Identification and Solution Structure of a Highly Conserved C-terminal Domain within ORF1p Required for Retrotransposition of Long Interspersed Nuclear Element-1*

Kurt Januszyk, Patrick Wai-Iun Li, Valerie Villareal, Dan Branciforte, Haihong Wu, Yongming Xie, Juli Feigon, Joseph A. Loo, Sandra L. Martin, and Robert T. Clubb

From the Department of Chemistry and Biochemistry, UCLA-Department of Energy Institute for Genomics and Proteomics, Molecular Biology Institute, **David Geffen School of Medicine, UCLA, Los Angeles, California 90095-1570 and the Human Medical Genetics Program and Department of Cell and Developmental Biology, University of Colorado School of Medicine, Aurora, Colorado 80045

Long interspersed nuclear element-1 (LINE-1 or L1) retrotransposons comprise a large fraction of the human and mouse genomes. The mobility of these successful elements requires the protein encoded by open reading frame-1 (ORF1p), which binds single-stranded RNA with high affinity and functions as a nucleic acid chaperone. In this report, we have used limited proteolysis, filter binding, and NMR spectroscopy to characterize the global structure of ORF1p and the three-dimensional structure of a highly conserved RNA binding domain. ORF1p contains three structured regions, a coiled-coil domain, a middle domain of unknown function, and a C-terminal domain (CTD). We show that high affinity RNA binding by ORF1p requires the CTD and residues within an amino acid protease-sensitive segment that joins the CTD to the middle domain. Insights in the mechanism of RNA binding were obtained by determining the solution structure of the CTD, which is shown to adopt a novel fold consisting of a three-stranded β sheet that is packed against three α-helices. An RNA binding surface on the CTD has been localized using chemical shift perturbation experiments and is proximal to residues previously shown to be essential for retrotransposition, RNA binding, and chaperone activity. A similar structure and mechanism of RNA binding is expected for all vertebrate long interspersed nuclear element-1 elements, since residues encoding the middle, protease-sensitive segment, and CTD are highly conserved.

* This work was supported by National Institutes of Health Grants GM57487 (to R. T. C.) and GM40367 (to S. L. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by the W. M. Keck Foundation for establishing the UCLA Proteomics Center and the UCLA Jonsson Comprehensive Cancer Center.

‡‡ To whom correspondence may be addressed: Dept. of Cell and Developmental Biology, University of Colorado School of Medicine, P.O. Box 6511, MS 8106, Aurora, CO 80045. Tel.: 303-724-3467; Fax: 303-724-3420; E-mail: sandy.martin@uchsc.edu.

‡‡¶ To whom correspondence may be addressed: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1570. Tel.: 310-206-2334; Fax: 310-206-4749; E-mail: rclubb@mbi.ucla.edu.

‡¶‡¶ The atomic coordinates and structure factors (code 2JRB) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡¶‡¶‡ Supported by the W. M. Keck Foundation for establishing the UCLA Proteomics Center and the UCLA Jonsson Comprehensive Cancer Center.

‡¶‡¶‡‡ To whom correspondence may be addressed: Dept. of Cell and Developmental Biology, University of Colorado School of Medicine, P.O. Box 6511, MS 8106, Aurora, CO 80045. Tel.: 303-724-3467; Fax: 303-724-3420; E-mail: sandy.martin@uchsc.edu.

‡¶‡¶‡‡‡ To whom correspondence may be addressed: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1570. Tel.: 310-206-2334; Fax: 310-206-4749; E-mail: rclubb@mbi.ucla.edu.

‡¶‡¶‡‡¶‡‡ Supported by the W. M. Keck Foundation for establishing the UCLA Proteomics Center and the UCLA Jonsson Comprehensive Cancer Center.

‡¶‡¶‡‡‡‡ To whom correspondence may be addressed: Dept. of Cell and Developmental Biology, University of Colorado School of Medicine, P.O. Box 6511, MS 8106, Aurora, CO 80045. Tel.: 303-724-3467; Fax: 303-724-3420; E-mail: sandy.martin@uchsc.edu.

‡¶‡¶‡‡‡‡‡ To whom correspondence may be addressed: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1570. Tel.: 310-206-2334; Fax: 310-206-4749; E-mail: rclubb@mbi.ucla.edu.

‡¶‡¶‡‡‡‡‡‡ Supported by the W. M. Keck Foundation for establishing the UCLA Proteomics Center and the UCLA Jonsson Comprehensive Cancer Center.

‡¶‡¶‡‡‡‡‡‡ To whom correspondence may be addressed: Dept. of Cell and Developmental Biology, University of Colorado School of Medicine, P.O. Box 6511, MS 8106, Aurora, CO 80045. Tel.: 303-724-3467; Fax: 303-724-3420; E-mail: sandy.martin@uchsc.edu.

‡¶‡¶‡‡‡‡‡‡‡ To whom correspondence may be addressed: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1570. Tel.: 310-206-2334; Fax: 310-206-4749; E-mail: rclubb@mbi.ucla.edu.

‡¶‡¶‡‡‡‡‡‡‡‡ Supported by the W. M. Keck Foundation for establishing the UCLA Proteomics Center and the UCLA Jonsson Comprehensive Cancer Center.

‡¶‡¶‡‡‡‡‡‡‡‡ To whom correspondence may be addressed: Dept. of Cell and Developmental Biology, University of Colorado School of Medicine, P.O. Box 6511, MS 8106, Aurora, CO 80045. Tel.: 303-724-3467; Fax: 303-724-3420; E-mail: sandy.martin@uchsc.edu.

‡¶‡¶‡‡‡‡‡‡‡‡‡ To whom correspondence may be addressed: Department of Chemistry and Biochemistry, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1570. Tel.: 310-206-2334; Fax: 310-206-4749; E-mail: rclubb@mbi.ucla.edu.
A Highly Conserved C-terminal Domain within ORF1p

RNA as the template. A double-stranded DNA copy of L1 is then synthesized and joined to the target site through a poorly understood process that may require DNA repair machinery supplied by the host (22).

