Article No. jmbi.1999.2818 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 289, 949–962

JMB

High Precision Solution Structure of the C-terminal KH Domain of Heterogeneous Nuclear Ribonucleoprotein K, a c-*myc* Transcription Factor

James L. Baber[1], Daniel Libutti[2], David Levens[2] and Nico Tjandra[1]*

[1]*Laboratory of Biophysical Chemistry, Building 3 National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda MD 20892-0380, USA*

[2]*Laboratory of Pathology National Cancer Institute National Institutes of Health Bethesda, MD 20892, USA*

**Corresponding author*

Among it's many reported functions, heterogeneous nuclear ribonucleoprotein (hnRNP) K is a transcription factor for the c-*myc* gene, a proto-oncogene critical for the regulation of cell growth and differentiation. We have determined the solution structure of the Gly26 → Arg mutant of the C-terminal K-homology (KH) domain of hnRNP K by NMR spectroscopy. This is the first structure investigation of hnRNP K. Backbone residual dipolar couplings, which provide information that is fundamentally different from the standard NOE-derived distance restraints, were employed to improve structure quality. An independent assessment of structure quality was achieved by comparing the backbone ^{15}N T_1/T_2 ratios to the calculated structures. The C-terminal KH module of hnRNP K (KH3) is revealed to be a three-stranded β-sheet stacked against three α-helices, two of which are nearly parallel to the strands of the β-sheet. The Gly26 → Arg mutation abolishes single-stranded DNA binding without altering the overall fold of the protein. This provides a clue to possible nucleotide binding sites of KH3. It appears unlikely that the solvent-exposed side of the β-sheet will be the site of protein-nucleic acid complex formation. This is in contrast to the earlier theme for protein-RNA complexes incorporating proteins structurally similar to KH3. We propose that the surface of KH3 that interacts with nucleic acid is comparable to the region of DNA interaction for the double-stranded DNA-binding domain of bovine papillomavirus-1 E2 that has a three-dimensional fold similar to that of KH3.

Keywords: hnRNP K; KH domain; dipolar coupling; nucleic acid-binding; c-*myc*

Introduction

Heterogeneous nuclear ribonucleoprotein (hnRNP) K was first identified as one of at least 20 major proteins that are part of hnRNP particles in mammalian cells and has been presumed to be involved in the transport and/or processing of heterogeneous nuclear and mature RNA. Although hnRNP K binds poly(rC), a synthetic RNA homopolymer, *in vitro, bona fide in vivo* RNA targets have yet to be ascertained (Matunis *et al*., 1992; Swanson & Dreyfuss, 1988). Subsequent investigations implicated hnRNP K was a positive-acting transcription factor for the human c-*myc* gene (Michelotti *et al*., 1996; Tomonaga & Levens, 1996). The c-*myc* gene encodes a DNA-binding protein important in cell growth and differentiation. Deregulation of this gene can lead to tumorgenesis. hnRNP K binds specifically with a single-stranded

Abbreviations used: hnRNP, heterogeneous nuclear ribonucleoprotein; KH, K homology; KH3, C-terminal KH domain of hnRNP K; ss, single-stranded; ds, double-stranded; CT repeat, CCCTCCCCA oligonucleotide; GST, glutathione-*S*-transferase; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RMSD, root-mean-square-deviation; RMS, root-mean-square; RNP, ribonucleoprotein; dsRBD, double-stranded RNA-binding domain; EMSA, electrophoretic mobility shift assay; IPAP, in-phase-anti-phase; HSQC, heteronuclear single quantum correlation; NOESY, nuclear Overhauser enhancement spectroscopy; CT, constant time; ppm, parts per million; TOCSY, total correlation spectroscopy.

E-mail address of the corresponding author: nico@helix.nih.gov

(ss) *cis*-element of c-*myc* comprised of five imperfect CT repeats (CCCTCCCCA) and located 100-150 bp upstream of the P1 promoter (Takimoto *et al*., 1993; Tomonaga & Levens, 1995).

Additional studies reviewed by Bomsztyk *et al*. (1997) suggest that hnRNP K interacts with a variety of proteins including tyrosine (Van Seuingen *et al*., 1995b) and serine/threonine kinases (Ostrowski *et al*., 1991; Van Seuningen *et al*., 1995a,b), the proto-oncoprotein Vav (Bustelo *et al*., 1995), zinc-finger transcriptional repressors (Bomsztyk *et al*., 1997; Denisenko *et al*., 1996), and TFIID TATA-binding protein (TBP; Michelotti *et al*., 1996). Protein-hnRNP K interactions often involve a partner's SH3 domain and one of the three proline-rich segments of hnRNP K (Bomsztyk *et al*., 1997). Co-immunoprecipitation of hnRNP K with both a serine/threonine kinase and c-Src or Vav led Bomszytk *et al*. to propose that hnRNP K might allow the formation of *in vivo* multi-enzyme complexes. Furthermore, they have suggested that hnRNP K might simultaneously engage both protein(s) and DNA or RNA bridging nucleic acids with proteins lacking nucleic acid binding capabilities (Bomsztyk *et al*., 1997). More recently, hnRNP K was shown to interact with the hepatitis C virus core protein (Hsieh *et al*., 1998) and with the C/EBPβ transcription factor which regulates a wide variety of genes (Miau *et al*., 1998). The binding and activities of Sp1 and Sp3 transcription factors with the promoter of a neuronal nicotinic acetylcholine receptor were shown to be affected by the presence of hnRNP K (Du *et al*., 1998). hnRNP K also mediates the translation of erythroid 15-lipoxygenase (LOX) mRNA (Ostareck *et al*., 1997) and L2 mRNA which codes for the L2 capsid protein of human papillomavirus type 16 (HPV-16) (Collier *et al*., 1998).

hnRNP K contains three K homology (KH) modules (Siomi *et al*., 1993a). These domains are also found in the following nucleic acid-binding proteins, to name a few: yeast MER1, Sam68, FMR1, and the c-*myc* FUSE binding protein (Ashley *et al*., 1993; Duncan *et al*., 1994; Gibson *et al*., 1993a,b; Siomi *et al*., 1993a,b, 1994; Wong *et al*., 1992). Several studies show that KH domains are directly responsible for the nucleic acid binding affinity of this class of proteins (Dejgaard & Leffers, 1996; Leffers *et al*., 1995; Siomi *et al*., 1994; Tomonaga & Levens, 1995; Urlaub *et al*., 1995). Two non-overlapping fragments of hnRNP K, one spanning the two N-terminal KH domains (KH1 and KH2) and the other encompassing the C-terminal KH domain (KH3), each bind specifically with an oligonucleotide comprised of three CT repeats *in vitro*. These same KH constructs as well as full-length hnRNP K bind specifically with the RNA analog of three CT repeats, albeit, to a lesser extent (Tomonaga & Levens, 1995). FMR1 and Sam68 also bind both RNA and DNA, though specific *in vivo* targets have not been demonstrated (Ashley *et al*., 1993). Do KH domains generally prefer DNA targets or RNA targets or do most KH domains bind specifically with both DNA and RNA and, if so, why? Does the KH motif just provide a scaffold that can be evolutionarily adjusted to recognize a specific DNA or RNA target? To help answer these questions, we have determined the solution structure of hnRNP Ks C-terminal KH domain (KH3) by nuclear magnetic resonance (NMR). This structural information should prove essential in unraveling the molecular mechanisms of the multifunctional hnRNP K.

Results

Structure determination

The fragment presented spans the last 85 C-terminal residues of hnRNP K (residues 379-463) and encompasses the third KH domain. A glutathione-*S*-transferase (GST) affinity column was used for protein purificatioin, resulting in the addition of four linker residues at the N terminus of the peptide. These extra residues are unstructured and the ^{15}N relaxation data (data not shown) show that they are highly flexible.

$^1D_{N\text{-}H}$, $^1D_{C\alpha\text{-}H\alpha}$, $^1D_{C\alpha\text{-}C'}$, and $^1D_{N\text{-}C}$, dipolar coupling restraints were utilized in addition to the distance and ϕ angle restraints derived from standard heteronuclear NMR experiments (see Table 1 and Materials and Methods). The dipolar coupling between two nuclei yields the orientation of the internuclear bond vector relative to the overall molecular alignment tensor. Residual dipolar couplings can be measured in high magnetic field for proteins with non-zero anisotropic magnetic susceptibility (Kung *et al*., 1995; Tjandra & Bax, 1997b; Tjandra *et al*., 1996; Tolman *et al*., 1995; Tolman & Prestegard, 1996). Dipolar coupling is reintroduced by a slight alignment of the molecule in the high field. However, it was recently shown that dissolving a macromolecule in a dilute liquid crystalline phase produces a considerably larger and adjustable degree of alignment (Tjandra & Bax, 1997a). The liquid crystal used is a mixture of phospholipids which form disc-shaped particles called bicelles that align with magnetic field (Sanders & Schwonek, 1992). The larger degree of alignment increases the sensitivity of the measurable dipolar couplings. In the case of directly bonded nuclei, dipolar couplings are distance-independent and contain long range order. Accurate orientations of secondary structural elements are difficult to discern when the elements are connected by only a few NOEs. Drohat *et al*. (1999) recently showed that the relative orientation of two structural modules in apo-S100B(ββ) can be accurately determined only after the inclusion of dipolar coupling information. This allowed the recognition of functionally important structural changes in this protein. Furthermore, the inclusion of dipolar coupling restraints accelerates the process of identifying regions of a molecule where structural information may be inconsistent.

