonds within the 14-residue zinc binding domain, and (2) the similarity of the first six residues to related residues in the iron domain of rubredoxin. The presence of a Gly residue at position 5 appears to stabilize Type-II NH-S folding [15]. This requirement is consistent with the conservative evolutionary substitution of glycine at position 5 in retroviral *gag* proteins. In addition, steric requirements

* Previously called Zn(p7$^{13-30}$); see Ref. 12.

† Hare Research, Inc., Woodinville, WA 98072.

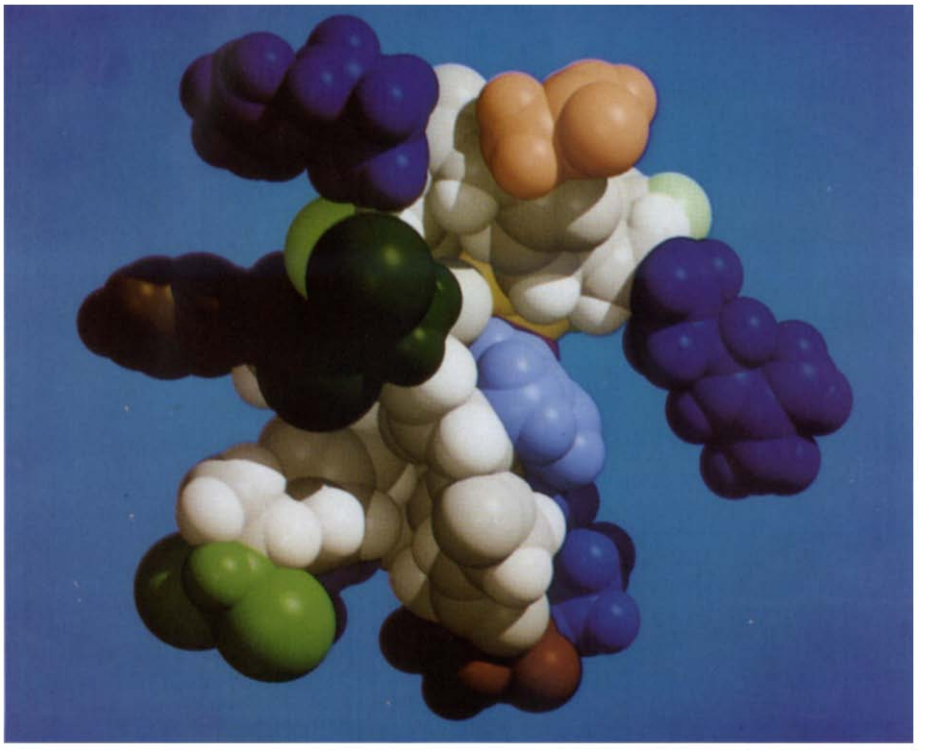

Fig. 3. Space filling representation of Zn(p55F1) DG structure 1, showing the spatial relationships between the side chains of the conservatively substituted amino acid residues that have been implicated in genomic RNA recognition. Methyl groups are displayed as spheres. The figure was generated with RASTER3D using the Shapely Model coloring scheme (backbone atoms, white; basic side chains, blue; acidic side chains, red; hydrophobic side chains, green; hydrophilic side chains, brown; sulfur atoms, yellow; zinc, purple). Reprinted with permission from *Biochemistry* **29**: 329–340, 1990. Copyright (1990) American Chemical Society. [Ref. 14].

