Structural and Functional Characterization of the TgDRE Multidomain Protein, a DNA Repair Enzyme from *Toxoplasma gondii*†

Karine Frénal,‡ Isabelle Callebaut,§ Karine Wecker,‡ Ada Prochnicka-Chalufour,‡ Najoua Dendouga,|| Sophie Zinn-Justin,⊥ Muriel Delepierre,‡ Stanislas Tomavo,|| and Nicolas Wolff*,‡

Unité de Résonance Magnétique Nucléaire des Biomolécules, CNRS URA 2185, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France, Département de Biologie Structurale, IMPMP, CNRS UMR 7590, Université Paris 6 et Paris 7, Case 115, 4 Place Jussieu, 75252 Paris Cedex 05, France, Equipe de Parasitologie Moléculaire, Unité de Glycobiologie Structurale et Fonctionnelle, CNRS UMR 8576, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, France, and Laboratoire de Structure des Protéines, Direction des Sciences du Vivant—Département d'Ingénierie et d'Etudes des Protéines (DSV—DIEP), CEA-Saclay F-91191, Gif-sur-Yvette, France

Received September 27, 2005; Revised Manuscript Received February 1, 2006

ABSTRACT: The parasite *Toxoplasma gondii* expresses a 55 kDa protein or TgDRE that belongs to a novel family of proteins characterized by the presence of three domains, a human splicing factor 45-like motif (SF), a glycine-rich motif (G-patch), and a RNA recognition motif (RRM). The two latter domains are mainly known as RNA-binding domains, and their presence in TgDRE, whose partial DNA repair function was demonstrated, suggests that the protein could also be involved in the RNA metabolism. In this work, we characterized the structure and function of the different domains by using single or multidomain proteins to define their putative role. The SF45-like domain has a helical conformation and is involved in the oligomerization of the protein. The G-patch domain, mainly unstructured on its own as well as in the presence of the SF upstream and RRM downstream domains, is able to bind small RNA oligonucleotides. We also report the structure determination of the RRM domain from the NMR data. It adopts a classical $\beta\alpha\beta\beta\alpha\beta$ topology consisting of a four-stranded β sheet packed against two α helices but does not present the key residues for the RNA interaction. In contrast, our analysis shows that the RRM of TgDRE is not only unable to bind small RNA oligonucleotides but it also shares the protein−protein interaction characteristics with two unusual RRMs of the U2AF heterodimeric splicing factor. The presence of both RNA- and protein-binding domains seems to indicate that TgDRE could also be involved in RNA metabolism.

Toxoplasma gondii is an obligate intracellular protozoan parasite infecting a broad range of warm-blooded vertebrates and up to 30% of the human population worldwide. *T. gondii* is recognized as an important opportunistic pathogen parasite associated with AIDS and congenital birth defects (*1*). The parasite life cycle is complex and involves vegetative multiplication in a wide variety of intermediate hosts. In mammalian nonfeline hosts, the parasite is found in two haploid asexual forms, the rapidly replicating virulent tachyzoites and the slowly dividing quiescent encysted bradyzoites. In response to immune system attacks and to survive within infected hosts, the tachyzoites differentiate into encysted bradyzoites that persist during the lifetime of the infected hosts and are characterized by their resistance to chemotherapy. This stage conversion involves profound metabolic and morphological changes to adapt to the environmental variations and should implicate a coordinated expression of genes. Molecular cloning of genes that encode stage-specific enzymes is of importance in understanding the mechanism of stage conversion between tachyzoites and bradyzoites. In this context, an in vitro encystation system followed by a complementary cDNA subtractive method have been developed to identify genes that are exclusively or preferentially expressed in the bradyzoite stage (*2*). Among the 12 isolated genes, we focused our study on the one that encoded *T. gondii* DNA repair enzyme (TgDRE),[1] a protein that is homologous to the SPF45 human splicing factor overexpressed in a variety of cancers (*3*). TgDRE orthologues of unknown structure and function are present in the genome sequences of the related apicomplexa parasites *Plasmodium falciparum* and *Plasmodium yoelii* as well as in those of *Drosophila melanogaster* and *Caenorhabditis elegans* (*4*). In agreement with the high sequence similarity that TgDRE presents with the DNA repair/tolerance protein DRT111 of

† This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS), and the Région Ile de France.

* To whom correspondence should be addressed. E-mail: wolff@pasteur.fr. Telephone: (33) 145688872. Fax: (33) 145688929.

‡ Institut Pasteur.

§ Université Paris 6 et Paris 7.

|| Université des Sciences et Technologies de Lille.

⊥ DSV—DIEP.