ORF1p is essential for retrotransposition and has at least two functions (21). First, it binds to L1 RNA during the cytoplasmic phase of the replication cycle and may thereby prevent its degradation. Evidence for this comes from immunoelectron staining of embryonal carcinoma cells, which has revealed punctate cytoplasmic structures when probed with anti-ORF1p antibodies (23). These structures presumably contain several ORF1 proteins bound to full-length L1 RNA, since large ribonucleoprotein particles containing both of these molecules can be fractionated from intact F9 (mouse) (24) and HeLa (human) cells (16). The ribonucleoprotein particle probably protects L1 RNA from degradation by cellular nucleases and via the RNA interference pathway (25). It may also facilitate RNA maturation into a retrotransposition-competent structure and/or nuclear import (21). ORF1p may also directly participate in the target site-primed reverse transcription reaction by facilitating nucleic acid strand transfers involved in reverse transcription. This is supported by in vitro studies, which have shown that it is a nucleic acid chaperone; the isolated protein accelerates strand annealing, promotes melting of mismatched duplexes, and in displacement assays facilitates strand exchange (17). Interestingly, the chaperone activity appears to be required for retrotransposition, since single amino acid ORF1p mutants have been identified that disrupt this function and retrotransposition (15) but are otherwise capable of forming ribonucleoprotein particles (16).

The structure of ORF1p from the retrotransposition-competent mouse Tf-type L1 element has been characterized in the greatest detail (42.9 kDa, 371 amino acids) (15). Sedimentation and atomic force microscopy (AFM) studies indicate that it is a dumbbell-shaped homotrimer that is presumably held together by a trimeric coiled-coil (26). Residues at the C-terminal end of ORF1p show the greatest degree of sequence conservation and bind and chaperone nucleic acids (27,28). In this report, limited proteolysis has been used to identify three distinct domains: the middle, and C-terminal domains. The latter two domains are connected by an ~40-amino acid protease-sensitive linker that in combination with the C-terminal domain (CTD) mediates high affinity RNA binding. NMR has been used to determine the solution structure of the CTD and has elucidated a positively charged surface that interacts with RNA.

MATERIALS AND METHODS

Proteolysis and Mass Spectrometry Analysis of Spa-type ORF1p—Full-length ORF1 protein from the mouse spa-type L1 element was purified from baculovirus-infected insect cells as described previously (15). Partial proteolysis was performed by adding sequencing grade trypsin (Sigma) to 10 μl of protein (~2 mg/ml) either on ice or at room temperature (buffer: 50 mM phosphate, pH 7.6, 250 mM NaCl, and 0.15 mM EDTA buffer). The reaction was quenched by the addition of either an equal volume of SDS-PAGE loading dye or 2% trifluoroacetic acid. Samples containing loading dye were loaded onto a NuPage 4–12% Bis-Tris gel (Invitrogen). After separation, the gel was stained with Coomassie Brilliant Blue staining buffer and fixed (29). Bands were excised, and the resulting gel cuts were washed multiple times with 100 mM ammonium bicarbonate and 50% acetonitrile, 50 mM ammonium bicarbonate. Trypsin (10 μl of a 20 ng/μl solution) was then added to each of the gel cuts for 4 h at 37 °C. The gel cut was washed multiple times with 50% acetonitrile, 1% trifluoroacetic acid. The supernatant was collected and dried, using a SpeedVac (Thermo). The dried sample was then re-suspended in 0.1% formic acid. Liquid chromatography (LC)-MS/MS of peptide mixtures was conducted on an LC Packings nano-LC system with a nanoelectrospray interface (Protana) and quadrupole TOF mass spectrometer (QSTAR XL; Applied Biosystems), as described in detail previously (30). Detected ions were then analyzed using the MASCOT data base search engine (Matrix Science) and the Pro ID program (Applied Biosystems) to determine the identity of the peptides. Positive protein identification was based on standard MASCOT criteria for statistical analysis of LC-MS/MS data. For matrix-assisted laser desorption/ionization (MALDI)-MS measurements, peptides from samples that were quenched with 2% trifluoroacetic acid were desalted using a C4 ZipTip (Millipore). MALDI-TOF (Voyager DE-STR; Applied Biosystems) was then employed to determine the masses of the proteolysis fragments (fragments 1–8) after co-crystallization with matrix (sinapinic acid).

Sample Preparation of Proteins Used in NMR and Proteolysis Studies—Expression plasmids encoding fragments of ORF1p from the Md-A-type L1 element were generated by PCR using L1 Md-A101 as a template (a gift from H. Kazazian; GenBankTM number AY053456). Nucleotide sequences encoding residues 230–357 and 261–347 were cloned into a pET11a expression vector (Novagen) to produce the ORF1CTD proteins, respectively. Both expression vectors were sequenced in the forward and reverse direction to confirm their fidelity. The ORF1CTD proteins were overexpressed and purified in an identical manner from Rosetta-DE3 cells (Novagen). Four-liter cell cultures were grown at 37 °C, and induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside at 30 °C for 4 h when the A500 nm reached ~0.7. The cells were then harvested at 8,000 r.p.m. (JA-10), resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 500 mM NaCl, 5 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride, 200 μg/ml lysozyme), and sonicated on ice. The lysate was centrifuged at 15,000 rpm for 30 min (JA-20 rotor), and the supernatant was dialyzed into buffer A (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl, 5 mM DTT, and 1 mM urea) and applied to a DEAE-Sepharose Fast Flow XK-50 column (Amersham Biosciences). Partially pure protein was eluted using a gradient of 1 mM NaCl (0–100%) in buffer A (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 2.5 mM DTT). Pooled protein fractions were then added to a SP-Sepharose Fast Flow XK-50 column (Amersham Biosciences) and eluted with a gradient of 1 mM NaCl (0–100%) in buffer B. The protein was then concentrated to ~1 ml and loaded onto a gel filtration column (Sephacryl S-100) equilibrated with buffer C (50 mM phosphate, pH 6.5, 200 mM NaCl, 2.0 mM DTT). Eluted protein was then concentrated using a Centricon concentrator (YM-3, Amicon, Inc.) for NMR studies. 15N-Labeled or 13N- and 15C-labeled proteins were

24894 JOURNAL OF BIOLOGICAL CHEMISTRY

VOLUME 282•NUMBER 34•AUGUST 24, 2007
produced by growing the cells in minimal media that contained 15NH4Cl or 13C0-labeled glucose, respectively. Protein NMR samples were ~1 mm in concentration and contained 50 mm phosphate buffer, pH 6.5, 200 mm NaCl, 5 mm dideuterated diothiothreitol, and 7% D2O.

