J Biomol NMR (2010) 47:79–83
DOI 10.1007/s10858-010-9412-1

NMR STRUCTURE NOTE

Solution structure of the RBD1,2 domains from human nucleolin

Sengodagounder Arumugam · M. Clarke Miller · James Maliekal · Paula J. Bates · John O. Trent · Andrew N. Lane

Received: 28 December 2009 / Accepted: 22 March 2010 / Published online: 8 April 2010

Biological context

Nucleolin is an abundant 77 kDa RNA binding protein found in the nucleolus of all eukaryotic cells, and is involved in rRNA packaging and transport (Tuteja and Tuteja 1998; Ginisty et al. 1999). The protein comprises several domains, including an acidic N-terminal domain, four RNA binding (RBD) domains, and a C-terminal RGG-rich "tail". Although primarily nucleolar, the protein is also found in the cytoplasm, and uniquely in cancer cells on the cell surface. The latter location is important as nucleolin acts as a receptor for the exogenous G-quartet DNA aptamer AS1411 that is currently in phase I/II clinical trials for renal cell carcinoma and acute myelogenous leukemia (www.antisoma.com) (Miller et al. 2006). In this form, the protein binds the aptamer, whereupon it is internalized, and delivers the drug to the cell where it interferes with NF-κB signaling (Girvan et al. 2006) and induction of tumor suppressor gene expression (Bates et al. 2009). AS1411 binds to recombinant fragments of nucleolin consisting of the RNA binding domains and RGG region as evidenced by a western blot (P.J. Bates, unpublished data).

An NMR structure of the first two RBD domains from the 84% homologous hamster protein and its complex with RNA hairpins has been solved (Allain et al. 2000a, b; Finger et al. 2003; Johansson et al. 2004). However, in order to understand the mechanism of binding the DNA aptamer in detail, it is necessary to have complete assignments and structures of the human version. As the sequence identity is only 84%, we resorted to a complete assignment and structural analysis of the human protein as part of an investigation of its interaction with AS1411. We have obtained essentially complete backbone and approx 90% of the side-chain assignments of the RBD1,2 construct (174 residues), by 3D heteronuclear NMR, and we have solved the three dimensional structure to high precision. This construct shows high affinity for AS1411, and will form the basis for determining the binding site on the protein for this DNA molecule, and an understanding of how this protein acts as a DNA aptamer receptor.

S. Arumugam · M. Clarke Miller · J. Maliekal · P. J. Bates · J. O. Trent · A. N. Lane (✉)
JG Brown Cancer Center, University of Louisville, 505 South Hancock St., Louisville, KY 40202, USA
e-mail: anlane01@louisville.edu

Methods and results

Materials

The synthetic RBD1,2 gene was purchased from Integrated DNA Technologies. The gene was delivered as an insert in the pZErO-2 plasmid. 2 μg of the pZErO-2/RBD1,2 plasmid was reconstituted with 20 μL of nuclease-free water for a final concentration of 0.1 μg/μL. The plasmid was then introduced into the NovaBlue (Novagen) host strain for replication. The resulting strain was plated on LB agar plates with 30 μg/mL kanamycin and grown overnight. Colonies were selected and 5 mL cultures were grown overnight. The pZErO-2/RBD1,2 plasmid was then harvested via a commercially available mini-prep kit (Sigma). The isolated plasmid was subjected to restriction digest using NdeI and XhoI (New England Bioscience) and the resulting fragments were isolated on a 1% agarose gel. The fragment of approximately 500 bp was sliced out of the gel, purified, and then ligated into similarly prepared pET21a (Novagen). The resulting pET21a/RBD1,2 construct was then replicated in NovaBlue and harvested as

described above. The resulting purified plasmid was introduced into Rosetta 2 (DE3)(Novagen), plated on LB agar with 50 μg/mL ampicillin and 30 μg/mL chloramphenicol. Colonies were selected and checked for expression. Once expression was confirmed, 1.5 L cultures were grown in minimal media containing 6 g/L glucose (unlabeled or [U-^{13}C]) and 1.5 g/L ($^{15}NH_4)_2SO_4$.

The cultures were then spun down and the cell pellets digested with Cellytic B (Sigma). The resulting supernatant was filtered and loaded onto Ni cartridges (Sigma) which had been prepared as directed. The cartridges were then subjected to washes with 8, 6, 4, and 2 M urea before final wash and elution with imidizole. The protein was further purified by size exclusion chromatography using a Superdex 75 10/300 column (GE Healthcare) installed on an Äkta-FPLC with UPC-900 UV absorbance monitor and Frac920 fraction collector (GE Healthcare) and prepared with a mobile phase consisting of 100 mM KCl, 25 mM K_2HPO_4 (pH = 8.0) at a flow rate of 0.10 mL/min. Crude RBD1,2 was injected at approximately 10 mg/mL in up to 500 μl aliquots. 0.2 mL fractions were collected with monitoring the absorbance at 280 nm. Fractions corresponding to an estimated molecular weight of approximately 20 kDa were combined and dialyzed into 100 mM KCl, 20 mM acetate (pH = 5.0) and concentrated to approximately 10 mg/mL using Microcon YM-10 (Millipore) centrifugal concentrators. The resulting material was then re-dialyzed into 100 mM KCl, 20 mM deuterated acetate (pH = 5.0), and 3 mM NaN_3. The final NMR sample was then concentrated as described above to approximately 1 mM.

