Role of Dimerization in KH/RNA Complexes: The Example of Nova KH3

Andres Ramos, David Hollingworth, Sarah A. Major, Salvatore Adinolfi, Geoff Kelly, Fred W. Muskett, and Annalisa Pastore*

NIMR, The Ridgeway, London NW7 1AA, United Kingdom

Received October 31, 2001; Revised Manuscript Received January 16, 2002

ABSTRACT: The K homology module, one of the most common RNA-binding motifs, is present in multiple copies in both prokaryotic and eukaryotic regulatory proteins. Increasing evidence suggests that self-aggregation of KH modules has a functional role. We have used a combination of techniques to characterize the behavior in solution of the third KH domain of Nova-1, a paradigmatic KH protein. The possibility of working on the isolated module allowed us to observe specifically the homodimerization and RNA-binding properties of KH domains. We provide conclusive evidence that self-association of Nova-1 KH3 occurs in solution even in the absence of RNA. Homodimerization involves a specific protein/protein interface. We also studied the dynamical behavior of Nova-1 KH3 in isolation and in complex with RNA. These data provide a model for the mechanism of KH/RNA recognition and suggest functional implications of dimerization in KH complexes. We discuss our findings in the context of the whole KH family and suggest a generalized mode of interaction.

Dimerization through selected domains is an important feature common to RNA-binding proteins, which, in this way, achieve cooperatively the necessary affinity and specificity for their targets (*1*). Among the main RNA-binding protein families, an increasing number of modular proteins containing KH domains[1] (*2*) have been shown to form dimers (*3*, *4*). KH modules consist of an $\alpha-\beta$ fold with a topology similar to that found in many ribosomal proteins (*5*−*10*) (Figure 1). KH motifs have been identified, usually in multiple copies, in more than 100 proteins, many of which take part in the complex network of protein/protein and protein/RNA interactions regulating gene expression in eukaryotes (*11*). It has been suggested that RNA-binding occurs through an induced-fit mechanism (*6*). Despite the frequent occurrence of KH modules and their possible implication in human genetic diseases, relatively little is known about the mechanism of dimerization of the KH protein family and the precise role of the KH domain in self-assembly. We address the question of whether KH dimerization occurs in solution and how it modulates the interaction of the motif with RNA. As a paradigm, we have characterized the behavior of the third KH domain of Nova-1 (Nova-1 KH3, Figure 1) both in isolation and in complex with RNA. Nova proteins are linked to the autoimmune disease paraneoplastic opsoclonus-mioclonus ataxia (POMA) (*12*). Although their function has not yet been fully clarified, the two close homologues Nova-1 and Nova-2 are known to be involved in regulation of neuronal pre-mRNA alternative splicing.

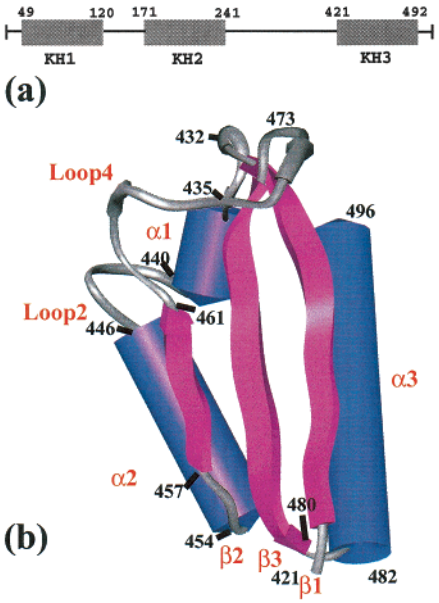

FIGURE 1: (a) Domain organization of the Nova-1 protein. (b) Ribbon representation of the structure of Nova-1 KH3 domain. The picture was prepared with the program InsightII and the pdb file 1dtj.pdb (*7*). The secondary structure elements are indicated in blue (helices) and magenta (β-strands) and labeled accordingly. The amino acid positions of the secondary structure elements are labeled.

Nova was selected for our studies because it is so far the best characterized of the KH-containing proteins both with regard to structure and binding specificity. By extensive SELEX studies, Nova proteins have been shown to recognize specifically UCAU repeats separated by short linkers (*12*−*15*). These sequences are present in Nova-1 and Nova-2 mRNAs and in the mRNA of a Gly receptor. Isolated Nova KH3 binds RNA with an affinity much higher than the other

* To whom correspondence should be addressed. Tel +44-20-8959-3666; fax +44−20−8906−4477; e-mail apastor@nimr.mrc.ac.uk.

[1] Abbreviations: AU: analytical ultracentrifugation; HSQC: heteronuclear single quantum coherence; KH: K homology; KH3: the third KH domain of Nova; NMR: nuclear magnetic resonance; NOESY: nuclear Overhauser effect spectroscopy; POMA: auto-immune disease paraneoplastic opsoclonus-mioclonus ataxia; T_1 and T_2: longitudinal and transversal relaxation times; TOCSY: total correlation spectroscopy; τ_c: correlation time.

10.1021/bi011994o CCC: $22.00
Published on Web 03/09/2002

two KH motifs (*13*). Structural studies on Nova-1 KH3 domains provide interesting and yet puzzling results of potential relevance for KH self-assembly. While uncomplexed Nova-1 and Nova-2 KH3s (which share 90% identity) crystallize as tetramers with two distinct symmetric interfaces, the complex of Nova-2 KH3 with a SELEX-derived RNA hairpin crystallizes as a dimer (*7*, *16*). The protein/protein interfaces involved are different from those observed in a NusA structure in which two KH modules from the same molecule are packed asymmetrically (*10*) shedding doubt on the in vivo significance of these findings. Each protein monomer binds one RNA hairpin, thus defining a 1:1 stoichiometry for the complex. Comparison of the crystal structure of Nova KH3 with results in solution for other KH domains also raises the question of whether a conformational change occurs upon RNA binding. In the crystal structure, the protein loop mainly involved in RNA binding retains essentially the same conformation in the isolated protein and in the complex, giving no support to the suggested induced-fit mechanism (*6*).

With the aim of clarifying some of the issues raised by the crystallographic structures, we have characterized the behavior of Nova KH3 in solution and its RNA-binding properties, specifically focusing on the self-assembly properties of the domain and the role that dimerization might have in RNA-binding. Furthermore, we have studied the dynamical properties of the KH fold in the absence and in the presence of RNA to understand whether a "conformational change" occurs in the protein on RNA binding. Comparison of our results with data on other KH domains leads to new working hypotheses about the RNA-binding characteristics of KH domains.