The dipolar coupling between two nuclei is given by (Tjandra & Bax, 1997a):

Table 1. Structural statistics

	⟨Sa⟩	Lowest energy
RMSDs from exptl. distance restraints (Å)		
All distance restraints (1302)	0.034 ± 0.003	0.029
Intraresidue (448)[a]	0.038 ± 0.005	0.033
Sequential ($\|i-j\|=1$) (322)	0.036 ± 0.004	0.031
Medium-range ($\|i-j\| \leqslant 5$) (198)	0.029 ± 0.005	0.026
Long-range ($\|i-j\|>5$) (279)	0.027 ± 0.002	0.026
H-bond (55)[b]	0.025 ± 0.004	0.027
RMSDs from dipolar coupling restraints (Hz)[c]		
$^1D_{NH}$ (42)	0.48 ± 0.05	0.44
$^1D_{NH}$ cat. 2 (7)	0.36 ± 0.04	0.46
$^1D_{CH}$ (49)	0.83 ± 0.04	0.82
$^1D_{CH}$ cat. 2 (10)	0.16 ± 0.04	0.16
$^1D_{C'N}$ (46)	0.40 ± 0.02	0.38
$^1D_{C'N}$ cat. 2 (2)	0.01 ± 0.03	0
$^1D_{C\alpha C'}$ (49)	0.68 ± 0.02	0.65
$^1D_{C\alpha C'}$ cat. 2 (4)	0.07 ± 0.08	0.08
RMSDs from exptl. φ-angle restraints (°) (56)[d]	0.28 ± 0.12	0.17
Deviations from idealized covalent geometry		
Bonds (Å) (1384)	0.0032 ± 0.0001	0.0031
Angles (°) (2508)	0.533 ± 0.004	0.529
Impropers (°) (692)	0.42 ± 0.02	0.418
Measures of structure quality		
Lennard-Jones energy (kcal mol^{-1})[e]	−311 ± 7	−307
PROCHECK[f]		
Residues in most favorable region of Ramachandran plot (%)	95.2 ± 1.1	95.0
Residues in allowed regions of Ramachandran plot (%)	4.8 ± 1.1	5.0
No. of bad contacts/100 residues	4.6 ± 2.1	5.7
RMSDs from mean structure (Å)[g]		
Backbone (N, C^α, C′, O)	0.16 ± 0.03	
All non-hydrogen	0.80 ± 0.06	

⟨SA⟩ represents the 20 lowest overall energy structures of 100 calculated. For ⟨SA⟩, the values shown are mean ± standard deviation with the number of restraints used to calculate these values shown in parentheses. The final values of the force constants used in the simulated annealing calculations are as follows: 1000 kcal mol^{-1} Å^{-2} for bond lengths, 500 kcal mol^{-1} rad^{-2} for angles and improper torsions, 4 kcal mol^{-1} Å^{-4} for the quartic vdW repulsion term (vdW radii are 0.8 times the CHARMM PARAM 19/20 values), 30 kcal mol^{-1} Å^{-2} for NOE-derived distance restraints, 55 kcal mol^{-1} Å^{-2} for generic hydrogen-bond restraints, 10 kcal mol^{-1} rad^{-2} for experimental φ angle restraints, 0.55 kcal mol^{-1} Hz^{-2} for $^1D_{NH}$, 0.36 kcal mol^{-1} Hz^{-2} for $^1D_{C\alpha H\alpha}$, 5.0 kcal mol^{-1} Hz^{-2} for $^1D_{C'N}$, and 2.2 kcal mol^{-1} Hz^{-2} for $^1D_{C\alpha C'}$.

[a] Only NOE-derived distance restraints between protons separated by more than three bonds are included in this category.

[b] For α-helices, two restraints were used per hydrogen-bond (e.g. $r_{NH\text{-}O} = 1.5\text{-}2.5$ Å and $r_{N\text{-}O} = 2.4\text{-}3.6$ Å).

[c] Dipolar coupling RMSDs denoted as cat. 2 are values calculated for residues in flexible regions of the molecule as discussed in the text (i.e. residues 1-12, 53-57, and 82-89).

[d] Backbone φ angles restraints were set to: 55(±35)° for $^3J_{HNH\alpha} < 5.5$ Hz, 120(±60)° for 7 Hz$<^3J_{HNH\alpha} < 8$ Hz, 120(±50)° for 8 Hz$\leqslant ^3J_{HNH\alpha} < 9$ Hz, and 120(±40)° for $^3J_{HNH\alpha} \geqslant 9$ Hz.

[e] The Lennard-Jones van der Waals energy was not incorporated into the simulated annealing calculation. CHARM PARAM 19/20 parameters were used to calculate these values.

[f] PROCHECK (Laskowski *et al*., 1993) and PROCHECK_NMR (Laskowski *et al*., 1996) were used to calculate these values. These values were calculated for residues 13-81 excluding the unstructured part of both the N as well as the C termini. There were no φ/ψ values in the disallowed or generously allowed regions for non-proline/non-glycine residues. The following PROCHECK *G*-factors were calculated: −0.03(±0.04) for φ-ψ, −0.05(±0.12) for χ_1-χ_2, −0.39(±0.12) for χ_1 only, and +0.60(±0.12) for χ_3-χ_4. The 20 lowest energy structures of 100 calculated without dipolar coupling restraints had 86.9(±2.7)% of their residues in the most-favored Ramachandran region and 4.9(±1.9) bad contacts/100 residues.

[g] Coordinates of residues 1-12, 53-57, and 82-89 were excluded from the calculation of these values. Structures were least-squares fit to the average structure by XPLOR3.1 (Brunger, 1993). The 20 lowest energy structures of 100 calculated without dipolar couplings produced structural RMSDs of 0.39(±0.05) Å and 0.92 (±0.06) Å for fits of the backbone and all non-hydrogen atoms, respectively.

$$D_{PQ}(\theta,\phi) = -S(\mu_0/4\pi)\gamma_P\gamma_Q h\{A_a(3\cos 2\theta - 1) + 3/2A_r \sin^2\theta\cos 2\phi\}/4\pi^2 r_{PQ}^3 \quad (1a)$$

or simplified (Clore *et al*., 1998a):

$$D(\theta,\phi) = D_a(3\cos^2\theta - 1) + 3/2D_r \sin^2\theta\cos 2\phi \quad (1b)$$

where S is the generalized order parameter for internal motion of vector **PQ**, r_{PQ} is the distance between nuclei P and Q, h is Planck's constant, γ_j is the gyromagnetic ratio for atom j, μ_0 is the magnetic permeability in a vacuum, θ and φ are the cylindrical coordinates of vector **PQ** relative to the molecular alignment tensor with principal axis system **A**, and A_a and A_r are the axial and rhombic components, respectively, of **A**. In the simplified format, all of the physical constants have been

absorbed into D_a and D_r, which now represent the magnitude of the alignment tensor in hertz.

XPLOR 3.1 (Brunger, 1993) was modified to incorporate pseudopotentials for dipolar coupling restraints (Clore *et al.*, 1998b). Dipolar couplings for residues 13 to 52 and 58 to 81 were placed in a quadratic harmonic potential (i.e. $E_{dip} = k(D_{PQ}^{calc} - D_{PQ}^{obs})^2$). Assuming that a dipolar coupling scales with respect to the order parameter S (see equation (1a)), dipolar couplings measured in flexible areas of the molecule will have lower magnitudes than expected from the orientation of the bond vector alone. Consequently, a quadratic half-open square well potential penalty function was used for residues with lower order parameters as indicated by the ^{15}N relaxation data (e.g. residues 1-12, 53-57, and 82-89 for KH3). This potential is defined as (Ottiger *et al.*, 1998b):

$$\begin{aligned} &\text{for} \quad D_{PQ}^{obs} > 0: \\ E_{dip} &= k(D_{PQ}^{calc} - D_{PQ}^{obs})^2 \quad \text{if} \quad D_{PQ}^{calc} < D_{PQ}^{obs} \\ E_{dip} &= 0 \qquad\qquad\qquad\quad \text{if} \quad D_{PQ}^{calc} > D_{PQ}^{obs} \end{aligned} \tag{2}$$

$$\begin{aligned} &\text{for} \quad D_{PQ}^{obs} < 0: \\ E_{dip} &= k(D_{PQ}^{calc} - D_{PQ}^{obs})^2 \quad \text{if} \quad D_{PQ}^{calc} > D_{PQ}^{obs} \\ E_{dip} &= 0 \qquad\qquad\qquad\quad \text{if} \quad D_{PQ}^{calc} < D_{PQ}^{obs} \end{aligned}$$

Dipolar couplings between −2 Hz and +2 Hz were excluded from the half-open square well restraint tables as these values conservatively represent the experimental uncertainty and one cannot reliably determine the true sign of the dipolar coupling. For glycine residues, only one value for the $C^{\alpha}H^{\alpha}$ dipolar couplings is observed. It is the sum of the two individual intraresidue $C^{\alpha}H^{\alpha}$ dipolar couplings. To account for this, the following penalty function was used in the case of glycine $C^{\alpha}H^{\alpha}$ dipolar coupling:

$$D_{dip} = k[(D_{C\alpha H\alpha 1} + D_{C\alpha H\alpha 2})^{calc} - (D_{C\alpha H\alpha 1} + D_{C\alpha H\alpha 2})^{obs}]^2 \tag{3}$$

where, again, $(D_{C\alpha H\alpha 1} + D_{C\alpha H\alpha 2})^{obs}$ is observed as a single value.