Table 1 Structure statistics for human RBD1,2

Number of cross peak assignments by CYANA	6,481	
Number of unique assignments	5,722	
Short range $\|i - j\| \leq 1$	4,329	
Medium range $1 \leq \|i - j\| < 5$	721	
Long range $\|i - j\| > 5$	1,431	
NMR structure determination statistics for RBD1,2		
Number of NOE derived upper distance restraints	3,793	
Short range $\|i - j\| \leq 1$	1,857	
Medium range $1 \leq \|i - j\| < 5$	620	
Long range $\|i - j\| > 5$	1,316	
Number of hydrogen bond restraints	60	
Number of dihedral angle restraints from TALOS+	274	
Total number of restraints	4,127	
Final structure statistics for RBD1,2		
Number of refined structures	20	
CYANA target function	5.61 ± 0.48	
Number of restraints violated in 6 or more structures:		
Distances	35	
Angle	1	
RMSD to the mean structure (Å):		
With out hydrogen bond restraints:	Backbone	Heavy atoms
RBD1,2 (6–167)	11.48 ± 3.55	11.97 ± 3.59
RBD1(6–81)	0.42 ± 0.08	1.11 ± 0.10
RBD2(93–167)	0.63 ± 0.15	1.30 ± 0.13
linker (82–92)	2.66 ± 1.09	3.98 ± 1.18
With hydrogen bond restraints:	Backbone	Heavy atoms
RBD1,2 (6–167)	9.89 ± 3.66	10.47 ± 3.80
RBD1(6–81)	0.32 ± 0.07	1.01 ± 0.24
RBD2(93–167)	0.71 ± 0.21	1.46 ± 0.24
Linker (82–92)	1.76 ± 0.75	3.09 ± 0.84
Procheck analysis		
Residues in most favored region	82.7%	
Residues in additional allowed regions	16.2%	
Residues in generously allowed region	0.1%	
Residues in disallowed region	0.9%	

NMR data collection and analysis

NMR spectra were recorded at 18.8 T on a Varian Inova spectrometer equipped with a 5 mm inverse triple resonance pfg probe. Additional spectra were recorded at 14.1 T using a cold probe. Spectra were processed using NMRPipe (Delaglio et al. 1995).

NMR assignments were obtained using a combination of 3D triple resonance experiments: HNCA, HN(CO)CA, CBCA(CO)NH, HNCO, HACACO, NOESY-^{15}NHSQC, NOESY-^{13}CHSQC, and HNCACB.

Structural restraints were obtained from ^{15}N and ^{13}C separated NOESY spectra recorded at different mixing times, along with torsion angles from chemical shift analysis using TALOS+ (Shen et al. 2009).

Structure calculations

99.7% of the backbone atoms have been assigned except for the N-terminal glycine and the nitrogen of the prolines. More than 89% of the side chain atoms have been assigned. The missing assignments are due to absence of stereo specific assignments and severe overlaps in the HCCH-TOCSY and ^{13}C-separated NOESY spectra. Amide nitrogen and proton resonances for T47 could not be assigned perhaps due to motional broadening. Also, one of the Hα protons of the G172 could not be assigned due to chemical shift degeneracy.

H–D exchange experiment was performed on the lyophilized RBD1,2 dissolved in 99.99% D_2O and amide peaks that remained after 16 h are presumed to be involved in hydrogen bonds.

Structure calculations were performed following automated NOESY assignment using CYANA (Guntert et al. 1997; Herrmann et al. 2002). In addition to the distance restraints arising from the automated NOESY assignments, torsion angles estimated using TALOS+ (Shen et al. 2009) and distance restraints from the possible hydrogen bonds based on the preliminary structure were included in the final structure calculations, as given in Table 1. Figures were created using MOLMOL (Koradi et al. 1996).