Sample Preparation for Chemical Shift Perturbation Studies—The 43-mer RNA fragment (5'-GGG AGG UUG UUA AUG AAA AGG AUU UGA GAA UAA UAG C-3'), 32-mer RNA fragment (5'-GGG AUU CCA UGU UCA GAA GAA CGU GGU AUC UC-3'), and 9-mer RNA fragment (5'-GGG UAA UGA UUA-3') were prepared as previously described (31, 32). 1H-15N HSQC spectra were acquired using 300 GH11032GGAUAGUUA-3

Sample Preparation for Chemical Shift Perturbation Studies—The 43-mer RNA fragment (5'-GGG AGG UUG UUA AUG AAA AGG AUU UGA GAA UAA UAG C-3'), 32-mer RNA fragment (5'-GGG AUU CCA UGU UCA GAA GAA CGU GGU AUC UC-3'), and 9-mer RNA fragment (5'-GGG UAA UGA UUA-3') were prepared as previously described (31, 32). 1H-15N HSQC spectra were acquired using 300 μM RNA-protein complexes with molar ratios of 1:4, 1:2, 1:1, 2:1, and 4:1. Soluble complexes were obtained by mixing the components in the presence of 1 m NaCl, followed by dialysis into no salt buffer (50 mm phosphate pH 6.5, 2.5 mM DTT, 7% D2O). The composite of the amide proton (ΔδHN) and nitrogen (ΔδN) chemical shift change was calculated using the equation, Δδtotal = (ΔδHN ŴHN)² + (ΔδN ŴN)²/2, with ŴHN = 1 and ŴN = 0.154 (33).

NMR Spectroscopy—NMR experiments were performed at 300 K on Bruker DRX 600- and 500-MHz spectrometers equipped with a z-gradient triple resonance cryoprobe and a room temperature probe, respectively. Protein resonance assignments were accomplished using the following three-dimensional experiments: HNCA, HNCO, HNCOCA (34), HCCH COSY (38), HACACO (39), three-dimensional HN(CO)C (46), 13C correlation spectra (49). Distance restraints were obtained using NMRPipe (50) and analyzed using the NMRView and PIPP programs (51, 52).

Structure Calculations—Structures were calculated using the program X-PLOR-NIH (version 2.14) (53). The final values for the different experimental terms in the target function employed for simulated annealing were 200 kcal mol⁻¹ rad⁻² for the torsion angle restraints, 30 kcal mol⁻¹ Å⁻² for the experimental distance restraints (NOEs and hydrogen bonds), 1 kcal mol⁻¹ Hz⁻² for the 3JHNCO couplings, and 0.5 kcal mol⁻¹ ppm⁻² for the carbon chemical shifts. Hydrogen bond restraints were introduced in the final stages of refinement and were determined through an analysis of the NOE data. The structures were calculated using hybrid distance geometry-simulated annealing protocols, followed by an additional simulated annealing step on each of the initial distance geometry-

simulated annealing structures (54–56). The figures were prepared using the program MOLMOL and PYMOL (57, 58). A total of 200 structures were calculated, of which 110 had no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. 40 conformers with the lowest energies were chosen to represent the structure of ORF1CTD and have been deposited to the Protein Data Bank (accession code 2JR8).

Nitrocellulose Filter Binding Assay—The filter binding assay was performed as previously described (27). Briefly, ORF1p truncation constructs were incubated with a 76-nucleotide 32P-labeled RNA molecule for 30 min on ice in binding buffer (20 mM Hepes, pH 7.5, 250 mM NaCl, 1 mM dithiothreitol, 5 mM EDTA, and 100 μg/ml bovine serum albumin). The reactions were filtered through nitrocellulose and DE81 and washed with ice-cold binding buffer, and the remaining labeled RNA was quantified by phosphorimaging analysis using ImageQuant software (Amersham Biosciences). The apparent dissociation constant (KD) was calculated by plotting the total concentration of protein ([Ptotal]) as a function of the fraction of the RNA that was bound (y). Two models were used to fit the data. Noncooperative binding was assumed by fitting the data to the equation, y = ([Ptotal]/KD(app))/(1 + [Ptotal]/KD(app)), where [Pfree] is approximately equal to [Ptotal]. Data were also fit to a cooperative binding model using the equation, y = (ymax)/(1 + e^([Ptotal] - KD(app))/[Ptotal]/h)), where h is the apparent Hill coefficient.

RESULTS

Mouse ORF1p Contains Three Protease-resistant Regions—The structure of full-length ORF1p from the mouse Tt-type spa L1 element was investigated using limited proteolysis followed by MS. Full-length ORF1p containing a 39-amino acid N-terminal histidine tag was partially proteolyzed with trypsin, and the products were separated by gel electrophoresis (Fig. 1a). The identity of each fragment was then determined using MALDI-TOF MS and LC coupled to tandem MS/MS (Fig. 1b). Immediately after the addition of trypsin at 4°C, the peptide bond following residue Arg-12 is cleaved, leaving a resistant fragment that is missing the histidine tag and 12 N-terminal residues (fragment 2 in lanes 2–6). Further low temperature incubation with trypsin yields six fragments (fragments 3–8 in lanes 3–5). Fragments 3–5 share a common N terminus that starts at residue 45 but differ at their C termini (fragments 3–5 are formed by residues 45–279, 45–247, and 45–238, respectively). Extended digestion reveals that these fragments contain two resistant segments within them, residues 45–169, called the coiled-coil (C-C) domain, and residues 176–238, called the middle (M) domain. The C-C domain is the most stable portion of ORF1p, since it resists trypsin cleavage at higher temperatures and longer incubation times (fragment 6 in lane 7). Residues 63–130 within it are predicted by the program MULTI-COIL to form a coiled-coil (59). The boundaries of the middle domain (M domain) were defined indirectly, because it was only weakly visible on the SDS-gel after 5 h of incubation at 4°C and because its position was overlapped with fragment 8. The C-terminal boundary of the M domain can be deduced from the sequence of fragment 5, whereas its N terminus was deter-
A Highly Conserved C-terminal Domain within ORF1p