MATERIALS AND METHODS

RNA and Protein Preparation. Three RNA oligonucleotides were synthesized in vitro using T7 polymerase and synthetic DNA templates using standard techniques (*17*). The KH construct, which spans amino acids 421−496 of the Nova-1 sequence, was expressed in BL21 pLysS cells as a His-tagged protein and isolated using a two-step standard purification procedure. After the cells were broken, the soluble fraction was loaded onto a nickel-affinity column. The protein was eluted in 0.3 M imidazole, concentrated, and loaded on a gel filtration column (G75 High Load 16/60 Pharmacia). The final protein solution was concentrated and dialyzed against 10 mM potassium phosphate buffer at pH 6.8 and 150 mM NaCl (for most of the measurements). The ionic strength effect on dimerization was checked using 10 mM potassium phosphate buffer at pH 6.8 and 10 mM NaCl. ^{15}N and ^{15}N, ^{13}C labeled protein was obtained by growing the cells in M9 minimal media supplemented with ^{15}N labeled amonium sulfate and ^{13}C labeled glucose. The protein molecular weight was confirmed by electrospray mass spectrometry. The calculated molar extintion coefficient of the construct is 2560 M^{-1} cm^{-1} (DNAStar).

Gel Filtration Assays. Gel filtration assays were performed on a Superdex 75 HR 10/30 Pharmacia column run at 0.5 mL/min using a Pharmacia AKTA purifier system. The column was calibrated using ovalbumin, chemotrypsinogen A, and RNAseA from the Pharmacia low molecular weight gel filtration calibration kit. These proteins provided the slope of the partition coefficient versus the logarithm of the molecular weight curve. The partition coefficient value was obtained from the elution volume of the species (*18*), assuming the elution volume of the dextran blue to be equal to the void volume of the column. The logarithm of the molecular weight for the first three proteins was obtained from the values reported on the calibration kit. The estimated error (R2) was evaluated to be 0.94.

Nuclear Magnetic Resonance Spectroscopy. NMR spectra were recorded at 25 °C on Varian INOVA-600 and Unity-600 spectrometers operating at 600 MHz ^{1}H frequency. The concentration of the free protein samples used for resonance assignment was 0.9 mM. Samples at different concentrations (0.04, 0.08, 0.32, 0.75, 1.3, and 2.6 mM) were used to follow the concentration-dependent chemical shift perturbation. Protein/RNA complex formation was studied by titrating a 0.3 mM sample of Nova-1 KH3 with RNA1−3. 2D ^{1}H NOESY and TOCSY spectra were acquired with mixing times of 50 and 150 ms for the NOESY and 40 and 70 ms for the TOCSY spectra. 3D NOESY- and TOCSY-HSQC spectra were recorded with mixing times of 150 and 70 ms, respectively (*19*, *20*). The dimensions of the respective matrixes were 1024 × 150 × 64 and 1024 × 128 × 32. HNCA and CBCANH (*21*, *22*) spectra were recorded using 42 and 48 increments in the ^{13}C dimension and 32 and 45 increments in the ^{15}N dimension. Water suppression was achieved by the WATERGATE pulse-sequence (*23*, *24*). The spectra were processed using the NMRPIPE program (*25*). Zero-filling was performed at least to the next power of two. Baseline correction was applied when necessary. The spectra were analyzed using the Felix (MSI) and XEASY programs (*26*).

Resonances Assignment. 95% percent of the HN, N, Cα, and Hα resonances were assigned using 2D TOCSY and NOESY experiments and their ^{15}N edited 3D versions (for the α-helix and β-sheet regions) as well as HNCA and CBCANH experiments (for the loop regions). Only the amide resonances of residues G443, L447, and T480 could not be assigned, although a tentative assignment is possible for residue 443. Partial assignment of the side chain resonances was also achieved.

Relaxation Data. Relaxation data were recorded at 500 MHz on a Varian Unityplus-500 spectrometer using standard sequences (*27*). The sample concentrations used were 0.08 and 2.6 mM for Nova-1 KH3 protein and 0.9 mM for the Nova-1 KH3/RNA1 complex. The acquisition time of the dilute sample was approximately 48 h. Peak intensities as a function of delay time were extracted from the spectra and normalized to the intensity of the first time point using the NMRPIPE/NMRDRAW package. The values of T_1 and T_2 were then determined by least-squares fitting to a single-exponential decay for each peak. The experimental errors were estimated by Monte Carlo sampling of the data-set and are reported in the additional material. The heteronuclear ^{15}N-^{1}H NOE values were obtained by recording interleaved ^{15}N-^{1}H correlation experiments, extracting the values of peak intensities and calculating the ratio of these intensities with and without presaturation. Correlation times were calculated from the T_1/T_2 ratios according to the so-called model-free approach (*28*−*30*). The values obtained for the individual residues were averaged in the well-defined secondary structure regions.

Analytical Ultracentrifugation. Sedimentation equilibrium experiments were carried out using a Beckman XL-A analytical ultracentrifugation equipped with UV absorption optics. Protein and RNA concentrations were varied in the different experiments but were chosen so that the absorbance at 260 nm was in the 0.2−1 absorbance units range (pathlength 1.2 cm). Data were recorded at 18 000, 24 000, 29 000, and 39 000 rpm and at ca. 15-h intervals to determine the protein molecular weight. In all datasets, the absorbance of the depleted area at a final speed of 39 000 rpm provided an experimental value for the baseline. 80 and 200 μM samples were studied for the free protein. Data from the complex were recorded using 60 μM protein and 3 μM RNA concentrations. A two-component self-association mechanism was assumed to allow for dimerization. The data were analyzed with the Origin XL-A/XL1 package (Beckman).

RESULTS

Characterization of Nova-1 KH3 Monomer/Dimer Equilibrium in Solution. The Nova-1 KH3 used in this study, which spans amino acids 421−496 of the full-length protein, is directly comparable to the construct used for crystal structure determination of the isolated Nova-1 KH (*7*). Although this is 15 amino acids shorter than the construct used for structure determination of the complex (*16*), the additional sequence is neither involved in protein/protein interaction nor in contacts with the specific UCAY RNA sequence. A combined approach using gel filtration, analytical ultracentrifugation, and NMR was used to determine the oligomeric state of Nova-1 KH3 in solution.