The force constants for the dipolar coupling penalty functions were incrementally increased during the cooling process of the simulated annealing calculation which is analogous to the treatment of generic hydrogen-bond and NOE-derived distance restraints. At the initial structure refinement stage, the force constants were chosen so that the final RMSDs between the observed and calculated dipolar couplings were approximately equal to the uncertainty in the dipolar coupling measurement.

The magnitude of the alignment tensor (i.e. D_a and D_r of equation (1b)) must be properly estimated in order to refine a NMR structure using dipolar coupling information. The distribution method proposed by Clore *et al*. (1998a) proved the least sensitive towards biases in the distribution of the dipolar vector orientations. One would expect a highly non-uniform distribution of the dipolar vectors in the case of KH3 due to the relatively small number of residues and the high content of regular secondary structure. The alignment tensor for KH3 was estimated by linearly normalizing $^1D_{C\alpha\text{-}H\alpha}$, $^1D_{C\alpha\text{-}C'}$ and $^1D_{N\text{-}C'}$ to those of $^1D_{N\text{-}H}$ by the relevant gyromagnetic ratio and bond distance dependent factors. All four sets of normalized dipolar couplings were compiled into a final histogram (Figure 1), which allowed immediate determination of the axial and rhombic components of the molecular alignment tensor (Clore *et al.*, 1998a). For our present case, the axial component (D_a^{NH}) is −19.6 Hz and the rhombicity ($R^{NH} = D_r^{NH}/D_a^{NH}$) is

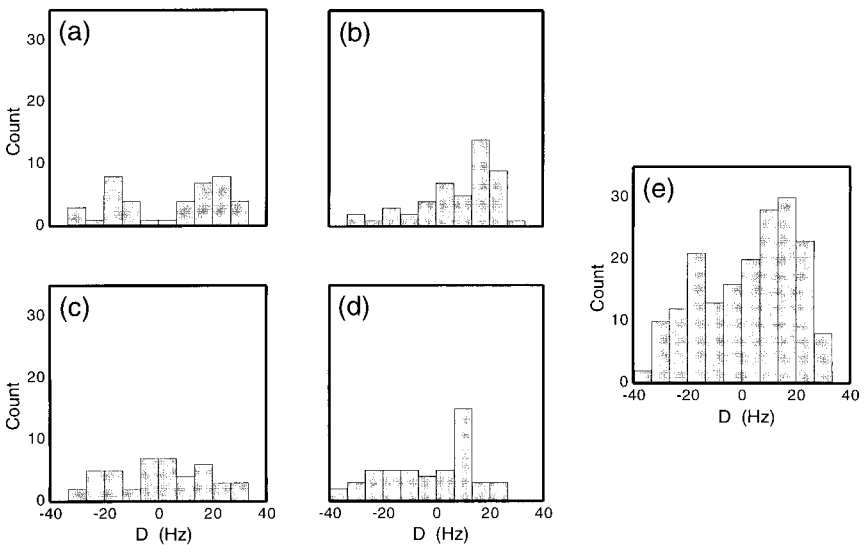

Figure 1. Histograms of the dipolar couplings. It is apparent that the individual distributions shown in the histograms for N-H (a), C^{α}-H^{α} (b), C′-N (c), and C^{α}-C′ (d) do not reflect a powder pattern distribution as a result of the high content of regular secondary structure. The collective distribution obtained by normalizing the C^{α}-H^{α}, C′-N, and C^{α}-C′ dipolar couplings to those of N-H by the appropriate factor containing their bond length and gyromagnetic ratio (i.e. $\gamma_N\gamma_H\langle r_{NH}^{-3}\rangle / \gamma_A\gamma_B\langle r_{AB}^{-3}\rangle$ where r_{AB} is the bond distance between A and B and γ_i is the gyromagnetic ratio of i, provides a better representation of a powder pattern distribution shown in (e). The factors used for each set of dipolar couplings were 0.46, 8.3, and 5.0 for C^{α}-H^{α}, C′-N, and C^{α}-C′, respectively.

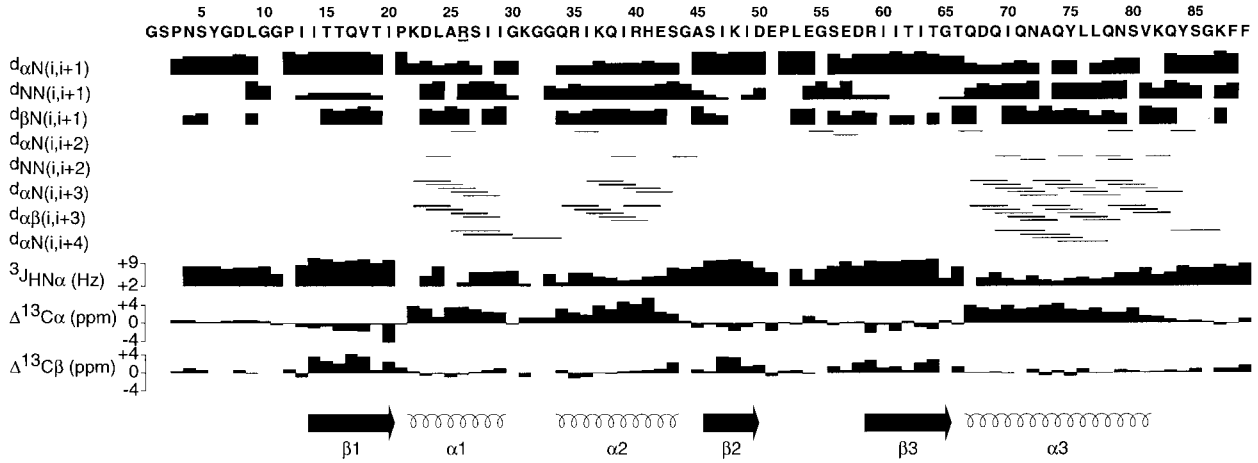

Figure 2. Diagram of data used to establish the secondary structure of KH3. The first four residues of the sequence are non-native. They are part of the glutathione-*S*-transferase linker. Arg26 (underlined) is a glycine in the wild-type protein.

0.3. Further refinement of these values by a grid search method (Clore *et al.*, 1998b) did not yield lower energy structures indicating that the tensor is close to optimal.

The orientation of the alignment tensor coordinate system used in the XPLOR simulated annealing calculation is defined by four atoms placed 50-100 Å from the molecule. The origin of this pseudomolecule is harmonically constrained and the axes must remain orthogonal. However, the orientation of this four atom-coordinate system floats freely during the structure calculation since the direction of the molecular alignment tensor is not known *a priori* (Clore *et al.*, 1998b).

Structure description

Data that establish the secondary structure of KH3 are tabulated in Figure 2. The secondary structural elements are shown at the bottom of the Figure. Ribbon representations of KH3 are shown in Figure 3. KH3 adopts a βααββα fold. The three-stranded (residues 14-20, 59-65, 46-49) anti-parallel β-sheet has a left-handed twist. The solvent-exposed side of the β-sheet is comprised of hydrophilic residues with I60 as the only exception. The other side of the β-sheet is composed of hydrophobic residues except for the solvent accessible T16. Three amphipathic α-helices (residues 22-29, 34-43, 67-81) stack on the hydrophobic face of the β-sheet producing a hydrophobic core. Two of these helices, α2 and α3, are nearly anti-parallel to each other (interhelical angle = 159°) and nearly parallel with the long axis of the β-sheet. Helix α1 is connected to α2 by a turn motif (GxxG) that is highly conserved in KH domains. The interhelical angle between helices α1 and α2 is roughly 63°. A highly conserved glycine, G44, at the end of α2 sterically allows a sharp turn between α2 and β2. G65 at the end of β3 is also highly conserved and plays a similar role in the relative positioning of β3 and α3. A large loop (residues 50-58) connects β2 and β3. The ^{15}N relaxation data indicate that residues in the central region of this loop have lower order parameters than residues in regions of regular secondary structure (data not shown).

Structure quality

Figure 4 illustrates the superposition of the backbone of the final 20 structures calculated with and without dipolar couplings. It is apparent from Figure 4 that both families of structures are relatively well defined. Structures calculated with the aid of dipolar couplings have a root mean square deviation (RMSD) of 0.16 Å for the backbone atoms and 0.80 Å for all heavy atoms. Structures calculated without dipolar couplings are less precise with RMSDs of 0.39 and 0.92 Å for backbone and all heavy atoms, respectively. A diagram quantifying the backbone precision per residue is shown in Figure 5(a). As expected, the inclusion of

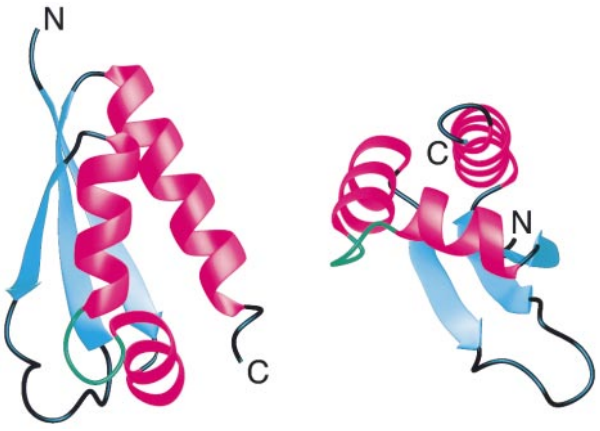

Figure 3. Two different views of the MOLMOL (Koradi *et al.*, 1996) ribbon diagram of the lowest overall energy KH3 structure. Residues 12-84 are shown. The GKGG loop (residues 30-33) is shown in green. The GxxG motif is highly conserved in the KH family.