A highly conserved C-terminal domain within ORF1p is involved in RNA binding and is structurally disordered. The protease-sensitive region within C-1/3 is involved in RNA binding when referring to the Md-A protein. Two polypeptides from Md-A ORF1p were studied by NMR, one containing the entire C-1/3 region (ORF1C-1/3; residues 244–371) and the other containing only the CTD (ORF1CTD; residues 275–361). Initially, triple resonance NMR methods were used to assign the backbone resonances of isopropyl-13C- and 15N-labeled ORF1CTD, which exhibits excellent spectral dispersion characteristic of a globular protein (Fig. 3a). The boundaries of the CTD were then further refined by measuring the heteronuclear 1H–15N NOE values of its backbone amide nitrogen atoms. A plot of these data as a function of residue number is shown in Fig. 3b and reveals that the structured CTD consists of residues Phe289–Ile353.

The protease-sensitive region within C-1/3 is involved in RNA binding—We investigated whether the unstructured polypeptide segments that surround the CTD were involved in RNA binding using a nitrocellulose filter binding assay (62). This work made use of ORF1p from the mouse L1 Md-A element (used in all of the NMR studies) and a 76-nucleotide RNA fragment that has previously been shown to bind full-length ORF1p with high affinity (27). The RNA binding properties of three ORF1p variants were investigated: 1) full-length ORF1p, 2) ORF1C-1/3, and 3) ORF1CTD. Since the stoichiometry of binding to the 76-nucleotide RNA fragment is not known, only apparent dissociation constants were determined. The data were fit to two different binding models: one that assumed fewer residues in its C-C domain, such that its primary sequence is numbered differently. Hereafter, unless otherwise indicated, the amino acid numbering system used for spa-type ORF1p will also be used when referring to the Md-A protein.

Two polypeptides from Md-A ORF1p were studied by NMR, one containing the entire C-1/3 region (ORF1C-1/3; residues 244–371) and the other containing only the CTD (ORF1CTD; residues 275–361). Initially, triple resonance NMR methods were used to assign the backbone resonances of isopropyl-13C- and 15N-labeled ORF1CTD, which exhibits excellent spectral dispersion characteristic of a globular protein (Fig. 3a). The boundaries of the CTD were then further refined by measuring the heteronuclear 1H–15N NOE values of its backbone amide nitrogen atoms. A plot of these data as a function of residue number is shown in Fig. 3b and reveals that the structured CTD consists of residues Phe289–Ile353.
binding cooperativity and one that assumed noncooperative
binding. Both models yielded similar values for the apparent
dissociation constant \(K_{D(app)}\). Interestingly, the ORF1C-1/3
and ORF1CTD binding data are best fit if cooperativity is
assumed, and in each case a similar Hill coefficient \(h\) of \(-1.7\) is
obtained (Fig. 4). Consistent with previous studies, full-length
ORF1p binds RNA with a \(K_{D(app)}\) of \(6.1 \pm 0.4 \text{ nM}\) (data not
shown). In contrast, the ORF1C-1/3 polypeptide binds RNA
30-fold less tightly; fitting of the ORF1C-1/3 binding data to
cooperative and noncooperative models yields \(K_{D(app)}\) values of
\(181 \pm 8\) and \(180 \pm 3 \text{ nM}\), respectively. The reduction in affinity
is compatible with a simple avidity effect, since ORF1C-1/3 is

FIGURE 2. Sequence alignment of ORF1 proteins from different species. Sequence alignments were created using ClustalW (66). Identical amino acids are shaded black, and similar residues are shaded gray. The C domain (black rectangle), the M domain (white rectangle), and the secondary structure topology of the CTD are shown above the aligned sequences. Single asterisks indicate residues that are \(-30\%\) solvent-accessible in the NMR structure. Accession numbers are as follows: Mus musculus ORF1p (from retrotransposon L1 Spa) (AACS5341.1), M. musculus ORF1p (from retrotransposon L1 Md-A) (M13002), Homo sapiens (from retrotransposon L1.2) (AAC51278), Rattus norvegicus (S21345), Orzrias latipes (AF055640), Danio rerio (CAD61093), and Fundulus heteroclitus (AAS33199). The first three ORF1 sequences are from retrotransposons previously determined to be active in vivo (14, 15, 67, 68). ORF1C-1/3 contains residues 244–371, and ORF1CTD contains residues 275–361 (spa numbering). The partial proteolysis studies made use of the spa sequence, whereas the NMR structure work used Md-A ORF1p. These proteins share 83\% sequence identity over their entirety.
monomeric, whereas full-length ORF1p is trimeric (26). Interestingly, ORF1CTD binds RNA 20-fold less tightly than ORF1C-1/3, with noncooperative and cooperative $K_D^{(app)}$ values of 4,130 ± 580 and 3,500 ± 420 nM, respectively. This indicates that amino acids within the disordered tails that surround the CTD participate in RNA binding. Additional binding experiments using a protein construct that deleted only the unstructured N-terminal residues from ORF1C-1/3 revealed that it binds RNA with similar affinity as ORF1CTD (data not shown). This indicates that amino acids within the N-terminal unstructured tail collaborate with the CTD to mediate high affinity RNA binding.