The apparent molecular weight of KH3 as estimated by gel filtration at a concentration of 0.6 mM of the loaded protein is 14.5 kDa, a value intermediate between the molecular mass of the protein monomer (9262 Da) and that of the dimer (18 524 Da). This intermediate value cannot be explained by an elongated shape of the protein, since the monomer is approximately spherical (*7*). The presence of a nonnegligible dimeric population or nonspecific aggregation are more likely under the experimental conditions (150 mM NaCl) than hydrophobic interaction with the gel filtration matrix: isolated protein domains are often subject to nonspecific aggregation phenomena (*31*).

A more direct determination of the molecular weight of the free protein was obtained by equilibrium centrifugation absorbance. Relatively high protein concentrations could be used because of the low absorbance of Nova-1 KH3. The measured molecular masses at 0.08 mM and at 0.2 mM protein concentrations (assuming a monomeric species) were 11.5 and 12 kDa, respectively (Figure 2). These values differ significantly from the calculated molecular weight, confirming that the protein is not a monomer. The experimental data could be fitted under the assumption of a monomer/dimer equilibrium with the presence of ca. 10−20% dimeric form, consistent with a K_d in the millimolar range.

The effect on the chemical environment of the monomer/dimer equilibrium was further explored by NMR. This technique operates in the millimolar concentration range (thus in the same order of magnitude of the estimated K_d) and allows us to explore a range of concentrations wider than absorbance AU. We used it to discriminate between specific and nonspecific self-association. Assignment of the backbone

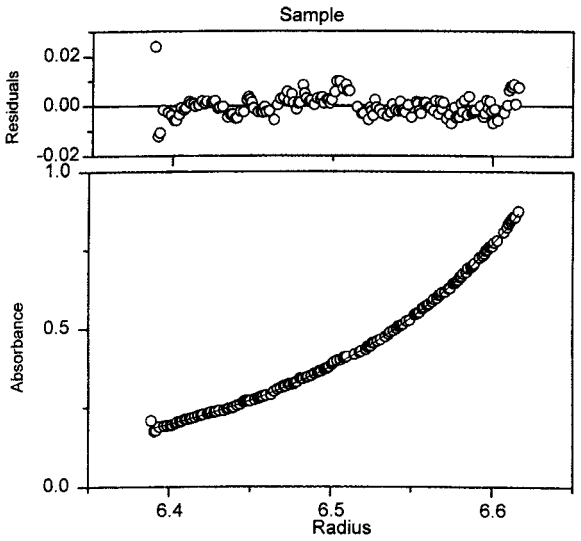

FIGURE 2: Sedimentation equilibrium data of Nova-1 KH3 in the analytical ultracentrifuge at 29 000 rpm and 15 °C. The data are fitted assuming a self-association equilibrium.

resonances of the construct was performed, as prerequisite for any further interpretation. Analysis of several independent NMR observables (i.e., secondary chemical shifts, NOE connectivities) confirmed that the construct has the expected fold. Evidence of a monomer/dimer equilibrium was confirmed by an appreciable concentration dependence of the rotational correlation times, well above that expected for the increase of viscosity caused by the higher protein concentration (τ_c values of 6.6 and 8.8 ns were obtained for 0.08 and 2.6 mM concentrations, respectively).

The surface involved in dimer formation was observed directly by concentration-dependent chemical shift perturbation. ^{15}N-^{1}H HSQC spectra were recorded in the 0.04−2.6 mM concentration range. The effect of high and low salt concentrations was also checked (see Materials and Methods). Concentration, but not salt-sensitive shifts were observed. The resonances affected were mapped onto the crystal structure of the protein (Figure 3). They cluster in a well-defined area, which must therefore correspond to a specific dimerization interface, ruling out the possibility of nonspecific aggregation or tetramerization. Interestingly, this surface involves residues in the last turn of α2, in the first two turns and in the last turn of α3 and is consistent with the dimerization interface common to the X-ray structure of free and RNA-complexed protein (*7*, *16*).

We conclude that, in the range of concentration explored, the free Nova-1 KH3 is in a monomer/dimer equilibrium with a binding constant in the submillimolar range. Dimer formation is specific, and the interface is directly described by our data.

Choice and Characterization of Oligonucleotides that Mimic Nova in Vivo Partners. To study the properties of the RNA/KH3 complex in solution, three RNA oligonucleotides (RNA1−3, Figure 4) with and without the specific UCAU target sequence were produced. They have different expected secondary structures. RNA1 contains two UCAU sequences intercalated by an eight-nucleotide spacer, and is designed to be single-stranded in solution to mimic the postulated in vivo target (*13*). RNA2 and RNA3 were selected for comparison: RNA2 includes the sequence of a