Discussion and conclusions

The NMR spectrum of the present protein construct is well dispersed (Fig. 1). The high number of assigned residues and structural restraints (Table 1) gave rise to well defined structures at both the backbone and side chain levels, as shown in Fig. 2 and Table 1. The secondary structure elements are essentially the same in both domains (Fig. 2a), and match those previously reported for the hamster protein (Allain et al. 2000a, b). The individual

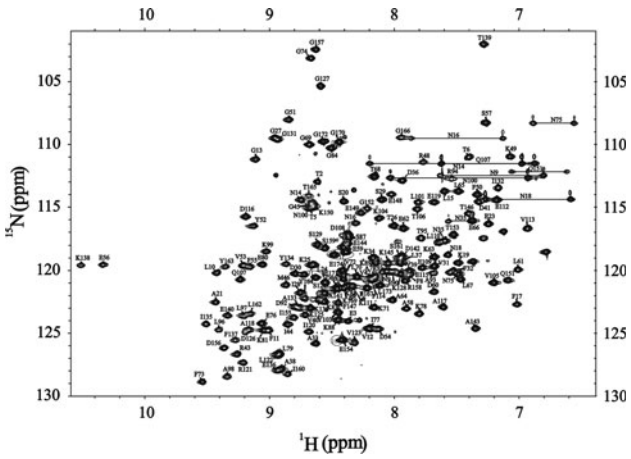

Fig. 1 ^{1}H-^{15}N HSQC spectrum of the RBD1,2 domain. The NMR spectrum was recorded at 18.8 T at 293 K with acquisition times of 0.15 s in t_2 and 0.021 s in t_1. The data were processed by zerofilling once in t_2, twice in t_1 and apodized using an unshifted Gaussian and a 1 Hz line broadening exponential in both dimensions

domains are very well determined by the data, especially that of domain 1 (Fig. 2b, Table 1). As expected, the domain structures are overall very similar, though with the higher restraint density, the human structure is better defined than the hamster orthologue (Allain et al. 2000a, b). This is especially so for the side chains, which except for some mobile surface residues, are well determined (cf. Table 1), which will be critical for analysis of future ligand binding experiments.

No interdomain NOEs or hydrogen bonds were observed; when domains 1 are superimposed, the relative orientation of domain 2 is not well determined (and vice versa) (Fig. 2b, Table 1). However, the linker region itself is relatively well determined according to superposition of this region (Table 1), implying that it is not a completely flexible, unstructured oligopeptide segment. Indeed, Ser87 to Asp92 adopt a helical conformation (Fig. 2c) according to their ϕ, ψ angles, though the torsions of the remaining residues lie well outside of the helical region. However, at least in the free protein, there is insufficient information to determine the relative orientation of the two domains, if any. Nevertheless, the amide ^{15}N line widths determined in high-resolution HSQC spectra were all in the range 6–8 Hz, which is larger than expected for a single domain of 10 kDa, suggesting that there is restricted motion of the two domains about the linker.

Similar structural statistics were reported for hamster RBD1,2 that was bound to a small piece of RNA (Allain et al. 2000b). The overlay of the hamster RBD1 (PDB ID:1FJ7) and RBD2 (PDB ID:1FJC) onto the human RBD1,2 are shown in Fig. 3. The rmsd for the backbone atoms to the individual domains of the isolated, RNA-free hamster structures are 1.72 and 2.16 Å respectively for the

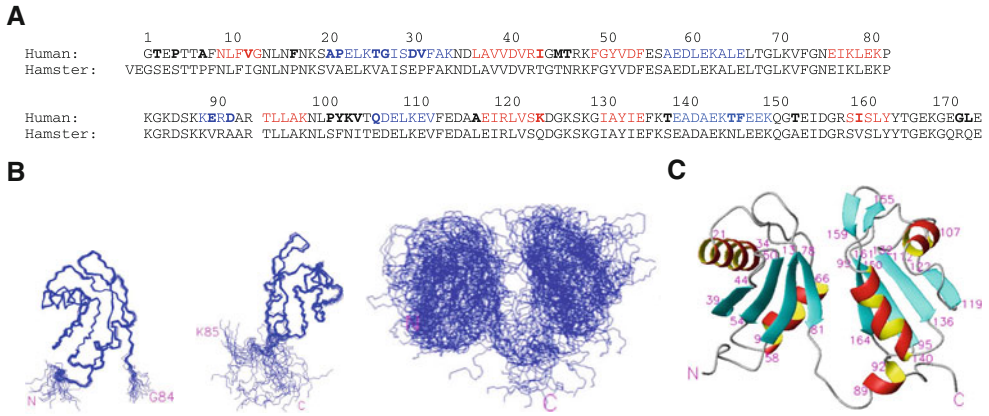

Fig. 2 Structures of nucleolin RBD1,2. **a** Sequence alignment of human with golden hamster nucleolin RBD1 and 2. Sequences were obtained from the European Bioinformatics Institute (http://www.uniprot.org/uniprot/). Numbering is according to the human construct. *Bold* denotes deviations between human and hamster sequence. *Upper* RBD1, *lower* linker, RBD2. Helices are shown in *blue* and β strands in *red*. Secondary structure elements were defined for the human sequence in Molmol (Koradi et al. 1996), and TALOS+ (Shen et al. 2009). **b** Ensemble of 20 backbone structures of RBD1,2 calculated with NOE, dihedral and hydrogen bond restraints. *Left* RBD1, *middle* RBD2, *right* RBD1,2. **c** RBD1,2 illustrated as a ribbon diagram with the lowest target function as calculated by CYANA