**Solution Structure of the CTD (ORF1CTD)**—NMR was used to solve the structure of the CTD within ORF1p from the L1 Md-A element (ORF1CTD; residues 275–361). A total of 1,136 experimentally determined structural restraints (940 NOEs, 120 dihedral angles, 56 hydrogen bonds, and 203 $J_{HN}$ scalar couplings) were obtained using standard NMR methods applied to uniformly $^{15}$N-enriched and $^{15}$N- and $^{13}$C-enriched samples of ORF1CTD. These data were then used as input in simulated annealing calculations to generate an ensemble of 40 conformers that represent the structure. As shown in Fig. 5a, Phe$^{289}$–Ile$^{353}$ form a well ordered domain whose backbone and heavy atom coordinates can be superimposed to the mean structure with a root mean square deviation of 0.42 ± 0.09 and 0.93 ± 0.07 Å, respectively. All 40 of the calculated structures possess good covalent geometry and have no NOE, dihedral angle, or scalar coupling violations greater than 0.5 Å, 5°, or 2 Hz, respectively. A summary of the structural and restraint statistics is presented in Table 1.

The structure of ORF1CTD contains three $\alpha$ helices that are packed against a three-stranded $\beta$ sheet ($\alpha\beta\beta\alpha\alpha$ secondary
structure topology) (Fig. 5b). A DALI analysis reveals that this is a novel RNA binding fold, since it is related to only one protein structure with a Z-score greater than 2.0, the hypothetical Escherichia coli protein Ykff (Protein Data Bank code 2HJJ) that has no known function and is unrelated to ORF1p at the primary sequence level. The structure is initiated by residues Thr293–Leu307, which form a long amphipathic α-helix (helix α1) that leads into a three-stranded anti-parallel β-sheet that is constructed from contiguous residues (strand β1, residues Arg315–Leu317; strand β2, residues Lys321–Ile324; and strand β3, residues Lys331–Phe333). Two α helices positioned at an angle of ~110°, α2 (residues Lys336–Ser344) and α3 (residues Pro347–Ile352), complete the fold and are separated by a two-residue turn. All of the helices are packed onto the same side of the sheet, with the long axis of helix α3 positioned perpendicular to the strands of the sheet and the axes of helices α1 and α2 positioned nearly parallel to strands β1 and β2, respectively. Each of the secondary structural elements contributes residues to the hydrophobic core of the protein (Fig. 5c). Residues located in the hydrophobic core are the following: Ala296, Ala299, Trp300, Val303, and Leu307 from α1; Leu322 and Ile324 from the β1 strand; Ile326 from the loop between strands β1 and β2; Phe333 from β2; Phe339, Tyr342, and Leu343 from α2; and Leu349 from helix α3. As compared with the CTD from the spa element discovered by limited proteolysis (Fig. 1a), the primary sequence of Md-A ORF1CTD differs at only four positions (Fig. 2). None of these differences are expected to alter the structure of the CTD, since residues Leu312 (in the β1 and β2 loop), Ile325 (in loop β1 and β2 loop), Asp327 (in the β2 and β3 loop), and His341 (in helix α2) in Md-A ORF1CTD studied by NMR are solvent-exposed or conservatively replaced in spa ORF1CTD.

NMR Localization of a RNA Binding Surface on ORF1p—In order to localize the RNA binding surface on ORF1p—1/3, we performed a series of NMR titration experiments using 15N-labeled ORF1p and three different RNA oligonucleotides: 1) a 43-mer RNA molecule from the L1 genome that binds the full-length mouse ORF1p (27), 2) a single stranded 9-mer RNA molecule, and 3) a 32-mer RNA molecule that forms a hairpin and thus contains both double-stranded RNA and an exposed tetraloop (63). The choice of the RNA molecules is somewhat arbitrary, since ORF1p presumably binds RNA nonsequence specifically (27). The 43-mer was chosen because it had previously been shown to bind ORF1p (27). The 9-mer RNA molecule was chosen because ORF1p preferentially interacts with single-stranded oligonucleotides, and based on the structure of the CTD, it should be long enough to fully contact any potential binding surface on the protein.
Finally, the 32-mer hairpin was chosen because it contains both double- and single-stranded features, thereby increasing the chances of it favorably interacting with the protein. In separate experiments, each of the RNA molecules were added to the protein, and $^1$H-$^{15}$N HSQC spectra were recorded. The addition of the 9-mer single-stranded RNA molecule to a 200 μM solution of ORF1$^{C-1/3}$ dissolved in low salt buffer (50 mM PO$_4$, pH 6.5) caused negligible spectral changes. ORF1$^{C-1/3}$ binds this oligonucleotide very weakly, since no spectral changes occurred even when the 9-mer RNA is present at a 10-fold molar excess (data not shown). In contrast, the addition of the longer 32- and 43-mer RNA fragments to the protein dissolved in low salt buffer results in immediate precipitation, indicating that ORF1$^{C-1/3}$ exhibits some degree of length or sequence specificity in its binding.

In order to obtain samples suitable for NMR analysis, protein-RNA complexes containing the longer 32- and 43-mer RNA molecules were first formed in high salt buffer (50 mM phosphate, pH 6.5, 2.5 mM DTT, 1 M NaCl), since under these conditions, the components can be mixed together without immediately precipitating. Each complex was then dialyzed into low salt buffer and concentrated for NMR analysis. For each RNA molecule, five RNA complexes were studied that differed in the ratio of RNA to protein (RNA/protein 1:4, 1:2, 1:1, 2:1, and 4:1). Although the RNA-ORF1$^{C-1/3}$ complexes containing the 32- and 43-mer molecules were soluble, their $^1$H-$^{15}$N HSQC spectra exhibited extensive resonance line broadening at all RNA/protein ratios. Resonances arising from atoms in both the CTD and the disordered N-terminal linker disappeared completely, suggesting that both elements in ORF1$^{C-1/3}$ are in contact with RNA, and/or the complex forms a soluble higher order aggregate.