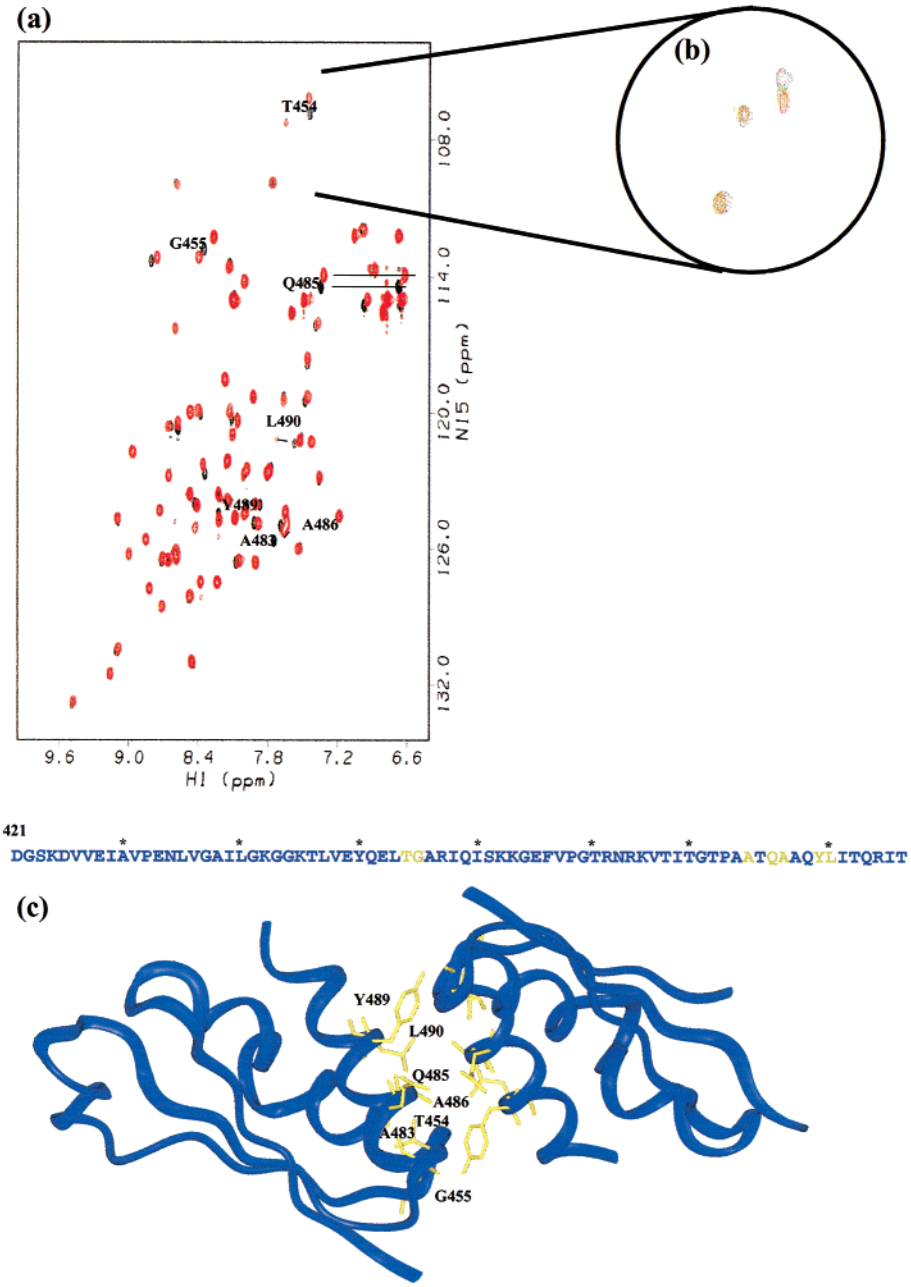

FIGURE 3: Effect of protein concentration on monomer/dimer equilibrium of Nova-1 KH3. (a) Superposition of the HSQC spectra of Nova-1 KH3 recorded at 0.08 mM and 1.3 mM concentration. The spectra were acquired at 600 MHz and 25 °C. (b) Enlargement of a region containing the resonance of T454. The spectra recorded at 0.08, 0.32, and 1.3 mM are superimposed. (c) Map of the effects of the concentration-dependent chemical shift perturbation experiments onto the sequence and the X-ray structure of Nova-1 KH3 (*7*). Residues involved in the dimerization interface are shown in yellow.

random pyrimidine-rich ssRNA. RNA3 contains the same sequence as RNA2 but in a sequence expected to form a hairpin as demonstrated in ref *32*. An exceptionally stable UUCG tetra-loop ensures that no dimerization will take place. 1D NMR spectra confirmed the secondary structures of the oligonucleotides: imino signals were observed for the expected base-pairs in RNA3, but not for RNA1 and RNA2, which should be single-stranded. The low dispersion of the resonances in RNA1 and RNA2 confirmed the absence of stable secondary structures. RNA1 was then studied by AU to determine its oligomeric state. A value of 6 kDa was obtained for its molecular mass in good agreement with the calculated molecular mass (6.3 kDa), indicating that RNA1 is a monomer in solution.

Stoichiometry of the Nova-1 KH3/RNA Complex in Solution. NMR spectroscopy was used to obtain the binding stoichiometry for the KH3/RNA1 complex and to estimate the dissociation constant. HSQC data recorded at different titration points of Nova-1 KH3 indicate that most protein resonances are in a moderately fast regime on a chemical shift time-scale and that the dissociation constant is in the low micromolar range (Figure 5). Chemical shift variations for a set of clearly distinguishable amide resonances were observed during titration up to but not beyond a 2:1 protein/RNA ratio, indicating that, at this ratio, all the protein is bound to RNA and that each RNA binds two proteins.

Support for a KH3 dimer/RNA complex was also obtained by AU (data not shown) and by ^{15}N relaxation data. The

— UUGCAGUCUCAUCA—UCAUUU——UCAUUUUGUUUA — **Gly Receptor mRNA**
— GUACAGUAUCAUCCUCUCAUCUACACCAUACUUUCAC — **Nova−1 mRNA**
— GUACAGUAUCAUCCUGUCAUCUACACCAUACUUUCAC — **Nova−2 mRNA**

(a)

GGUCAUCUCUUCUCUCAUU **RNA1**
GGUCCCUCCCACAC **RNA2**
GGAGCUUCGGCUCCACCCCUCCACAC **RNA3**

(b)

FIGURE 4: (a) In vivo target sequences of Nova-1 (*13*). These sequences include three UCAU repeats joined by short variable linkers. (b) Sequences of the RNA oligonucleotides used in our studies. The position of the expected stem-loop is indicated below the relative RNA sequence (RNA3).

correlation time obtained for the complex (10.8 ns) is higher than the one expected for a 1:1 protein/RNA complex, indicating that the RNA oligonucleotide is binding to the protein dimer. If the two KH3 domains were bound to the unfolded oligonucleotide without interacting with each other, we would expect their correlation time to be between the values observed for the protein monomer and the dimer.

Dynamical Studies of the Three Species of Nova-1 KH3 in Solution. Relaxation data were analyzed to extract information about the dynamical behavior of the KH fold during dimerization and complex formation (Figure 6). The longitudinal and transversal relaxation times (T_1 and T_2) and the NOEs are sensitive parameters that can provide a description of the molecular motions along the sequence (*28*, *30*). Comparison of the T_1 values of the ^{15}N backbone resonances for the free protein at two concentrations (0.08 and 2.6 ms) highlights substantial differences. A general increase of the average T_1 values is observed for the protein at high concentration, reflecting the higher molecular weight of the protein dimer, which becomes the predominant species in solution. However, while at high concentration (where the dimeric species is highly populated) there is very little variation of the T_1 values from the average, the values are much more scattered at low concentration. In particular, several resonances have T_1 significantly shorter than the average: the T_1 values of E433, V436, I439, G441, G444, I458, T470, R473, V475, A486, A487, and I491 are shorter than 400 ms. Most of these residues are conserved and structurally important. A486, A487, and I491 are part of α3, on the dimerization interface. Their side chains, however, are not directly involved in dimerization since these residues are in the interface between α1, α2, and α3. The other nine residues define the packing of α1 and α2 against the β-sheet. The anomalous T_1 values cannot easily be attributed to anisotropy since they differ from other ^{15}N backbone resonances that are also part of the same secondary structure elements. Protein dimerization seems therefore to lead to a reduction of high frequency motions in residues involved in the packing of secondary structure elements in the RNA binding interface. This in turn directly correlates with an increase of the order parameter and a decrease of the entropy of these residues.