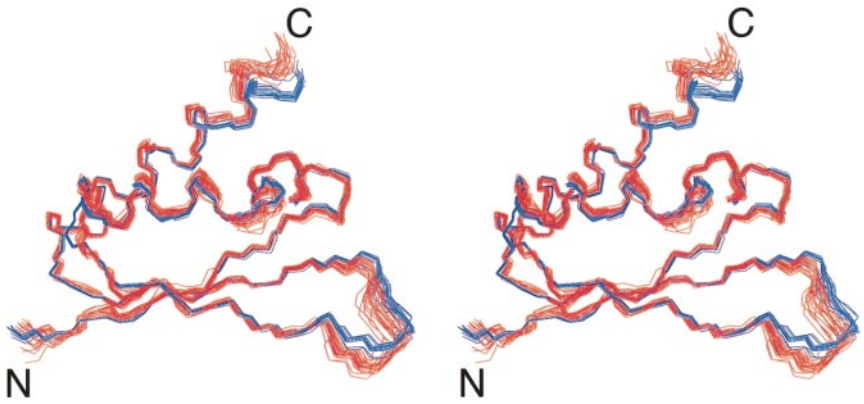

Figure 4. A stereoview of the backbone superposition of the 20 lowest energy structures calculated with (blue) and without (red) dipolar couplings. Residues 12-84 are shown.

high quality dipolar coupling restraints improved the precision of the structures, especially in areas not so well defined by NOEs like the GKGG loop (residues 30-33) which extends out into solution. Indeed, even flexible regions of the molecule (e.g. residues 53-57) are better defined due to the restrictions conferred by the half-open penalty functions (equation (3)) used for the dipolar couplings of residues in these areas. The agreement between the families of structures calculated with and without dipolar couplings is good as seen in Figure 4. Excluding the flexible regions (residues 1-12, 53-57, 82-89), the difference between the backbones of the mean structures of the two families is almost negligible with a pairwise root mean square (RMS) of 0.34 Å. The residue-specific pairwise RMSs calculated for the two families are presented in Figure 5(b).

There is still a need to have an independent evaluation of NMR structure quality since the precision of the structures also depends on the force constants used to minimize all of the different energy terms in the structure calculation. Dipolar couplings were used to assess structure quality by the recently proposed *Q*-factor approach (Drohat *et al.*, 1999; Ottiger *et al.*, 1998b). Briefly, a structure is calculated excluding one type of dipolar coupling. The excluded set is then compared to the final structure. The RMSD between the observed and calculated dipolar couplings is divided by the RMS value of the observed dipolar couplings. The *Q*-factor is a highly sensitive measure of any irregularities in the structure since a dipolar coupling is strongly dependent on the angle between the internuclear vector and the alignment tensor axis. Typical *Q*-factors for NMR structures not refined with any dipolar couplings range from 0.6 to 0.9. The $^1D_{C\alpha\text{-}H\alpha}$ and $^1D_{N\text{-}H}$ *Q*-factors for the refined KH3 structures are 0.26(±0.02) and 0.34(±0.03), respectively. A weighted average of these values equals 0.30(±0.02). This is in contrast to $^1D_{C\alpha\text{-}H\alpha}$, $^1D_{N\text{-}H}$, and weighted *Q*-factors of 0.32(±0.03), 0.57(±0.04), and 0.44(±0.03), respectively, for KH3 structures calculated without dipolar coupling restraints.

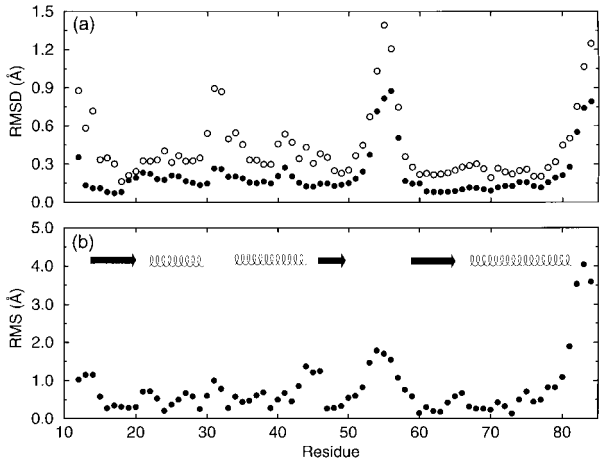

Figure 5. Backbone precision plots. (a) The RMSD per residue of the backbone atoms (C′, O, C^α, and N) for the 20 lowest overall energy structures of 100 calculated with (filled circles) and without (open circle) dipolar coupling constraints. (b) The residue-specific pairwise RMSs calculated from the mean structures of the two families in (a). Backbone atoms for residues 13-52 and 58-81 were used for all least squares fits. For consistency, only values for residues 12-84 are shown.

The ratio of backbone ^{15}N T_1 and T_2 values can be used as structural constraints when a molecule tumbles anisotropically ($D_\parallel/D_\perp \neq 1.0$) (Tjandra *et al.*, 1997). A T_1/T_2 ratio depends on the orientation of the internuclear vector relative to the rotational diffusion tensor. The alignment coordinate frame cannot generally be assumed to be the same as the diffusion frame. Since the ratios of the backbone relaxation times are physical quantities independent of the dipolar couplings and represent an independent coordinate frame, they are an obvious choice to use as an additional measure of structure quality. The ^{15}N T_1/T_2 ratios of KH3 were fit to the lowest energy structure calculated with and without dipolar couplings using a global fitting procedure described previously (Tjandra *et al.*, 1995). A fully asymmetric diffusion tensor model was assumed. The results of the fit are shown in Figure 6 as correlation plots between the observed and calculated T_1/T_2 ratios. The fitted diffusion anisotropy ($D_z/0.5(D_x + D_y)$) is 1.37 with a rhombicity factor η ($1.5(D_y - D_x)/(D_z - 0.5(D_y + D_x))$) of 0.47 and an effective correlation time of 6.87 ns. It is important to

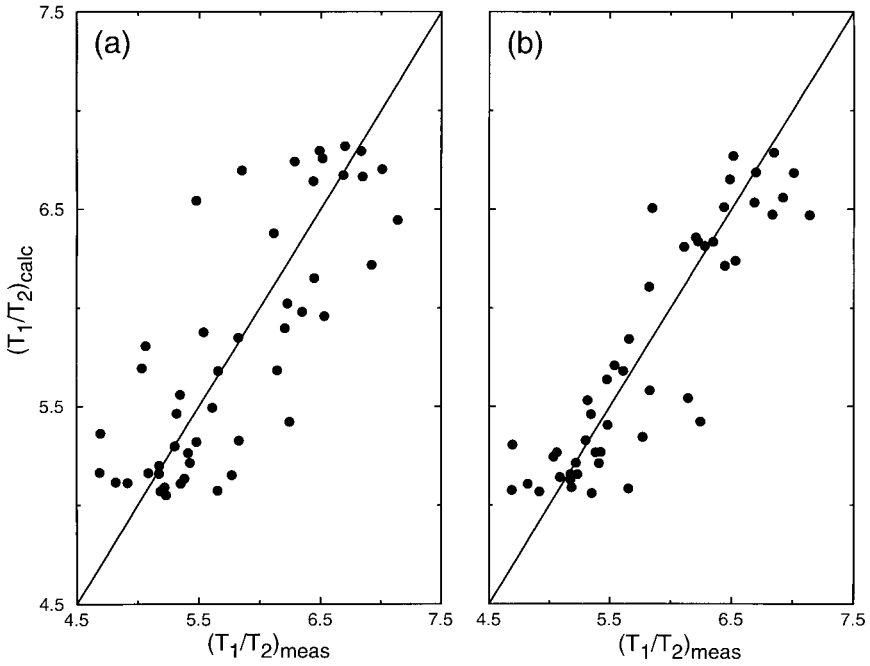

Figure 6. Comparisons of calculated and measured ^{15}N T_1/T_2 ratios. The ratios were fit to the lowest energy structure of 100 calculated for each case. The correlation of measured *versus* calculated ratios using the structure derived without dipolar couplings is shown in (a). The correlation factor (r) is 0.8. A higher correlation factor of 0.9 is obtained when the ratios were fit to the lowest energy structure derived with the aid of dipolar couplings (b).

emphasize that the structures used for this evaluation were not refined with the aid of ^{15}N T_1/T_2 restraints. The RMS differences between the observed and calculated T_1/T_2 ratios are 0.42 and 0.30 for structures calculated without and with dipolar coupling constraints, respectively. For comparison, the RMS difference between the observed and calculated ratios is 0.94 for the N-terminal domain of enzyme I (EIN) after the T_1/T_2 ratios were incorporated into the simulated annealing calculation. In contrast, EIN structures calculated without T_1/T_2 constraints had a RMS difference of 2.7 for the T_1/T_2 ratios (Tjandra *et al.*, 1997). This method of evaluating structure quality clearly shows that structures calculated with dipolar couplings are ''better'' overall.