Interpretative chemical shift perturbation data were obtained when ORF1$^{C-1/3}$ and the 32-mer RNA molecule were studied. ORF1$^{C-1/3}$ removes the amino-terminal tail and in complex with the 32-mer yields $^1$H-$^{15}$N HSQC spectra that exhibit selective chemical shift perturbations and line broadening. Fig. 6a shows residues exhibiting composite chemical shift ($\Delta\delta_{\text{composite}}$) changes greater than 0.12 mapped onto the structure of ORF1$^{C-1/3}$. This analysis reveals a continuous surface that is composed of residues in helix $\alpha_1$, strand $\beta_1$, and residues in the turns connecting strands $\beta_1$ to $\beta_2$ and $\beta_3$ to helix $\alpha_2$. An inspection of the electrostatic surface of ORF1$^{C-1/3}$ reveals that these residues form a positively charged surface that is well suited for RNA binding (Fig. 6b). The distinct behavior of ORF1$^{C-1/3}$ and ORF1$^{C-1/3}$ upon titration of the 32-mer RNA is compatible with the stronger affinity of ORF1$^{C-1/3}$ for RNA measured by filter binding (Fig. 4) and further supports the notion that the N-terminal flexible tail collaborates with the CTD for optimal RNA binding.

**DISCUSSION**

L1 retrotransposons have successfully colonized a large fraction of the human and mouse genomes and require L1-encoded ORF1p to retrotranspose (2–5, 21). In this study, we investigated the structure and RNA binding properties of mouse ORF1p, which has been characterized in the greatest detail and shares significant sequence identity with human ORF1p. Previous sedimentation and AFM studies have revealed that it adopts a dumbbell-shaped homotrimeric structure in which...
A Highly Conserved C-terminal Domain within ORF1p

FIGURE 6. Solvent-exposed surface representation of the ORF1CTD protein. a, RNA-dependent chemical shift changes mapped onto the structure of ORF1CTD. A ribbon diagram of ORF1CTD and transparent surface is shown. Residues exhibiting significant chemical shift perturbations upon the addition of a 32-mer RNA molecule are indicated by yellow spheres. Residues shown have chemical shift perturbations greater than one S.D. from the mean change (0.12 < Δδ < 0.62) and are Glu292, Ala320, Arg298, Gln305, and Asp335. Also shown are residues with chemical shift perturbations that could not be determined either because they are significantly broadened or experience chemical shift perturbations in excess of 0.3 ppm (Phe289, Met294, Arg298, Gln305, and Arg308). b, the electrostatic surface potential diagram of ORF1CTD. Basic and acidic regions are colored in red, respectively. Hydrophobic regions are colored white. c, surface structure of ORF1CTD showing the positioning of residues required for in vivo retrotransposition mutants (red). ORF1p mutants harboring changes in Arg292 and Arg298 as well as residues Tyr218, Pro315, Ala520, Lys321, Ser323 are unable to promote retrotransposition in vivo (69). In all panels, two images are shown that differ by a rotation of 180° about the vertical axis.

due to the hydration of the protein molecule. The C-terminal third of ORF1p binds RNA and accelerates annealing of complementary oligonucleotides (61). The protease cleavage data indicate that this region is formed by the CTD and a ~40-residue protease-sensitive linker that connects the CTD to the M domain. Four lines of evidence indicate that both the linker and the CTD collaborate in RNA binding. First, we have shown that the isolated CTD binds RNA ~20-fold less tightly than a polypeptide containing the entire C-1/3 region (Fig. 4). Second, amino acid mutations within both of these elements disrupt retrotransposition and ribonucleoprotein particle formation in vivo (15, 16). Third, our NMR titration experiments indicate that the C-1/3 region binds RNA more tightly than the isolated CTD. Finally, residues within the linker and CTD are highly conserved, compatible with an important role in RNA binding (Fig. 2).

Since the linker is highly sensitive to digestion in the absence of RNA, ORF1p may adopt a "hydralike" structure in which the CTD RNA binding module in each chain of the trimer is connected to the M-domain via a flexible linker (Fig. 7a). This model is compatible with sedimentation studies of isolated ORF1C-1/3 that have shown that it does not form a hydrated compact sphere (64) and our NMR experiments that indicate that the polypeptides are arranged in parallel (26, 64). The handle of the dumbbell in the AFM image is 15.6 nm in length, whereas the lobes differ in size and have diameters of ~11 and ~4.5 nm (26). The domain structure deduced from our limited proteolysis study is compatible with this model and suggests which segments within ORF1p are likely to contribute to different parts of this structure (Fig. 7a). Mouse ORF1p contains at least three structured regions: 1) a segment that contains a predicted coiled-coil (C-C domain), 2) an acidic region in the middle of the primary sequence we refer to as the M domain, and 3) a CTD. Presumably, the handle of the dumbbell is formed by C-C domain, which is the most trypsin-resistant fragment in ORF1p. An analysis of its primary sequence using the program MULTICOIL predicts that residues 63–130 form a coiled-coil with a confidence level that exceeds 50% with a 21-residue prediction window. This is compatible with studies that have shown that this region contains determinants required for oligomerization (26, 28, 64). At the N- and C-terminal ends of the C-C fragment, 18 and 39 amino acids surround the predicted coiled-coil, respectively. Although these residues are predicted to form helices, there is a significantly lower probability that they form a coiled-coil. If all of the residues in the C-C protease-resistant fragment form a coiled-coil, its predicted length would be 18.8 nm (assuming a rise of 1.5 Å per residue (65)), which is slightly larger than the value measured from the AFM image. The small lobes probably contain residues within the protease-sensitive N-terminal tail (residues 1–45), whereas the large lobe in the AFM image is presumably formed by the M domain and the CTD. An inspection of the primary sequences of ORF1 proteins from vertebrate L1 elements reveals a high degree of sequence conservation in residues that encode the M domain, linker, and CTD (Fig. 2). In contrast, residues within the C-C domain show little sequence conservation and in human ORF1p have been predicted to form a leucine zipper (21).
A Highly Conserved C-terminal Domain within ORF1p