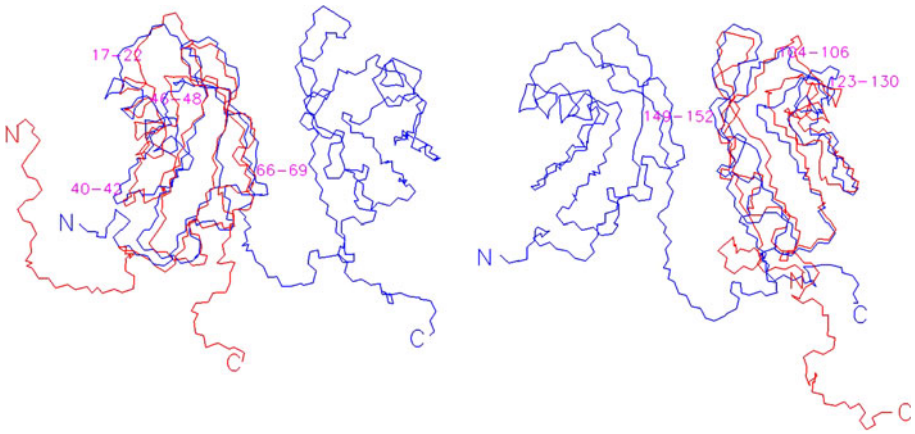

Fig. 3 Comparison of the human and hamster RBD. Backbone structures of the hamster domains (*red*) were superposed on the human domains (*blue*). *Left* superposition of RBD1, *right* superposition of RBD2

domains 1 and 2. Nevertheless, there are some significant differences especially in the loop regions comprising residues 17–22, 40–42, 46–48 and 66–69 for domain 1 and 104–106, 123–130, and 149–152 for domain 2. Many, but not all of these differences correlate with differences between the sequences. On removing these loop regions, the rmsd values became 1.47 and 1.79 for domains 1 and 2, respectively.

Deposition of assignments and structure coordinates

The coordinates of the 20 refined structures of RBD1,2 have been deposited in the Protein Data Bank PDB (entry 2KRR) and the chemical shift data have been deposited at the BMRB (entry 16646).

Acknowledgments The work was supported by NIH/NCI RO1 CA113735-01 (to JOT) and the Kentucky Challenge for Excellence. NMR spectra were recorded in the JG Brown Cancer Center NMR Facility supported by grants NIH Grant Number RR018733 from the National Center for Research Resources, NSF EPS-0132295 and the Brown Foundation.

References

Allain FHT, Bouvet P, Dieckmann T, Feigon J (2000a) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. EMBO J 19(24):6870–6881

Allain FHT, Gilbert DE, Bouvet P, Feigon J (2000b) Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. J Mol Biol 303(2):227–241

Bates PJ, Laber DA, Miller DM, Thomas SD, Trent JO (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. Exp Mol Pathol 86(3):151–164

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) Nmrpipe—a multidimensional spectral processing system based on Unix pipes. J Biomol NMR 6(3):277–293

Finger LD, Trantirek L, Johansson C, Feigon J (2003) Solution structures of stem-loop RNAs that bind to the two N-terminal

RNA-binding domains of nucleolin. Nucleic Acids Res 31(22): 6461–6472

Ginisty H, Sicard H, Roger B, Bouvet P (1999) Structure and functions of nucleolin. J Cell Sci 112(6):761–772

Girvan AC, Teng Y, Casson LK, Thomas SD, Juliger S, Ball MW, Klein JB, Pierce WM, Barve SS, Bates PJ (2006) AGRO100 inhibits activation of nuclear factor-kappa B (NF-kappa B) by forming a complex with NF-kappa B essential modulator (NEMO) and nucleolin. Mol Cancer Ther 5(7):1790–1799

Guntert P, Mumenthaler C, Wuthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273(1):283–298

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227

Johansson C, Finger LD, Trantirek L, Mueller TD, Kim S, Laird-Offringa IA, Feigon J (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. J Mol Biol 337(4):799–816

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:29–32

Miller DM, Laber DA, Bates PJ, Trent JO, Taft BS, Kloecker GH (2006) Extended phase I study of AS1411 in renal and non-small cell lung cancers. Ann Oncol 17:147–148

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44(4):213–223

Tuteja R, Tuteja N (1998) Nucleolin: a multifunctional major nucleolar phosphoprotein. Crit Rev Biochem Mol Biol 33(6): 407–436