Discussion

Structure comparisons

A DALI (Holm & Sander, 1993) search revealed the DNA-binding modules of bovine papillomavirus-1 E2[326-410] (E2) (Hedge *et al.*, 1992) and Epstein-Barr nuclear antigen 1[459-607] (EBNA1) (Bochkarev *et al.*, 1996, 1998) to be structurally similar to KH3. Both are homodimeric in their double-stranded (ds) DNA-bound form. EBNA1 contains a N-terminal helix that precedes the first β-strand and plays a role in nucleic acid-binding, but is otherwise structurally similar to E2. The general area of EBNA1 that interacts with DNA is similar to that of E2, though significant differences become apparent when the structures are critically compared. A backbone least squares fit of E2 in its bound form to KH3 (2.6 Å pairwise backbone RMSD) is shown in Figure 7. E2 has a fourth, C-terminal β-strand and there are only two helices covering one side of the β-sheet and thus possesses a βαββαβ fold. The N-terminal helix of E2 was shown to be a dsDNA recognition helix (Hegde *et al.*, 1992). The strong similarities between E2, EBNA1, and KH3 suggest that α1 and/or α2 of KH3 are possible sites of nucleic acid interaction.

The DALI search also showed a reasonable match between KH3 and another dsDNA-bound protein, the TATA-box binding protein (TBP) (Kim *et al.*, 1993). TBP possesses a βαββββα topology in the region matching KH3. However, unlike E2 and EBNA1, the region of TBP that interacts with the

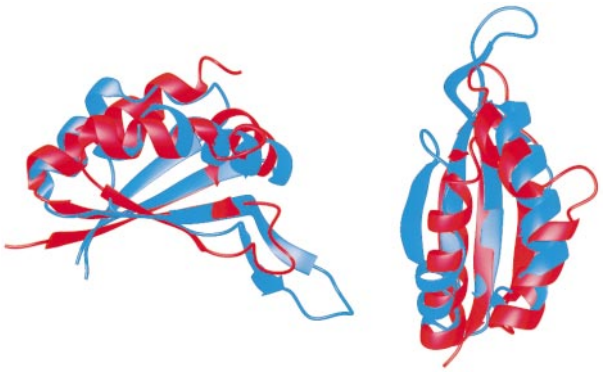

Figure 7. Two different views of superimposed ribbon diagrams of the dsDNA-bound form of bovine papillomavirus-1 E2 (Hegde *et al.*, 1992) (residues 326-406 shown) and KH3 (residues 12-84 shown). A backbone least squares fit of the two structures was performed before the diagrams were rendered. KH3 is shown in red and E2 is shown in blue. The atomic coordinates for bovine papillomavirus-1 E2 were downloaded from the Protein Data Bank at Brookhaven National Laboratory (accession number: 2bop). Note that the secondary structural elements were identified by MOLMOL (Koradi *et al.*, 1996).

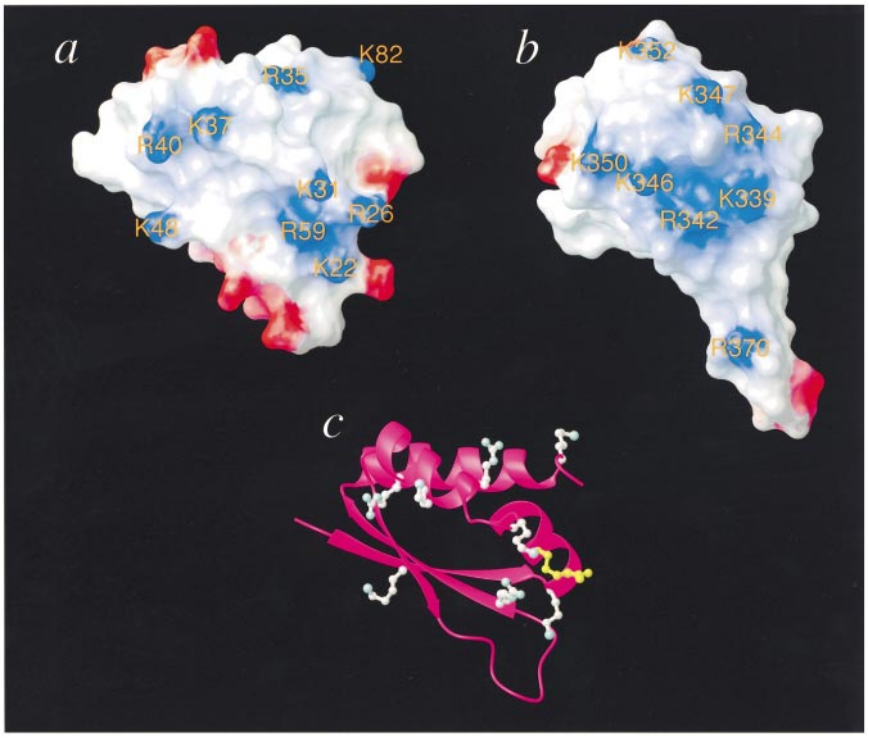

Figure 8. MOLMOL (Koradi *et al.*, 1996) surface representations of (a) KH3 (residues 12-84; lowest energy of 100 calculated structures) and (b) bovine papillomavirus-1 E2 (residues 326-406 shown). Positively charged areas are colored blue and negatively charged areas are in red. These representations were generated by first least squares fitting E2 to KH3 and then rotating the bundle to best display the dsDNA-binding region of E2 (i.e. the relative orientation between the two surfaces is determined by a least squares fit of the protein backbones). (c) The ribbon diagram of KH3 with basic residue side-chain heavy atoms and bonds displayed. R26, the mutated residue, is shown in yellow. The orientation of this diagram is equivalent to that of the structure used to generate (a).

nucleic acid is a β-sheet with helices covering the side opposite the interaction surface. The ribonucleoprotein (RNP) domain of U1A is structurally similar to KH3 and also binds nucleic acid, albeit RNA, at a β-surface (Allain *et al.*, 1996, 1997; Oubridge *et al.*, 1994). For completeness, we note that the double-stranded RNA-binding domain (dsRBD) fold also has two helices stacked on top of a β-sheet, but has αββββα topology (Nagai, 1996). No dsRBD matches were found in the DALI search for KH3.

In contrast to TBP and the RNP of U1A, it is unlikely that the solvent-exposed side of KH3's β-sheet will be a site of nucleic acid interaction. There are no aromatic groups and only one basic residue, K48, extending from the β-sheet side of KH3 that is not shielded by the helices. This latter observation is not generally true for all KH domains. Based on sequence alignment, there is a wide variation throughout the KH family in terms of the number of aromatic, basic, and hydrophobic residues extending from the side of the β-sheet not covered with the helices. Regardless, it seems plausible to rule out the β-sheet mode of nucleic acid binding in at least the case of KH3.

Two previously published KH domains, KH1 of FMR1 (Musco *et al.*, 1997) and KH6 of vigilin (Musco *et al.*, 1996), are both structurally similar to KH3 with pairwise backbone RMSDs of 2.8 and 2.6 Å, respectively (Note: the loop connecting β2 and β3 has a variable number of residues throughout the KH family and is neglected for this comparison.) This similarity is not illuminating since specific *in vivo* nucleic acid targets for KH1 and KH6 have not been identified nor has a protein-nucleic acid complex for either of these two KH domains been presented.

Possible nucleic acid binding sites of KH3

It is useful to consider the surface charge distribution of KH3 in order to derive a hypothesis for nucleotide interaction sites. This distribution is illustrated in Figure 8(a). All of KH3's most basic residues (i.e. Arg and Lys) with the exception of K87 are evident in this Figure. Residues on helices α1 and α2 (residues K22, R35, K37, R40), an arginine at the N-terminal end of β3 (R59), and K31 in the GKGG loop define a positive surface for KH3. These residues provide a biased distribution of positive charges on one surface of the protein and represent possible sites for electrostatic interaction with nucleotides. The other positively charged residues are K82 and K87 at the end of the C-terminal helix and K48, which is on the solvent-exposed side of β2. Residues S27 on α1 and S43 on α2 pose as possible hydrogen-bond donors. Of the residues listed above, K37, R40, S43, and R59 are the most highly conserved in the KH domain. R35 can be characterized as semi-conserved as there is typically a hydrogen-bond donor (e.g. Thr) in that position. While making an analogy to the conventional helix-hairpin-helix motif (Doherty *et al.*, 1996), Musco *et al.* (1996) have noted that the second residue away from the GxxG loop towards the C terminus (e.g. R35 in KH3) might form hydrogen-bonds with nucleotide backbone phosphates. Y75 is the only aromatic residue present in the KH module of our construct and is positioned on the hydrophilic side of α3. It is surprising not to find greater conservation of the basic and hydrogen-

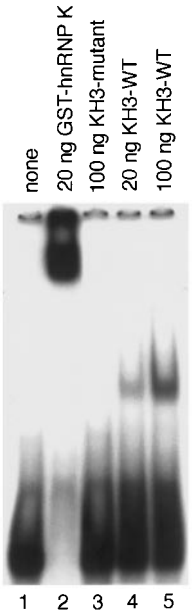

Figure 9. Wild-type KH3 binds dCT3. The G26R KH3 mutant shows no appreciable affinity for this same oligonucleotide. Lane 1, probe alone; lane 2, 20 ng of GST-hnRNP K; lane 3, 100 ng of the G26R KH3 mutant; land 4, 20 ng of wild-type KH3; lane 5, 100 ng of wild-type KH3.

bond donor residues in the KH family. On the other hand, as first suggested by Dejgaard & Leffers (1996), perhaps the variation at these seemingly crucial positions creates the diverse binding specificities observed for this class of proteins.

Missing from the discussion of putative binding sites is R26 of KH3. A conserved glycine occupies this position in wild-type KH3. This G26R substitution does not alter the overall fold of the protein as reflected by the good agreement between KH3 and the two previously published KH structures (Musco *et al*., 1996, 1997). In contrast to the high specificity and tight binding of wild-type KH3 to an oligonucleotide known to specifically bind hnRNP K, the KH3 mutant exhibits almost no affinity for ssDNA as demonstrated in the EMSA shown in figure 9. This clearly suggests that R26 abrogates binding through steric effects and/or charge interference. Consequently, this structurally benign mutation localizes a site of nucleic acid interaction.