Figure 7. Model of trimeric ORF1 protein and a comparison of ORF1pCTD with the KH type-1 RNA-binding protein. a, model of trimeric ORF1p based on AFM imaging and the domain structure deduced from limited proteolysis. In each chain of the trimer, a CTD (blue sphere) is connected to the M domain (pink sphere) via a protease-sensitive linker (jagged blue line). Six amino acids then connect each M domain to a coiled-coil segment that contains a predicted coiled-coil (C-C, pink tube). The first 24 amino acids of ORF1p are sensitive to protease and are indicated by a black jagged line. b, comparison of the structures of ORF1pCTD (left) and the type-1 KH domain bound to single-stranded RNA (right) (69). The proteins adopt generally similar structures, although they are unrelated at the primary sequence level and their secondary structural elements are arranged differently. Each KH domain has a β1-α1-α2-β2-B3-a3 configuration as compared with ORF1pCTD with a α1-β1-β2-β3-a2-a3 arrangement. In the KH domain-RNA complex, a cleft formed by the sheet and the α-helix interacts with RNA (right). A similar cleft exists in the structure of ORF1pCTD, and the chemical shifts of residues within it are significantly perturbed when RNA is added (left, residues most affected by the addition of RNA are colored yellow). The protein and the RNA in the KH domain-RNA complex are colored pink and gray, respectively (Protein Data Bank code 1EC6).

that the linker is unstructured in ORF1C-1/3. However, it is also conceivable that in the context of the full-length trimeric protein, the linker is structured as a result of inter-protomer interactions. Cleavage of one or more exposed sites within the structured linker could destabilize it sufficiently to lead to its complete and rapid digestion under the experimental conditions we have used.

Interestingly, ORF1C-1/3 and ORF1pCTD bind RNA with a modest degree of cooperativity, as evidenced by Hill coefficients of ~1.7 (Fig. 4). This may be a result of protein-protein interactions that facilitate binding and could represent residual interactions between the CTDs that are also present in the intact homotrimer. It is also possible that binding of ORF1p promotes additional protein binding by altering the structure of RNA, which would be compatible with the demonstrated nuclear chaperone activity of ORF1p (15). In the future, it will be important to further explore this issue by determining the stoichiometry and nucleotide sequence and length dependence of protein binding. NMR and binding studies that progressively shorten the 76-mer RNA molecule that we have shown binds ORF1C-1/3 with high affinity may also lead to the identification of RNA subdomains whose protein complexes are better suited for structural analysis.

NMR studies indicate that the CTD is a structurally unique binding module that may interact with single-stranded oligonucleotides like type-1 KH domains. Although they are unrelated at the primary sequence level and have a distinct arrangement of secondary structural elements, the CTD and type-1 KH domains adopt superficially similar structures in which three α helices are packed against a three-stranded antiparallel β sheet (Fig. 7b). Interestingly, both proteins may preferentially bind single-stranded RNA using a structurally similar cleft. In the structure of the protein-RNA complex containing the Nova type-1 KH domain, a single-stranded RNA molecule binds to a cleft formed by two tandem arranged α helices (α1 and α2) and β strand β2 (Fig. 7b, right). The chemical shift mapping data of ORF1p are compatible with a similar general mode of binding in which the RNA contacts the cleft formed by residues in helix α1, strand β2, and the β1–β2 loops. Attractively, the location of this binding surface is consistent with the demonstrated importance of the linker, which immediately precedes helix α1 in the primary sequence and is therefore well positioned to make additional contacts to RNA. However, the chemical shift mapping data should be interpreted with caution, since only the addition of the 32-mer hairpin molecule caused interpretable changes in the NMR spectrum. These changes are minor and presumably result from weak interactions with nucleotides in the hairpin portion of the RNA. In contrast, the 43-mer appears to bind the protein with high affinity and exhibits uniform line broadening, suggesting that multiple proteins bind, and/or a well defined complex is not formed. It is unclear why minimal spectral changes are observed when the 9-mer is added to ORF1C-1/3, but it is possible that it is not long enough to effectively bind the protein.

The structure of the CTD and its postulated RNA binding mode are compatible with mutagenesis studies of the mouse and human ORF1 proteins. This work has shown that Arg297 and Arg298 (spa numbering) play an important role in retrotransposition (14–16). Nonconservative substitution of the diarginine peptide with Ala-Ala in the mouse or human proteins impairs RNA binding and ribonucleoprotein particle formation as well as in vivo retrotransposition. These residues also play an important role in the nucleic acid chaperone activity of ORF1p, since mutants containing more conservative alterations (Lys-Arg, or Lys-Lys) retain RNA binding affinity but, unlike the wild-type protein, have lost their nucleic acid chaperone activity and are unable to support retrotransposition (14–16). In the structure, the side chains of Arg297 and Arg298 are located at the end of helix α1 and are well positioned to interact with nucleic acids, since they are proximal to the interaction surface identified by NMR (Fig. 6b). The structure of CTD also sheds light onto the effects of more radical mutations in the CTD. Mutants of human ORF1p that replace residues Tyr318-Pro319-Ala320-Lys321-Leu322-Ser323 (L1 spa numbering) with Ala-Ala-Ala-Leu-Ala are inactive in vivo and incapable of forming ribonucleoprotein complexes. These residues are completely conserved in mouse ORF1p and are located in strand β2 and the loop that connects strands β1 and β2. It seems likely that replacement of Pro319 with alanine will disrupt...
the structure of the CTD by destabilizing the turn that connects these β strands. In the closely related human protein, a four-amino acid substitution in the linker preceding the CTD impairs retrotransposition and ribonucleoprotein particle formation (Arg271-Glu272-Lys273-Gly274 to Ala-Ala-Ala-Ala) (14). Based on our results, it seems likely that at least part of the deleterious effect of this mutation results from a reduction in RNA binding affinity.

In conclusion, we have shown that the most highly conserved region in vertebrate ORF1 proteins consists of two domains that are connected by a protease-sensitive linker. Both the linker and the CTD are required for high affinity binding to RNA. The CTD adopts a unique three-dimensional structure that, based on chemical shift mapping experiments, interacts with RNA via a cleft formed by the N-terminal α-helix and the three-stranded β sheet. Future studies will need to determine if the M domain contributes to binding, whether the linker is structurally disordered in the full-length protein, and how these motifs work in concert in the full-length trimer protein to interact with and chaperone nucleic acids.