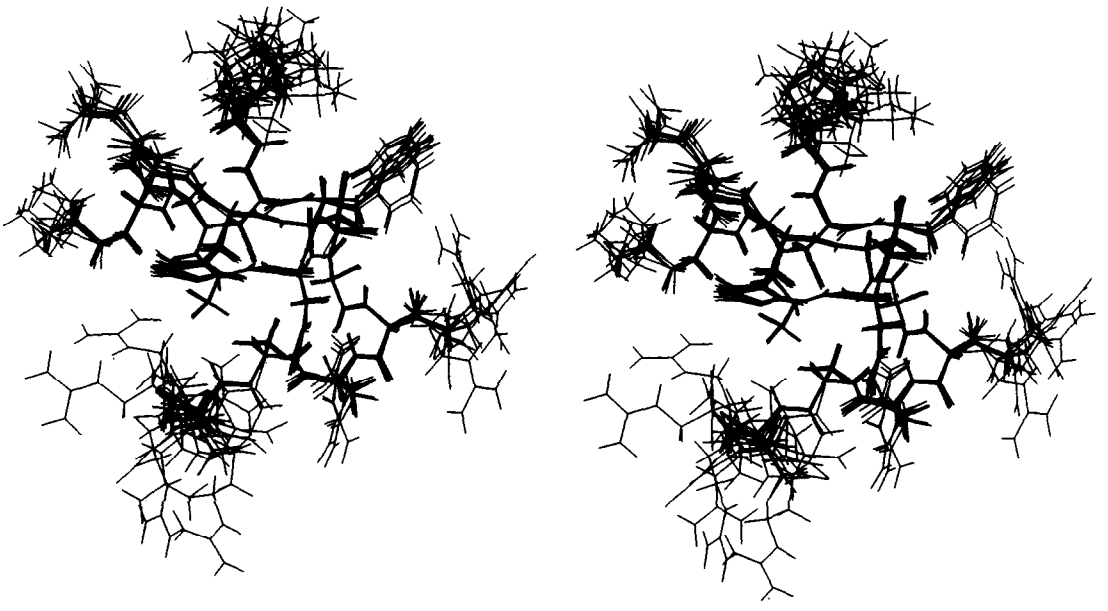

Fig. 1. Superposition (in stereo, obtained using DSPACE) of eight independently generated Zn(p55F1) structures. Reprinted with permission from *Biochemistry* **29**: 329–340, 1990. Copyright (1990) American Chemical Society. [Ref. 14].

associated with the 3_{10} turn at position 8 explain the conservative substitution of glycine at this position. *Thus, the structural constraints associated with the zinc-binding motif provide a rationale for the conservative placement of Gly residues at positions* 5 *and* 8 [14].

The backbone folding implicitly positions the side chains of the conservatively substituted F(2), I(10), R(12) and N(13) residues in the same general spatial location, forming what may be best described as a hydrophobic patch with adjacent positively charged and amide-containing groups (see Fig. 3). The orientation of these conservatively substituted side chains is consistent with the proposal that these residues participate directly in gene recognition by binding to specific RNA structures [8, 11]. In this regard, we have

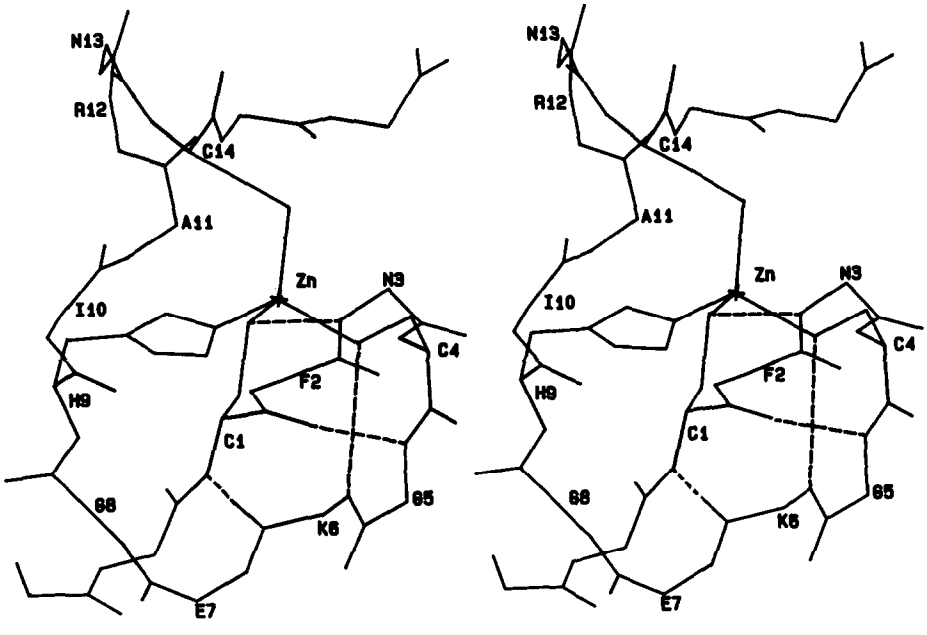

Fig. 2. Stereo diagram of DG structure 1 showing all backbone atoms and the side chain atoms (except protons) for residues C(1), C(4), H(9) and C(14). Dashed lines indicate hydrogen bonds. Reprinted with permission from *Biochemistry* **29**: 329–340, 1990. Copyright (1990) American Chemical Society. [Ref. 14].