The analysis so far indicates, in a common-sense fashion, possible nucleotide binding sites of KH3. For additional corroboration, we examined surface-charge representations of some of the proteins previously discussed. The surface most similar to KH3 is shown in Figure 8(b). It is the dsDNA-binding surface of E2. The two surfaces shown in Figure 8 were generated after first least squares fitting KH3 to E2 and orienting the bundle to best display the nucleic acid-binding region of E2. R342, R344, and R370 terminal amide protons of E2 are in close proximity to the dsDNA backbone while K339 is involved in specific base recognition. Three other residues of E2 not as evident in this diagram, N336, C340, F343, are also involved in specific base recognition. Backbone fits of KH3 and E2 show that the side-chain termini of R59 and R342 are spatially similar even though R59 extends up from the β-sheet of KH3 whereas R342 extends from the first helix of E2. The side-chain position of K31 is not well-defined in KH3, but is seemingly comparable to K339. Hence, KH3 is not only structurally similar to E2, but its putative binding surface has a charge distribution similar to that of the binding region of E2. In contrast, similarities between the surface charge distributions of EBNA1 and KH3 are not so readily apparent. KH6 of vigilin (Musco *et al*., 1996) exhibits a surface charge distribution only moderately comparable to the regions shown in Figure 8(a) whereas that of KH1 of FMR1 (Musco *et al*., 1997) shows much less similarity.

This discussion on KH3 putative binding sites is consistent with the proposal by Musco *et al*. that the GxxG loop, which is highly conserved and connects helices α1 and α2 in KH3, is critical for binding in KH domains (Musco *et al*., 1996, 1997). There is always at least one basic or hydrogen-bonding residue in the center of this loop. Musco *et al*. presented backbone dynamics data to show that the GxxG loop in KH1 of FMR1 is flexible with motions in the microsecond to millisecond time-scale (Musco *et al*., 1997) and predicted this loop to be flexible in KH6 of vigilin (Musco *et al*., 1996). This conformational freedom is thought to be essential for nucleic acid binding. There is notable precedence for this idea. Several studies (Doherty *et al*., 1996; Markus *et al*., 1997; Puglisi *et al*., 1995; Thayer *et al*., 1995) have shown local flexibility in the binding region of the uncomplexed protein for the helix-hairpin-helix (HhH) motif.

In contrast to the GxxG loop of FMR1 KH1 (Musco *et al*., 1997), the GKGG loop in KH3 does not show an appreciable decrease in the ^{15}N T_2 values. Perhaps performing the NMR experiments at a substantially lower pH influenced the exchange contribution to the ^{15}N T_2. Indeed, absence of the cross-peak of G32 from the ^{1}H-^{15}N correlation experiment indicates that this particular region of the protein undergoes an exchange process. This is consistent with the hypothesis of Musco *et al*. (1997) regarding the importance of loop flexibility for nucleotide binding.

The suggested involvement of helices α1 and α2 of KH3 in nucleic acid binding is supported by recently published structures of ribosomal RNA-binding proteins ErmAM (Yu *et al*., 1997), VP39 (Hodel *et al*., 1996), S15 (Berglund *et al*., 1997), and L11 (Markus *et al*., 1997; Xing *et al*., 1997). In each of these proteins, the RNA-binding domain is comprised of helices. Yu *et al*. (1997) propose that ErmAm will bind to a relatively large region of its 23 S rRNA target based on a chemical footprinting analysis of a protein homologous to ErmAM,

ErmC′ (Su & Dubnau, 1990). Similarly, hnRNP K may exploit a large surface comprised of multiple KH domains to recognize repeated CT elements. The existence of multiple KH domains in hnRNP K and its specificity to a nine base CT repeat suggests the possibility of inter-KH domain cooperativity in nucleotide binding.

Conclusion

Is the overall architecture of a nucleic acid-binding module critical for discerning between DNA and RNA or does it merely provide a scaffold to which the appropriate recognition elements are attached? Can a protein possessing a general nucleic acid-binding fold therefore be mutated to alter its nucleic acid preference? The idea of a modifiable scaffold is the most consistent with current literature. E2 (Hedge *et al*., 1992), EBNA1 (Bochkarev *et al*., 1996, 1998), TATA-box binding protein (Kim *et al*., 1993), and KH3 have folds similar to those of the RNA-binding RNP domains, yet the three former bind dsDNA and the latter binds ssDNA. L11 (Markus *et al*., 1997; Xing *et al*., 1997) and S15 (Berglund *et al*., 1997) have folds comparable to helix-turn-helix DNA binding domains, yet they bind RNA. Furthermore, can a particular nucleic acid-binding module have similar affinities for both specific DNA and RNA targets as proposed for KH3? The well-known *Xenopus laevis* transcription factor TFIIIA, a zinc-finger protein, exhibits comparable affinities for both specific DNA and RNA targets (Honda & Roeder, 1980; Pelham & Brown, 1980). Hence, the idea that select proteins exhibit dual binding roles is not too surprising. It would be interesting to choose a specific nucleic acid-binding protein and make select, subtle mutations that alter the protein's partitioning between DNA and RNA targets without altering the protein's fold. We are currently determining the structure of the KH3-CT complex to aid in such investigations. It is anticipated that the enhanced precision and accuracy afforded by dipolar coupling restraints will be a significant help in such endeavors.

Materials and Methods

Sample preparation

KH3 was expressed as a fusion protein with glutathione-*S*-transferase (GST). The DNA coding sequence comprising amino acids 380 to 463 of hnRNP K was cloned into a pGEX-4T-3 plasmid. *Escherichia coli* BL21(DE3) was used as the bacteria expression host. Uniform labeling (>99 %) was accomplished by growing the cells in a minimal medium at 37 °C with $^{15}NH_4Cl$ and/or [$^{13}C_6$]glucose as the sole source of nitrogen and/or carbon, respectively. Protein expression was induced for four hours at 37 °C with 0.5 mM isopropyl-D-thiogalactoside. The cells were harvested, resuspended in phosphate buffered saline (PBS) (pH 7.4), lysed by sonication, and centrifuged for 45 minutes at 27,000 *g*. The supernatant was loaded onto a glutathione-Sepharose column equilibrated with the PBS buffer. The column was rinsed with a large amount of PBS to remove impurities. KH3 was cleaved from the column with biotinylated thrombin. Cleavage took place at room temperature for three to four hours. Thrombin remaining in the eluted fraction was removed with streptavidin agarose. The sample was further purified using a C-4 reversed-phase high performance liquid chromatography (HPLC) column with a 20 to 60 % acetonitrile gradient in 0.1 % (v/v) aqueous trifluoroacetic acid and then lyophilized.

All NMR samples were 1 mM protein at a pH (or pH*) of 5.5 with a maximum of 5 mM of KCl. Sodium azide was added to a final concentration of 0.1 %. All NMR samples are either in 90 % H_2O/10 % ^2H_2O or 99.9 % ^2H_2O.

DNA binding

EMSA was performed on a 8 % (w/v) polyacrylamide gel using 10 fmol of ^{32}P-labeled dCT3 (AAT TCT CCT CCC CAC CTT CCC CAC CCT CCC CA) as the probe. All proteins were expressed as fusion proteins with GST. GST was removed for the KH3-mutant and KH3-wild-type samples. Additional conditions used for the EMSA are similar to those described by Tomonaga & Levens (1995).

NMR spectroscopy

NMR experiments were performed on a Bruker AMX600 operating at the ^{1}H frequency of 600 MHz equipped with a shielded *z*-gradient triple resonance probe. All NMR experiments except those involving the anisotropic (liquid crystal) sample were carried out at 27 °C. NMRPipe (Delaglio *et al*., 1995) and PIPP/STAPP (Garrett *et al*., 1991) software were used to process and analyze the spectra, respectively. The following 3D heteronuclear NMR experiments were used to make ^{1}H, ^{15}N, and ^{13}C resonance assignments: CBCANH, CBCA(CO)NH, HBHA(CO)NH, HNCO, HNHA, and HCCH-TOCSY (Bax & Grzesiek, 1993; Clore & Gronenborn, 1991, 1998a). Backbone ϕ angle restraints were derived from $^3J_{HNHA}$ values measured in a quantitative *J* correlation HNHA experiment (Vuister & Bax, 1993). A HNHB experiment aided in the stereospecific assignment of methylene H^β protons (Archer *et al*., 1991) for 12 residues. The following NOE experiments provided interproton distance restraints: 3D ^{15}N separated NOESY (150 ms mixing time), 3D ^{15}N separated ROESY (40 ms mixing time), 4D ^{15}N/^{13}C separated NOESY (120 ms mixing time), and 4D ^{13}C/^{13}C separated NOESY (110 ms mixing time) (Bax & Grzesiek, 1993; Clore & Gronenborn, 1991, 1998a).

Dipolar coupling measurements

Dipolar couplings were calculated from the difference in corresponding *J* splittings measured in the anisotropic (liquid crystalline) phase and in the isotropic phase. A KH3 sample in water as described above was used to measure all of the isotropic *J* values. The liquid crystal sample contained 3.2 % dimyristoyl phosphatidylcholine/dihexanoyl phosphatidylcholine (3.5:1 molar ratio), 0.6 mM protein, 0.1 % NaN_3, and 10 % ^2H_2O. The pH of this sample was adjusted to 6.1. All measurements were made at 29 °C, which is close to the minimum temperature for a stable liquid crystalline phase.