Larger variations are observed for T_2 values at both concentrations. No heteronuclear NOE data are available at low concentration because of signal-to-noise limitations. At high concentration, the residues with low values of heteronuclear NOE do not correspond to residues with low T_2 values. This indicates that the low T_2 values observed for

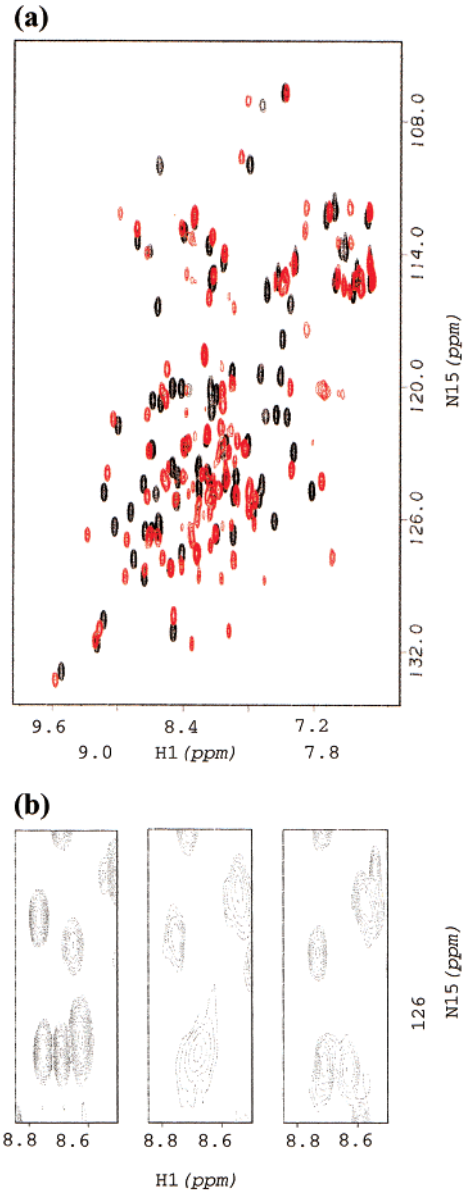

FIGURE 5: Nova-1 KH3 titration with RNA1 as followed by HSQC. The data were recorded at 600 MHz and 25 °C. (a) Superposition of the data recorded without RNA (in black) and with an excess of RNA. (b) Enlargement of a representative region and comparison of the spectra at three different protein:RNA ratios. From left to right are reported the spectra without RNA, with a 1:6 ratio, and with a 1:1 ratio (excess of RNA).

some of the protein resonances are due to chemical exchange rather than fast internal motions.

The relaxation parameters of the high concentration data (which correspond to predominance of the protein dimer) were then compared with those of the Nova-1 KH3/RNA complex. The T_1 values of the complex cluster around 600 ms, in agreement with the further increase of the molecular weight. Contrary to what was observed for the monomer to dimer transition, the distribution of T_1 values does not vary significantly on binding to RNA and remains rather flat. Large variations of T_2 values are instead found in the protein/RNA interface, i.e., loop2 and loop4, and, to a lesser extent, α2 and parts of β2 and β3. NOE values are more disperse and difficult to interpret. However, the difference between the NOEs measured in the complex and the ones measured for the free protein is roughly constant along the sequence