Acknowledgments—We thank members of the Clubb and Martin laboratories for useful discussions and support.

REFERENCES

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczyk, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Graham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Showkneen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Wendl, M. L., Wendl, M. D., Deininger, P. L., Moran, J. V., Batzer, M. A., and Kazazian, H. H. (2001) Nature Genet. 27, 11–20.

2. Brouha, B., Schustak, J., Badger, M. D. L., and Moor, J. C. (2004) National Acad. Sci. U. S. A. 100, 5280–5285.

3. Chen, J. M., Stenson, P. D., Cooper, D. N., and Ferec, C. (2005) Hum. Genet. 117, 411–427.

4. Deininger, E. M., and Kazazian, H. H. (2001) Mol. Cell 7, 35–42.

5. Deininger, E. M., and Kazazian, H. H. J. (2003) Annu. Rev. Genet. 37, 528–558.

6. Martin, S. L. (1991) J. Mol. Biol. 221, 429–475.

7. Martin, S. L., Branciforte, D., Keller, D., and Bain, D. L. (2003) Cell 112, 315–325.

8. Martin, S. L., and Branciforte, D. (1993) J. Biol. Chem. 268, 2674–2679.

9. Martin, S. L. (2006) Nat. Struct. Mol. Biol. 13, 1155–1165.

10. Martin, S. L., and Kazazian, H. H. J. (2003) Annu. Rev. Genet. 37, 528–558.

11. Martin, S. L., and Kazazian, H. H., Jr. (2003) Curr. Opin. Genet. Dev. 13, 651–658.

12. Martin, S. L., and Kazazian, H. H., Jr. (2003) Nat. Genet. 35, 41–48.

13. Nishikura, K., Takada, N., Kobayashi, M., and Maekawa, H. (2004) Cell 120, 77–87.

14. Oh, J., Yu, D., Madura, M. J., and Boeke, J. D. (2004) J. Virol. 78, 13815–13820.

15. Okumura, K., Sato, Y., Miyazawa, Y., Shizuya, H., Choi, S., and Chen, Y. J. (2001) Nature 409, 860–921.

16. Ostertag, E. M., and Kazazian, H. H. (2001) J. Virol. 75, 1555–1565.

17. Ostertag, E. M., and Kazazian, H. H. (2001) Annu. Rev. Genet. 35, 521–558.

18. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

19. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

20. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

21. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

22. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

23. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

24. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

25. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

26. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

27. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

28. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

29. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

30. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

31. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

32. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

33. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

34. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.

35. Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D. T., and Williams, C. S. (2000) Genomics 65, 241–252.
A Highly Conserved C-terminal Domain within ORF1p

36. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc.* 114, 6291–6293
37. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Magn. Reson.* 88, 425–431
38. Ikura, I., Kay, L. E., and Bax, A. (1991) *J. Biomol. NMR* 1, 299–304
39. Zhang, W., and Gmeiner, W. H. (1996) *J. Biomol. NMR* 7, 247–250
40. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993) *J. Magn. Reson. B* 101, 333–337
41. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry (Mosc.)* 28, 6150–6156
42. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc.* 115, 7772–7777
43. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson.* 95, 636–641
44. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* 13, 289–302
45. Clore, G. M., Bax, A., and Gronenborn, G. M. (1991) *J. Biomol. NMR* 1, 13–22
46. Hu, J. S., and Bax, A. (1997) *J. Am. Chem. Soc.* 119, 6360–6368
47. Hu, J. S., Grzesiek, S., and Bax, A. (1997) *J. Am. Chem. Soc.* 119, 1803–1804
48. Grzesiek, S., Vuister, G. W., and Bax, A. (1993) *J. Biomol. NMR* 3, 487–493
49. Bax, A., Max, D., and Zax, D. (1992) *J. Am. Chem. Soc.* 114, 6923–6925
50. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* 6, 277–293
51. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J. Magn. Reson.* 95, 214–220
52. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* 4, 603–614
53. Schwiers, C. D., Kuszewski, J., I., Tjandra, N., and Clore, G. M. (2003) *J. Magn. Reson.* 160, 65–73
54. Nilges, M., Clore, G., and Gronenborn, A. (1988) *FEBS Lett.* 229, 129–136
55. Nilges, M., Kuszewski, J., and Brunger, A. T. (1991) *Computational Aspects of the Study of Biological Macromolecules by NMR*, Plenum Press, New York
56. Kuszewski, J., Nilges, M., and Brunger, A. T. (1992) *J. Biomol. NMR* 2, 33–56
57. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph.* 14, 51–55
58. Delano, W. L. (2006) *The PyMOL Molecular Graphics System*, Version 0.99, DeLano Scientific, South San Francisco, CA
59. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci.* 6, 1179–1189
60. Garnier, J., Gibrat, J. F., and Robson, B. (1996) *Methods Enzymol.* 266, 540–553
61. DeBerardinis, R. J., Goodier, J. L., Ostertag, E. M., and Kazazian, H. H., Jr. (1998) *Nat. Genet.* 20, 288–290
62. Wong, L., and Lohman, T. M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 5428–5432
63. Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004) *Proc. Natl. Acad. Sci. U. S. A.* 101, 8307–8312
64. Basane, S., WAi-lun Li, P., Howard, G., Branciforte, D., Keller, D., and Martin, S. L. (2006) *J. Mol. Biol.* 357, 351–357
65. Hakansson-McReynolds, S., Jiang, S., Rong, L., and Caffrey, M. (2006) *J. Biol. Chem.* 281, 11965–11971
66. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* 22, 4673–4680
67. Goodier, J. L., Ostertag, E. M., Du, K., and Kazazian, H. H., Jr. (2001) *Genome Res.* 11, 1677–1685
68. Naas, T. P., DeBerardinis, R. J., Moran, J. V., Ostertag, E. M., Kingsmore, S. F., Seldin, M. F., Hayashizaki, Y., Martin, S. L., and Kazazian, H. H. (1998) *EMBO J.* 17, 590–597
69. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) *Cell* 100, 323–332
70. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* 8, 477–486