A 2D IPAP ^{15}N-^{1}H HSQC experiment was used to measure $^{1}J_{\text{N-H}}$ values. This experiment reduces spectral overlap by separately recording the in-phase and anti-phase components of the ^{15}N-^{1}H doublets that result when ^{1}H coupling is allowed during ^{15}N evolution (Ottiger *et al.*, 1998a). 256 complex t_1 (^{15}N) and 768 complex t_2 (^{1}H) points were recorded with 12 scans per t_1 increment for the isotropic sample. A total of 20 scans per t_1 increment were used for the anisotropic sample. Spectral widths were 32.92 ppm and 18.54 ppm for t_1 and t_2, respectively. It took four hours to acquire both sets of interleaved data for the isotropic sample and seven hours for the anisotropic one. Both dimensions were zero-filled to 2048 points in the final data matrix. A squared, 72° phase-shifted sine-bell window function was used to process t_2. A 72° phase-shifted sine-bell window function was used for t_1 processing.

$^{1}J_{\text{N-C}'}$ values were measured from a simple 2D ^{15}N-^{1}H HSQC without C′ decoupling during nitrogen evolution (Delaglio *et al.*, 1991). 300 complex t_1 (^{15}N) and 1024 (or 768) complex t_2 (^{1}H) points were recorded. A total of 48 scans per t_1 increment were used for the anisotropic measurements and 16 scans per t_1 increment were used to record the spectrum of the isotropic sample. Spectral widths were 25.69 and 15.15 ppm for t_1 and t_2, respectively. The total experiment time was ten hours for the anisotropic sample and 3.3 hours for the isotropic one. t_1 was zero-filled to 2048 points and t_2 was zero-filled to 8192 points. 54° and 72° phase-shifted sine-bell window functions were used on t_1 and t_2, respectively.

$^{1}J_{\text{C}\alpha\text{-C}'}$ values are measured by CT-HNCO experiments (Grzesiek & Bax, 1992, 1993; Kay *et al.*, 1990) without C$^{\alpha}$ decoupling during C′ evolution. A total of 24 complex t_1 (^{15}N), 100 complex t_2 (^{13}C), and 512 complex t_3 (^{1}H) points were recorded for the isotropic sample. There were 32 complex points for t_1 in the anisotropic case. There were eight scans per t_2 increment for both samples. Spectral widths were 19.57, 10.35, and 14.17 ppm for t_1, t_2, and t_3, respectively. The total experiment time was 29 and 39 hours for the isotropic and anisotropic samples, respectively. t_1, t_2, and t_3 were zero-filled to 128, 512, and 1024 points, respectively. A squared, 72° phase-shifted sine-bell window function was used to process the data in t_3. A 72° phase-shifted sine-bell window function was used for the two remaining dimensions.

$^{1}J_{\text{C}\alpha\text{-H}\alpha}$ couplings were measured by two different methods. The first experiment was a 3D CT-(H)CA(CO)NH experiment without H$^{\alpha}$ decoupling during C$^{\alpha}$ evolution. A total of 45 complex t_1 (^{15}N), 76 complex t_2 (^{13}C), and 512 complex t_3 (^{1}H) points were recorded. There were eight scans per t_2 increment. Spectral widths were 29.93, 24.12, and 18.27 ppm for t_1, t_2, and t_3, respectively. The total experiment time was 38 hours. t_1, t_2, and t_3 were zero-filled to 128, 512, and 1024 points, respectively. A squared, 72° phase-shifted sine-bell window function was used to process the data in each dimension. The second method used to measure $^{1}J_{\text{C}\alpha\text{-H}\alpha}$ couplings was a *J*-modulated [^{13}C-^{1}H] CT-HSQC. The ^{13}C constant-time period ($=1/J_{\text{CC}} \approx 28$ ms) for a [^{1}H-^{13}C] CT-HSQC was varied. A total of 20 *J*-modulated CT-HSQC planes were obtained. Peak intensities for each correlation were fit to yield the $^{1}J_{\text{C-H}}$ couplings (Ottiger *et al.*, 1998b). For each plane, the following parameters and methods were used: 128 complex t_1 (^{13}C) and 384 complex t_2 (^{1}H) points; eight scans per t_1 increment; spectral widths were 30.12 and 12.51 ppm for t_1 and t_2, respectively; t_1 and t_2 were zero-filled to 256 and 1024 points, respectively; a 72° phase-shifted sine-bell window function was used to process the data in both t_1 and t_2. It took 14.2 hours to collect all of the data for this experiment. The dipolar coupling values determined from the two different methods, 3D CT-(H)CA(CO)NH and *J*-modulated [^{13}C-^{1}H] CT-HSQC, agree quite well with a pair-wise RMSD of 2.3 Hz. The final $^{1}D_{\text{C}\alpha\text{-H}\alpha}$ set of restraints was constructed by averaging the results from the two different experiments.

Structure calculations

Peak intensities from the NOE experiments were translated into a continuous distribution of interproton distance restraints. A summation averaging $[(\Sigma r^{-6})^{-1/6}]$ was utilized for ambiguous and non-stereospecifically assigned NOEs (Nilges, 1993). Generic hydrogen bond distance restraints were utilized for regions of regular secondary structure which were based on NOE patterns and secondary ^{13}C$^{\alpha}$ and ^{13}C$^{\beta}$ chemical shifts. Structures were calculated by a distance geometry/simulated annealing protocol (Clore & Gronenborn, 1998b; Nilges *et al.*, 1988) using a modified version of XPLOR 3.1 (Brunger, 1993) in which pseudopotentials for dipolar coupling restraints were incorporated (Clore *et al.*, 1998b). The starting structure was heated to 4000 K and cooled in 40,000 steps of 0.002 ps for the final simulated annealing calculations.

PDB accession codes

The coordinates of 20 simulated annealing structures and chemical shift assignments for KH3 of hnRNP K have been deposited in the Protein Data Bank (accession codes 1khm and 1khmmr).

Acknowledgments

We thank Dan Garrett, John Marquardt, and Marcel Ottiger for helpful software scripts and conservations.

References

Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature,* **380**, 646-650.

Allain, F. H., Howe, P. W., Neuhaus, D. & Varani, G. (1997). Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* **16**, 5764-5772.

Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991). An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonance in larger proteins. *J. Magn. Reson.* **95**, 636-641.

Ashley, C. T., Jr, Wilkinson, K. D., Reines, D. & Warren, S. T. (1993). FMR1 Protein: Conserved RNP Family Domains and Selective RNA Binding. *Science,* **262**, 563-566.

Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Acc. Chem. Res.* **26**, 131-138.

Berglund, H., Rak, A., Serganov, A., Garber, M. & Hard, T. (1997). Solution structure of the ribosomal RNA

binding protein S15 from *Thermus thermophilus*. *Nature Struct. Biol.* **4**, 20-23.

Bochkarev, A., Barwell, J., Pfuetzner, R., Bochkareva, E., Frappier, L. & Edwards, A. (1996). Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA. *Cell,* **84**, 791-800.

Bochkarev, A., Bochkareva, E., Frappier, L. & Edwards, A. (1998). The 2.2 Å structure of a permanganate-sensitive DNA site bound by the Epstein-Barr virus origin binding protein, EBNA1. *J. Mol. Biol.* **284**, 1273-1278.

Bomsztyk, K., Van Seuningen, I., Suzuki, H., Denisenko, O. & Ostrowski, J. (1997). Diverse molecular interactions of the hnRNP K protein. *FEBS Letters,* **403**, 113-115.

Brunger, A. T. (1993). *XPLOR Manual Version 3.1*, Yale University, New Haven, Connecticut.

Bustelo, X., Suen, K., Michael, W., Dreyfuss, G. & Barbacid, M. (1995). Association of the vav proto-oncogene product with poly(rC)-specific RNA-binding proteins. *Mol. Cell Biol.* **15**, 1324-1332.

Clore, G. M. & Gronenborn, A. M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science,* **252**, 1390-1399.

Clore, G. M. & Grenenborn, A. M. (1988a). Determining the structures of larger proteins and protein complexes by NMR. *Trends Biotech.* **16**, 22-34.

Clore, G. M. & Gronenborn, A. M. (1998b). New Methods of structure refinement for macromolecular structure determination by NMR. *Proc. Natl Acad. Sci. USA,* **95**, 5891-5898.

Clore, G. M., Gronenborn, A. M. & Bax, A. (1998a). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133**, 216-221.

Clore, G. M., Gronenborn, A. M. & Tjandra, N. (1998b). Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159-162.

Collier, B., Goobar-Larsson, L., Sokolowski, M. & Schwartz, S. (1998). Translational inhibition *in vitro* of human papillomavirus type 16 L2 mRNA mediated through interaction with heterogeneous ribonucleoprotein K and poly(rC)-binding proteins 1 and 2. *J. Biol. Chem.* **273**, 22648-22656.

Dejgaard, K. & Leffers, H. (1996). Characterisation of the nuclei-acid-binding activity of KH domains. *Eur. J. Biochem.* **241**, 425-431.

Delaglio, F., Torchia, D. A. & Bax, A. (1991). Measurement of ^{15}N-^{13}C J couplings in staphylococcal nuclease. *J. Biomol. NMR,* **1**, 439-446.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* **6**, 277-293.

Denisenko, O., O'Neill, B., Ostrowski, J., Van Seuningen, I. & Bomsztyk, K. (1996). Zik1, a transcriptional repressor that interacts with the heterogeneous nuclear ribonucleoprotein particle K protein. *J. Biol. Chem.* **271**, 27701-27706.