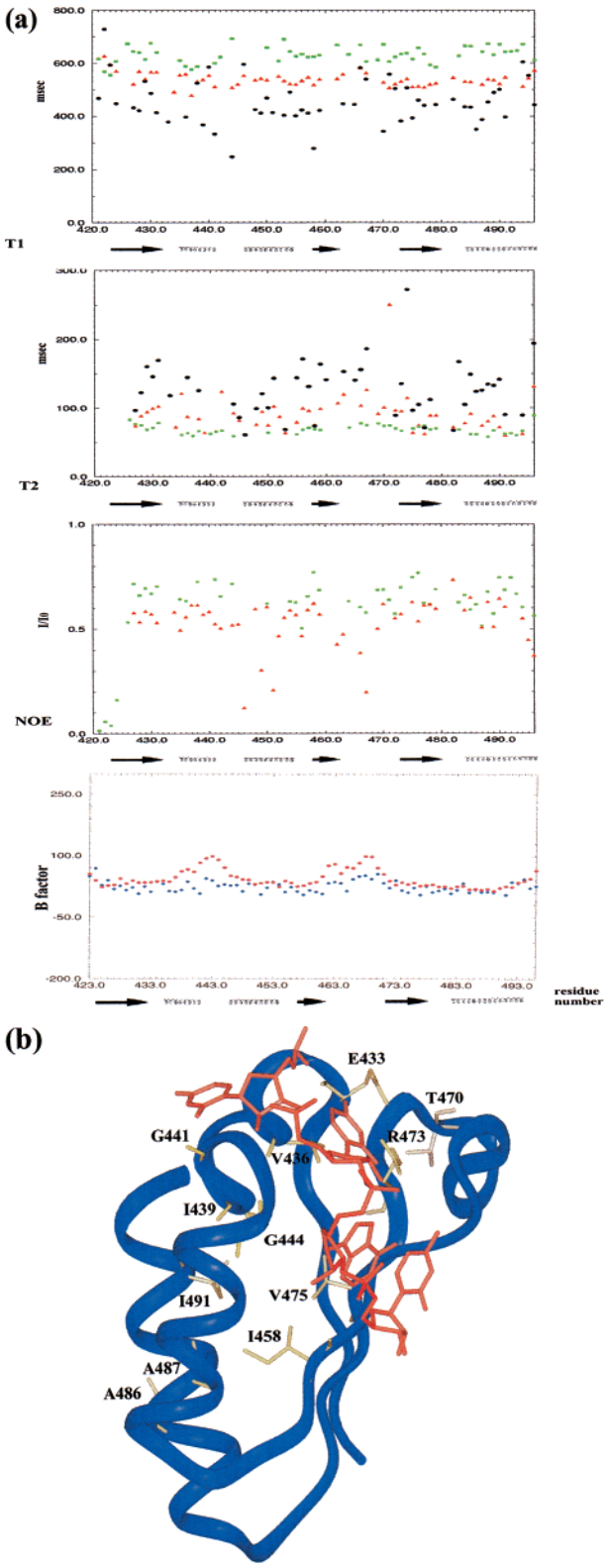

FIGURE 6: (a) From top to bottom, comparison of the T_1, T_2, NOE, and B factor values of the free Nova-1 KH3 and its RNA complex. Black and red are used to indicate T_1, T_2, and NOE data obtained for the low and high concentration of the free protein (no NOE values are available for the low concentration because of signal-to-noise limitations). The T_1, T_2, and NOE data of the complex are indicated in green. The B factors of the crystallographic structures of Nova-2 KH isolated and in the complex are reported in red and blue respectively (bottom panel). (b) The side chains of the residues with T_1 values smaller than 400 ms are mapped onto the KH3 structure (in yellow) (*7*, *16*). The U12−C15 oligonucleotide is shown in red.

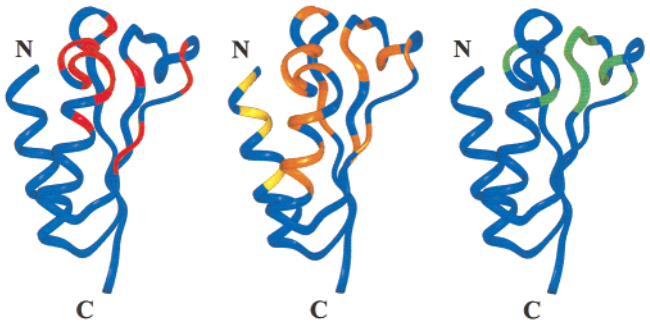

FIGURE 7: Comparison of the residues involved in protein/RNA interactions as observed in the X-ray structure of the complex (left) (*16*) with those indicated by chemical shift perturbation, by RNA titration (middle), and by cross-saturation experiments (right) (*34*). The residues directly in contact with RNA in the X-ray structure are shown in red, while in green are shown the ones affected in cross-saturation experiments. The results from chemical shift perturbation include, in addition to residues identified by X-ray and cross-saturation experiments (in orange), a further patch on α3 (in yellow). These include residues involved in dimerization according to our study. A similar effect was observed by Baber et al. (*39*).

with the exception of two regions: loop2 and the beginning of loop4. This difference is due to unusually low values of the NOEs in the free protein, while the values in the complex are close to the average. The T_2 and NOE data are consistent with the presence of high-frequency motions in α1-loop2 and loop4 in the free protein, which disappear on RNA complexation.

A more complete analysis of the motions present in the protein monomer, dimer, and in the complex using the so-called model-free approach was not attempted, because of the absence of NOE data for the low concentration free protein and because the monomer/dimer equilibrium implies that, even at high concentration, a fraction of the protein can be in a monomeric form. Further analysis is meaningless under these conditions (*29*, *33*). However, two major effects can be clearly observed from the raw data. First, dimer formation leads to a general stiffening of the protein and in particular of the groove where RNA binds. Second, high frequency motions are lost in relatively flexible regions upon interaction with RNA.

Characterization of RNA/Protein Interface in Solution. The protein/RNA binding surface in solution was determined by two independent methods: chemical shift perturbation and a novel application based on cross-saturation experiments (*34*, *35*). The former method detects the effect on the chemical shifts of the molecular environment change due to complexation of one of the two components, whereas the latter describes the effects that irradiation of a molecule has on spatially close regions of a second molecule. We have recently demonstrated that a combination of the two methods provides a powerful tool for the study of protein/RNA complex interfaces (*34*). In particular, the two methods must be considered as complementary and are therefore very useful for distinguishing direct from indirect effects of ligand binding.

The effects observed upon addition of RNA1 to Nova-1 KH3 were mapped onto the structure of the protein/RNA complex (Figure 7). They cluster in two regions of the protein. The first region (which shows larger effects) comprises α1 and α2, loop2 and 4 and part of β3. These