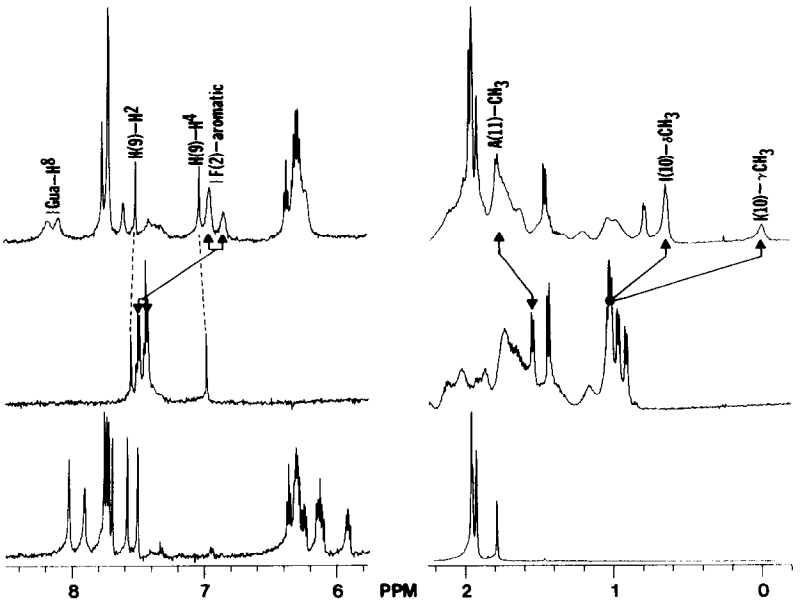

Fig. 4. Portions of the ^{1}H NMR spectra obtained for (A) d(TTTGGTTT), (B) Zn(p55F1), and (C) a mixture of the solutions used to obtain (A) and (B). Signal assignments were made using 2D NOESY and HOHAHA data (not shown).

obtained 1D (Fig. 4) and 2D (not shown) ^{1}H NMR spectra for a solution containing Zn(p55F1) and d(TTTGGTTT) (Fig. 4). Signals for the His aromatic and most of the other protons of Zn(p55F1) are largely unaffected by the presence of oligonucleotide, but the F(2) aromatic and I(10) methyl proton signals exhibit significant upfield shifts of *ca.* 0.5 and 1.0 ppm respectively. In addition, the G-H_8 signals of d(TTTGGTTT) are broadened significantly and shifted to lower field, whereas the T-H_6 and T-CH_3 proton signals are affected only moderately. These preliminary studies confirm that nucleic acid binding via the conservatively substituted hydrophobic residues is possible, at least in isolated finger-like peptides. In fact, the upfield shift observed for the F(2)-aromatic and I(10) methyl protons (brown and dark green spheres, respectively; Fig. 3) and the downfield shift observed for the A(11) methyl protons (light green sphere, Fig. 3) are consistent with a tentative binding model involving intercalation of a nucleobase between the F(2) aromatic and I(10) methyl groups.

The structure of Zn(p55F1) differs considerably from structures predicted [22] and found [23, 24] for the classical DNA-binding zinc finger motif. In particular, the short amino acid sequence between C(4) and H(9) precludes the formation of a helical nucleic acid-binding "finger" structure within the major loop. Instead, it appears that the functionally important nucleic acid binding regions of the retroviral motif are the four-residue tight turns, which may be best described as "knuckles" rather than "fingers".

Acknowledgements—This work was supported by ACS Institutional Research Grant IN-147F, MD Cancer Program/University of Maryland and by NIH Grant GM42561-01 (to M.F.S.).

* Authors to whom correspondence should be addressed. Reprint requests should be addressed to M.F.S.

Department of Chemistry and Biochemistry
University of Maryland
Baltimore County Campus
Baltimore, MD 21228; and
† *Hare Research, Inc. 14810 216th Ave. N.E.*
Woodinville, WA 98072
U.S.A.