Doherty, A. J., Serpell, L. C. & Ponting, C. P. (1996). The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucl. Acids Res.* **24**, 2488-2497.

Drohat, A. C., Bax, A., Tjandra, N., Baldisseri, D. & Weber, D. J. (1999). The use of dipolar couplings for determining the solution structure of rat Apo-S100B(ββ). *Protein Sci.* **8**, 800-809.

Du, Q., Melnikova, I. N. & Gardner, P. D. (1998). Differential effects of heterogeneous nuclear ribonucleoprotein K on Sp1- and Sp3-mediated transcriptional activation of a neuronal nicotinic acetylcholine receptor promoter. *J. Biol. Chem.* **273**, 19877-19883.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M. & Levens, D. (1994). A sequence-specific, single-strand binding protein activates the far upstream element of c-*myc* and defines a new DNA-binding motif. *Genes Dev.* **8**, 465-480.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214-220.

Gibson, T., Rice, P., Thompson, J. & Heringa, J. (1993a). KH domains within the FMR1 sequence suggest that fragile X syndrome stems from a defect in RNA metabolism. *Trends Biochem. Sci.* **18**, 331-333.

Gibson, T. J., Thompson, J. D. & Heringa, J. (1993b). The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. *FEBS Letters,* **324**, 361-366.

Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432-440.

Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment of uniformly $^{13}C/^{15}N$-enriched proteins. *J. Biomol. NMR,* **3**, 185-204.

Hegde, R., Grossman, S., Laimins, L. & Sigler, P. (1992). Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature,* **359**, 505-512.

Hodel, A. E., Gershon, P. D., Shi, X. & Quiocho, F. A. (1996). The 1.85 Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell,* **85**, 247-256.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.

Honda, B. & Roeder, R. (1980). Association of a 5S gene transcription factor with 5S RNa and altered levels of the factor during cell differentiation. *Cell,* **22**, 119-126.

Hsieh, T., Matsumoto, M., Chou, H., Schneider, R., Hwang, S. B., Lee, A. S. & Lai, M. M. C. (1998). Hepatitis C virus core protein interacts with heterogeneous nuclear ribonucleoprotein K. *J. Biol. Chem.* **273**, 17651-17659.

Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496-514.

Kim, Y., Geiger, J., Hahn, S. & Sigler, P. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365**, 512-520.

Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55.

Kung, H. C., Wang, K. Y., Goljer, I. & Bolton, P. H. (1995). Magnetic alignment of duplex and quadruplex DNAs. *J. Magon. Reson. ser. B,* **109**, 323-325.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thorton, J. M. (1993). PROCHECK: a program to

check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR,* **8**, 477-486.

Leffers, H., Dejgaard, K. & Celis, J. E. (1995). Characterisation of two major cellular poly(rC)-binding human proteins, each containing three K-homologous (KH) domains. *Eur. J. Biochem.* **230**, 447-453.

Markus, M. A., Hinck, A. P., Huang, S., Draper, D. E. & Torchia, D. A. (1997). High resolution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nature Struct. Biol.* **4**, 70-77.

Matunis, M. J., Michael, W. M. & Dreyfuss, G. (1992). Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell Biol.* **12**, 164-171.

Miau, L., Chang, C., Shen, B., Tsai, W. & Lee, S. (1998). Identification of heterogeneous nuclear ribonucleoprotein K (hnRNP K) as a repressor of C/EBPβ-mediated gene activation. *J. Biol. Chem.* **273**, 10784-10791.

Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D. (1996). Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell. Biol.* **16**, 2350-2360.

Musco, G., Stier, G., Joseph, C., Antonietta, M., Morelli, C., Nilges, M., Gibson, T. J. & Pastore, A. (1996). Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell,* **85**, 237-245.

Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T. J., Nilges, M. & Pastore, A. (1997). The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nature Struct. Biol.* **4**, 712-716.

Nagai, K. (1996). RNA-protein complexes. *Curr. Opin. Struct. Biol.* **6**, 53-61.

Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by ^{1}H-NMR. *Proteins: Struct. Funct. Genet.* **17**, 297-309.

Nilges, M., Gronenborn, A. M., Brunger, A. T. & Clore, G. M. (1988). Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.* **2**, 27-38.

Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M. & Hentze, M. W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3′ end. *Cell,* **89**, 597-606.

Ostrowski, J., Sims, J., Sibley, C., Valentine, M., Dower, S., Meier, K. & Bomsztyk, K. (1991). A serine/threonine kinase activity is closely associated with a 65-kDa phosphoprotein specifically recognized by the kappa B enhancer element. *J. Biol. Chem.* **226**, 12722-12733.

Ottiger, M., Delaglio, F. & Bax, A. (1998a). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373-378.

Ottiger, M., Delaglio, F., Tjandra, N. & Bax, A. (1998b). Measurement of dipolar couplings for methylene and methyl sites in weakly oriented macromolecules and their use in structure determination. *J. Magn. Reson.* **134**, 365-369.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature,* **372**, 432-438.

Pelham, H. & Brown, D. (1980). A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. Natl Acad. Sci. USA,* **77**, 4170-4174.

Puglisi, J. D., Chen, L., Banchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science,* **270**, 1200-1203.

Sanders, C. R. & Schwonek, J. P. (1992). Characterization of magentically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry,* **31**, 8898-8905.

Siomi, H., Matunis, M. J., Michael, W. M. & Dreyfuss, G. (1993a). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucl. Acids Res.* **21**, 1193-1198.

Siomi, H., Siomi, M. C., Nussbaum, R. L. & Dreyfuss, G. (1993b). The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell,* **74**, 291-298.

Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L. & Dreyfuss, G. (1994). Essential role for KH domains in RNa binding: impaired RNa binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell,* **77**, 33-39.

Su, S. L. & Dubnau, D. (1990). Binding of *Bacillus subtilis* ErmC′ methyltransferase to 23S ribosomal RNA. *Biochemistry,* **29**, 6033-6042.

Swanson, M. S. & Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.* **8**, 2237-2241.

Takimoto, M., Tomonaga, t. , Matunis, M., Avigan, M., Krutzsch, H., Dreyfuss, G. & Levens, D. (1993). Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-*myc* promoter, *in vitro*. *J. Biol. Chem.* **268**, 18249-18258.

Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. & Tainer, J. A. (1995). Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.* **14**, 4108-4120.

Tjandra, N. & Bax, A. (1997a). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science,* **278**, 1111-1114.

Tjandra, N. & Bax, A. (1997b). Measurement of dipolar contributions to $^{1}J_{CH}$ splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra. *J. Magn. Reson.* **124**, 512-515.

Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566.

Tjandra, N., Grzesiek, S. & Bax, A. (1996). Magnetic field dependence of nitrogen-proton J splittings in ^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* **118**, 6264-6272.

Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. (1997). Defining long range order in NMR structure determination from the dependence

of heteronuclear relaxation times on rotational diffusion anisotropy. *Nature Struct. Biol.* **4**, 443-449.

Tolman, J. R. & Prestegard, J. H. (1996). A quantitative J-correlation experiment for the accurate measurement of one-bond amide ^{15}N-^{1}H couplings in proteins. *J. Magon. Reson. ser. B,* **112**, 245-252.

Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegrad, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins - information for structure determination in solution. *Proc. Natl Acad. Sci. USA,* **92**, 9279-9283.

Tomonaga, T. & Levens, D. (1995). Heterogeneous nuclear ribonucleoprotein K is a dNA-binding transactivator. *J. Biol. Chem.* **270**, 4875-4881.

Tomonaga, T. & Levens, D. (1996). Activating transcription from single stranded DNA. *Proc. Natl Acad. Sci. USA,* **93**, 5830-5835.

Urlaub, H., Kruft, V., Bischof, O., Muller, E. C. & Wittmann-Liebold, B. (1995). Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J.* **14**, 4578-4588.

Van Seuningen, I., Ostrowski, J. & Bomsztyk, K. (1995a). Description of an IL-1-responsive kinase that phosphorylates the K protein. Enhancement of phosphorylation by selective DNA and RNA motifs. *Biochemistry*, **34**, 5644-5650.

Van Seuningen, I., Ostrowski, J., Bustelo, X., Sleath, P. & Bomsztyk, K. (1995b). The K protein domain that recruits the interleukin 1-responsive K protein kinase lies adjacent to a cluster of c-Src and Vav SH3-binding sites. Implications that K protein acts as a docking platform. *J. Biol. Chem.* **270**, 26976-26985.

Vuister, G. W. & Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond $J(H^{N}H^{\alpha})$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772-7777.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P. & McCormick, F. (1992). Molecular cloning and nucleic acid binding properties of the GAP- associated tyrosine phosphoprotein p62. *Cell,* **69**, 551-558.

Xing, Y., GuhaThakurta, D. & Draper, D. E. (1997). The RNA binding domain of ribosomal protein L11 is structurally similar to homoedomains. *Nature Struct. Biol.* **4**, 24-27.

Yu, L., Petros, A. M., Schnuchel, A., Zhong, P., Severin, J. M., Walter, K., Holzman, T. F. & Fesik, S. W. (1997). Solution Structure of an rRNA methyltransferase (ErmAM) that confers macrolide-lincosamide-streptogramin antibiotic resistance. *Nature Struct. Biol.* **4**, 483-489.

Edited by P. E. Wright

(Received 18 February 1999; received in revised form 19 April 1999; accepted 20 April 1999)