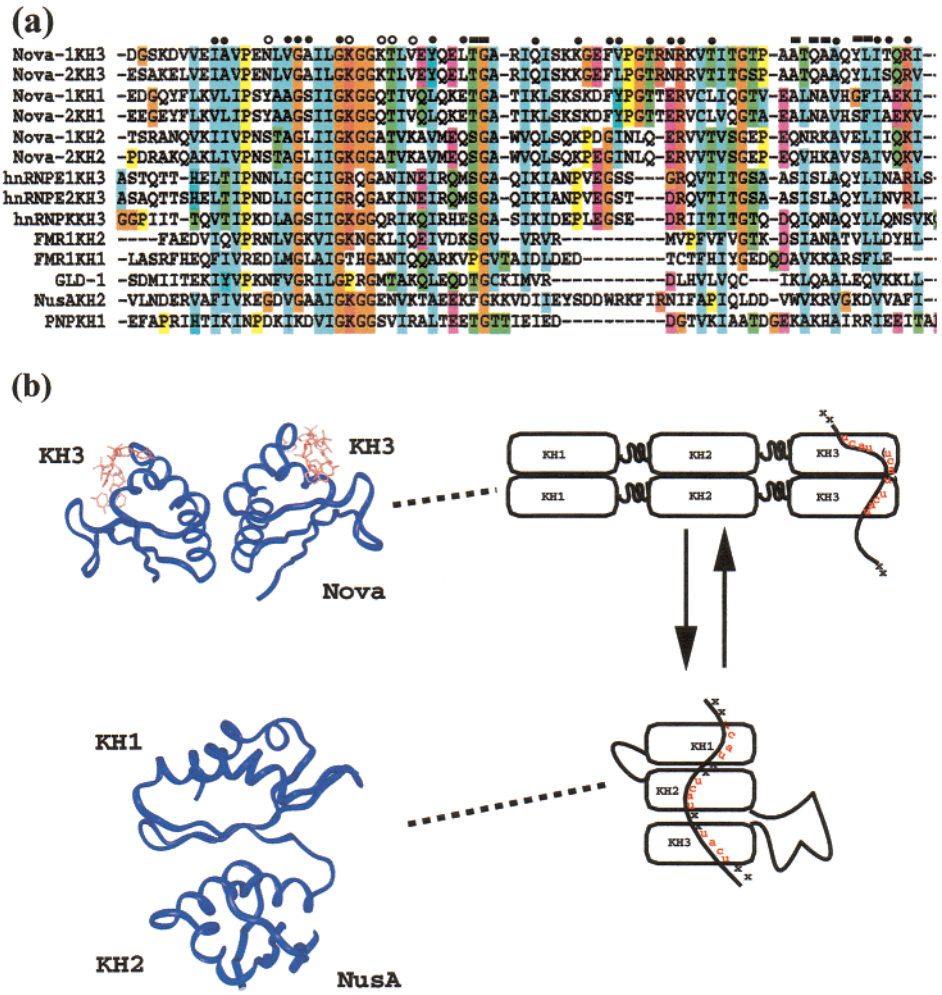

FIGURE 8: (a) Sequence alignment of Nova-1 and other representative KH domains displayed with Clustal X colors to emphasize conserved sequence features (*40*). On the top, a dash indicates the residues implicated in protein dimerization. The residues for which the resonances move (filled circles) or disappear (empty circles) during titration with RNA. (b) Possible model of KH/KH and KH/RNA interactions in the cell. We suggest that formation of the Nova-1/RNA complex could be based on an equilibrium between the protein homodimer (2:1 protein:RNA complex) and the protein monomer (1:1 protein:RNA complex). The arrangement of the protein domains involved would be consistent with the interfaces observed in the crystallographic structures of Nova-2 KH3 (top left) and NusA (bottom left) (*10*, *16*).

stretches form a narrow groove in the protein structure, which can easily accommodate an oligonucleotide chain. These solution data are in close agreement with the X-ray structure of the Nova-1/RNA complex (*16*). Smaller but clearly detectable chemical shift changes involve residues A487, I491, T492, and R494 on α3. These residues are either directly involved in protein dimer formation or close to this interface. Titration of Nova-1 KH3 with RNA2 and RNA3 shows a similar but smaller effect. This suggests that the surface of interaction remains the same for specific and nonspecific binding.

DISCUSSION

We have used a combination of biophysical and biochemical techniques to study the self-association of KH modules and its role in RNA-binding using the Nova-1 KH3 as a model system. We provide conclusive evidence that Nova-1 KH3 homodimerizes in solution even in the absence of RNA and without the contribution of other regions of the full-length protein. Dimerization occurs through a specific interface, with a dissociation constant in the submillimolar range and involves residues conserved in Nova-1 KH1 and KH3 and to a large extent also in KH2 (Figure 8a). In vivo, the self-association of Nova could therefore involve both KH1 and KH3, thus cooperatively increasing the dimerization affinity. The residues implicated in Nova dimerization are not conserved within the entire KH family but only within a subfamily that comprises hnRNP-K protein and shares with Nova the general architecture and a relatively higher sequence homology. Self-association has indeed been described for hnRNP-K and E2 proteins and has been suggested to mediate their biological function either by interacting with other effector proteins or by determining RNA-binding specificity (*4*). We therefore suggest that our findings can be extended to the whole subfamily. The presence of the two close Nova-1 and Nova-2 homologues is also suggestive of a possible modulation of Nova functions through homo- and heterodimerization of the proteins. Several of the KH-containing proteins have close homologues with which they are known to heterodimerize, so it might be tempting to suggest that the RNA-binding specificity of many members of the family could be regulated by homo- and heterodimerization. Such a mechanism is already known for DNA-binding proteins such as the b-LZ leucine zipper (e.g., Jun/Fos) and the HLH families (*36*): a similar strategy to

modulate nucleic acid specificity might be common to DNA and RNA binding proteins.

The relatively low affinity of the dimer formation in our model system allows us to observe a "snapshot" of each of the three components: the protein monomer, the dimer, and the RNA complex. It is worth stressing that in vivo dimerization and RNA binding will almost certainly be modulated by the presence of the other KH domains, increasing the affinity and the cooperativity of the binding and changing the kinetics. This would explain why most of the proteins of the KH family are either assembled from several tandem KH modules or from different RNA-binding motifs. It is thanks to the possibility of working with the isolated module that we are able to observe and characterize individually the three states. Our relaxation data indicate that two distinct phenomena occur, corresponding to the protein/protein and protein/RNA interaction processes. First, we observe a general stiffening of the protein upon protein dimerization. Stiffening of noninteracting regions of a protein upon interaction with nucleic acids has been proposed to be functional toward reducing the entropic cost associated with the interaction in the tRNA/aaRS (*37*) and GCN4/DNA (*38*) complexes. Second, high frequency motions still observable in the loops directly involved in RNA-binding are lost upon interaction with RNA. This implies a rigidification of the protein/RNA-binding surface mediated by protein dimerization and agrees with the reduction of the thermal B-factor observed in the crystal structure of the complex (Figure 6a). Dimerization has therefore two functional effects: presenting two recognition sites and facilitating the binding of each protein monomer to the RNA.

KH3 forms in solution a 2:1 complex with RNA1, an oligonucleotide that contains two copies of the UCAU motif. The surface of RNA recognition observed in solution is the same as in the crystal structure of the complex (*16*). The whole KH family seems to adopt the same protein surface for RNA recognition: RNA-binding studies conducted for hnRNP-K KH3 and for FMR1 KH1 show involvement of the same interface (*39* and Ramos and Pastore, unpublished results). This is supported by the strong sequence conservation of the exposed GkxG motif in loop2 and of a number of positive charges spatially contiguous to it (ca. 80% of the conserved positive charges are located in or around the experimentally determined RNA-binding surface of Nova KH1 and hnRNP-K KH3). As for the protein/protein interface in the dimer, the observed dimerization surface, together with the stoichiometry of the complex, are in agreement with the hypothesis that, in vivo, Nova binds RNA through dimerization of the KH3 domain (*16*). However, the presence of a third UCAU repeat conserved in the in vivo targets strongly argues in favor of at least two alternative modes of binding, possibly in mutual equilibrium (Figure 8b). This possibility is consistent with the different packing modes observed in the crystal structures of Nova-2 and NusA (*10*, *16*). Both modes could be represented in the Nova/RNA complex: a 2:1 protein homodimer/RNA complex in which the packing between KHs occurs through the interface observed in solution and in the Nova X-ray structure could be in equilibrium with a 1:1 protein/RNA complex. In this second complex, the packing arrangement of the three KH domains could be that observed in the NusA structure of NusA with two KH domains from the same molecule packed asymmetrically. Enthalpic and entropic contributions may favor one or the other interaction, which will be strongly modulated by the relative protein/RNA concentrations. The work described here exemplifies how low resolution structural information can be used in conjunction with biochemical data to elucidate aspects of the mechanism and the determinants of specificity of regulatory systems. Further studies will be needed to clarify the role of heterodimerization and of cooperativity in RNA/KH binding.