TERRI L. SOUTH
BO KIM
DENNIS R. HARE*†
MICHAEL F. SUMMERS*

REFERENCES

1. Dickson C, Eisenman R, Fan J, Hunter E and Teich N, *RNA tumor viruses* (Eds. Weiss R, Teich N, Varum H and Coffin J), 2nd Edn, Vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1985.
2. Dickson C, Eisenman R, Fan J, Hunter E and Teich N, *RNA tumor Viruses* (Eds. Weiss R, Teich N, Varmus H and Coffin J), 2nd Edn, Vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1985.
3. Bolognesi DP, Montelaro RC, Frank H and Schafer W, Assembly of type C oncornaviruses: A model. *Science* **199**: 183–186, 1978.
4. Henderson LE, Copeland TD, Sowder RC, Smythers GW and Oroszlan S, Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J Biol Chem* **256**: 8400–8406, 1981.
5. Copeland TD, Morgan MA and Oroszlan S, Complete amino acid sequence of the basic nucleic acid binding protein of feline leukemia virus. *Virology* **133**: 137–145, 1984.
6. Karpel RL, Henderson LE and Oroszlan S, Interactions of retroviral structural proteins with single-stranded nucleic acids. *J Biol Chem* **262**: 4961–4967, 1987.
7. Berg JM, Potential metal binding domains in nucleic acid binding proteins. *Science* **232**: 485–487, 1986.

8. Gorelick RJ, Henderson LE, Hanser JP and Rein A, 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, 1988.
9. Jentoft JE, Smith LM, Fu X, Johnson M and Leis J, Conserved cysteine and histidine residues of the avian myeloblastosis virus nucleocapsid protein are essential for viral replication but are not "zinc-binding fingers." *Proc Natl Acad Sci USA* **85**: 7094–7098, 1988.
10. Jentoft JE, Smith LM and Secnik J, Structure and RNA binding by the avian retroviral nucleocapsid protein. *Biophys J* **55**: 584a, 1989.
11. Meric C and Goff SP, 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, 1989.
12. South TL, Kim B and Summers MF, ^{113}Cd 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, 1989.
13. Green LM and Berg JM, A retroviral Cys-Xaa_2-Cys-Xaa_4-His-Xaa_4-Cys peptide binds metal ions: Spectroscopic studies and a proposed three-dimensional structure. *Proc Natl Acad Sci USA* **86**: 4047–4051, 1989.
14. Summers MF, South TL, Kim B and Hare DR, Structure of an HIV zinc finger via a new NMR-based distance geometry approach. *Biochemistry* **29**: 329–340, 1990.
15. Adman E, Watenpaugh KD and Jensen LH, NH . . . S hydrogen bonds in *Peptococcus aerogenes* ferredoxin, *Clostridium pasteurianum* rubredoxin, and *Chromatium* high potential iron protein. *Proc Natl Acad Sci USA* **72**: 4854–4858, 1975.
16. Wain-Hobson S, Sonigo P, Danos O, Cole S and Alizon M, Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**: 9–17, 1985.
17. Wuthrich K, *NMR of Proteins and Nucleic Acids*. John Wiley, New York, 1986.
18. Sklenar V and Bax A, Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J Magn Reson* **74**: 469–479, 1987.
19. Sklenar V, Brooks BR, Zon G and Bax A, Absorption mode two-dimensional NOE spectroscopy of exchangeable protons in oligonucleotides. *FEBS Lett* **216**: 249–252, 1987.
20. Bax A, Sklenar V, Clore GM and Gronenborn AM, Water suppression in two-dimensional spin-locked nuclear magnetic resonance experiments using a novel phase-cycling procedure. *J Am Chem Soc* **109**: 6511–6513, 1987.
21. Diakun GP, Fairall L and Klug A, EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature* **324**: 698–699, 1986.
22. Berg JM, Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc Natl Acad Sci USA* **85**: 99–102, 1988.
23. Parraga G, Horvath SJ, Eisen A, Taylor WE, Hood L, Young ET and Klevit RE, Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* **241**: 1489–1492, 1988.
24. Lee MS, Gippert GP, Soman KV, Case DA and Wright PE, Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**: 635–637, 1989.