ACKNOWLEDGMENT

We are indebted to Andrew Lane whose helpful discussions have provided constant inspiration for the present work. We would also like to thank R. Andrew Atkinson and Franca Fraternali for their help with some of the programs used, Yunghan Au for technical assistence, Giovanna Musco for helpful discussions, and David J. Thomas for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

T_1, T_2, and NOEs for Nova-1 KH1. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Perez-Canadillas, J. M., and Varani G. (2001) Recent advances in RNA-protein recognition. *Curr. Opin. Struct. Biol. 11*, 53−58.
2. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res. 21*, 1193−1198.
3. Chen, T., Damaj, B. B., Herrera, C., Lasko, P., and Richard, S. (1997) Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. *Mol. Cell. Biol. 17*, 5707−5718.
4. Kim, J. H., Hahm, B., Kim, Y. K., Choi, M., and Jang, S. K. (2000) Protein−protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol. 298*, 395−405.
5. Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T. J., Nilges, M., and Pastore, A. (1997) The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat. Struct. Biol. 4*, 712−716.
6. Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M. A., Nilges, M., Gibson, T. J., and Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell 85*, 237−245.
7. Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B., and Burley, S. K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Struct., Folding Des. 7*, 191−203.
8. Baber, J. L., Libutti, D., Levens, D., and Tjandra, N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol. 289*, 949−962.
9. Grishin, N. V. (2001) KH domain: one motif, two folds. *Nucleic Acids Res. 29*, 638−643.
10. Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R., and Wahl, M. C. (2001) An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol. Cell 7*, 1177−1189.
11. Gibson, T. J., Thompson, J. D., and Heringa, J. (1993) 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 Lett. 324*, 361−366.
12. Buckanovich, R. J., Posner, J. B., and Darnell, R. B. (1993) Nova, the paraneoplastic Ri antigen, is homologous to an

RNA-binding protein and is specifically expressed in the developing motor system. *Neuron 11*, 657−672.
13. Buckanovich, R. J., and Darnell, R. B. (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol. Cell. Biol. 17*, 3194−3201.
14. Jensen, K. B., Dredge, B. K., Stefani, G., Zhong, R., Buckanovich, R. J., Okano, H. J., Yang, Y. Y., and Darnell, R. B. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron 25*, 359−371.
15. Yang, Y. Y., Yin, G. L., and Darnell, R. B. (1998) The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl. Acad. Sci U.S.A. 95*, 13254−13259.
16. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell 100*, 323−332.
17. Price, S. R., Ito, N., Oubridge, C., Avis, J. M., and Nagai, K. (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol. 249*, 398−408.
18. Cantor, R. C., and Shimmel, R. P. (1980) *Biophysical Chemistry*, W. H. Freeman and Co., New York.
19. Marion, D., Driscoll P. C., Kay L. E., Wingfield P. T., Bax A., Gronenborn A. M., and Clore G. M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann−Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry 28*, 6150−6156.
20. Zuiderweg, E. R., and Fesik, S. W. (1989) Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry 28*, 2387−2391.
21. Grzesiek, S., and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson. 99*, 201−207.
22. Grzesiek, S., and Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc. 114*, 6291−6293.
23. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR 2*, 661−665.
24. Sklenar, V., Peterson, R. D., Rejante, M. R., and Feigon, J. (1993) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in 15N, 13C-labeled RNA oligonucleotides. *J. Biomol. NMR 3*, 721−727.
25. Delaglio, F., Grzesiek, S., Vuister, G. V., Zhu, G., Pfeifer, J., and Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.
26. Bartels, C., Xia, T. H., Billeter, M., Gunter, P., and Wuthrich, K. (1995) The program XEASY for computer-supported NMR spectral-analysis of biological macromolecules. *J. Biomol. NMR 6*, 1−10.
27. Kay, L. E. Nicholson, L. K., Delaglio, F., Bax, A., and Torchia D. A. (1992) Pulse sequences for removal of the effects of cross correlation between dipolar and chemical shift anisotropy relaxation mechanisms on the measureament of heteronuclear T1 and T2 values in proteins. *J. Magn. Reson. 97*, 359−375.
28. Lipari, G., and Szabo G. (1982) A model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc. 104*, 4546−4559.
29. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol. 246*, 144−163.
30. Palmer, A. G., Rance, M., and Wright, P. E. (1991) Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterised by proton-detected natural abundance C-12 heteronuclear NMR-spectroscopy. *J. Am. Chem. Soc. 113*, 4371−4380.
31. Pfuhl, M., Chen, H. A., Kristensen, S. M., and Driscoll, P. C. (1999) NMR exchange broadening arising from specific low affinity protein self-association: analysis of nitrogen-15 nuclear relaxation for rat CD2 domain 1. *J. Biomol. NMR 14*, 307−320.
32. Allain, F. H., and Varani, G. (1995) Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol. 250*, 333−353.
33. Mittermaier, A., Varani, L., Muhandiram, D. R., Kay, L. E., and Varani, G. (1999) Changes in side chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol. 294*, 967−979.
34. Ramos, A., Kelly, G., Hollingworth, D., Pastore, A., and Frenkiel, T. (2000) Mapping the interfaces of protein-nucleic acid complexes using cross-saturation. *J. Am. Chem. Soc. 122*, 11311−11314.
35. Takahashi, H., Nakanishi, T., Kami, K., Arata, Y., and Shimada, I. (2000) A novel NMR method for determining the interfaces of large protein−protein complexes. *Nat. Struct. Biol. 7*, 220−223.
36. Branden, C. Y., and Tooze, J. (1999) *Introduction to Protein Structure*, Garland, New York.
37. Ribas de Pouplana, L., Auld, D. S., Kim, S., and Schimmel, P. (1996) A mechanism for reducing entropic cost of induced fit in protein-RNA recognition. *Biochemistry 35*, 8095−8102.
38. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G. (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol. 285*, 2133-2146.
39. Baber, J. L., Levens, D., Libutti, D., and Tjandra, N. (2000) Chemical shift mapped DNA-binding sites and 15N relaxation analysis of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K. *Biochemistry 39*, 6022−6032.
40. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25*, 4876−4882.

BI011994O