[1] Abbreviations: TgDRE, *Toxoplasma gondii* DNA repair enzyme; SF45 or SF, splicing factor 45-like motif; G-patch or Gp, glycine-rich motif; RRM, RNA recognition motif; RNP, ribonucleoprotein; U2AF, U2snRNP auxiliary factor; UHM, U2AF homology motif; HSQC, heteronuclear single-quantum correlation; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; TROSY, tranverse relaxation-optimized spectroscopy; NOE, nuclear Overhauser effect.

10.1021/bi051948e CCC: $33.50
Published on Web 03/28/2006

Table 1: Domain Organization of TgDRE and Proteins Used in This Study

Name	Organization	Coding region (aa)
TgDRE	SF45 G-patch RRM	1 - 466
SF+Gp+RRM		195 - 466
Gp+RRM		284 - 466
SF		195- 261
Gp		284 - 343
RRM		356 - 466

Arabidopsis thaliana (*5*), it has been demonstrated that TgDRE cDNA can partially correct DNA damages induced by the mitomycin C in an *Escherichia coli* mutant strain lacking RuvC endonuclease and RecG helicase (*4*).

All of these homologous proteins share the same architecture with TgDRE and are therefore designated as the splicing factor 45-like motif (SF45) family (*4*). Bioinformatics studies have shown that they possess three domains localized in the C terminus: a SF45, a glycine-rich motif (G-patch), and a RNA recognition motif (RRM). The presence of these putative RNA-binding domains suggests that TgDRE might also be involved in other biological functions, such as RNA metabolism in addition to DNA repair. Furthermore, recent results have shown that various classes of RRM and G-patch domains exist and are involved in protein as well as in nucleic acid binding. Indeed, even though the RRM domain is the most abundant type of eukaryotic RNA-binding motif, the structures of the U2snRNP auxiliary factor (U2AF) heterodimeric splicing factor have revealed two examples of noncanonical RRMs with specialized features for protein recognition (*6*, *7*). Likewise, the G-patch motif of MIA-14 and MPMV proteinases is able to bind single-stranded DNA and RNA (*8*), whereas the G-patch motif of Spp2 is involved in the protein interaction with prp2 (*9*). These data reflect the multiple role of these domains and show that it is necessary to explore experimentally the structure and function of such domains in TgDRE.

To characterize the behavior and structure in solution of the entire C-terminal sequence of TgDRE, a combination of analytical techniques was applied to five different constructs encompassing one, two, or three domains (Table 1). We systematically analyzed the solubility, stability, degradation, oligomerization state, secondary-structure content, and tertiary fold of each construct. RNA-binding assays were also performed to define their putative biological functions in the parasite TgDRE.

MATERIALS AND METHODS

Cloning and Synthesis. The different constructs, named RRM, Gp + RRM, and SF + Gp + RRM, were subcloned by PCR from the TgDRE ORF contained in the TA-cloning vector (Invitrogen) into the pENTR/D-TOPO vector (Invitrogen) to create entry vectors as described in the Gateway system (Invitrogen). Target genes of entry clones were transferred into the destination vector pDEST15 via a LR recombination reaction to create expression clones that encode N-terminal glutathione-*S*-transferase (GST)-tagged proteins. The G-patch and SF domains, 60 and 63 residues long, respectively, were synthesized in solid phase using the Fmoc strategy. The synthetic peptides were purified by HPLC on a semipreparative C18 column. Molecular weights were measured by mass spectroscopy. The purity of the G-patch and the SF peptides is about 96 and 90%, respectively.

Expression and Purification. Recombinant proteins were overexpressed in a BL21Star(DE3) *E. coli* strain (Invitrogen). ^{15}N- or ^{13}C/^{15}N-labeled proteins were expressed in minimal medium (M9) supplemented with $^{15}NH_4Cl$ or $^{15}NH_4Cl$/^{13}C-glycerol, respectively. The cell culture was induced with 0.5 mM IPTG for 3 h at 30 °C. Cell pellets were sonicated in 50 mM sodium phosphate buffer, 400 mM NaCl, 5 mM β-mercaptoethanol, and 1 mg/mL lysozyme at pH 7.4. Proteins were purified by glutathione-affinity chromatography (Amersham Biosciences GSTrap) under standard conditions followed by an overnight His-tagged TEV protease cleavage. To remove the GST tag and the TEV protease, the samples were reloaded onto glutathione- and Ni^{2+}-affinity columns arranged in series (Amersham Biosciences GSTrap and Hi-Trap). The proteins were then further purified by size-exclusion chromatography (Amersham Biosciences Sephacryl S-100 HP). All of the purification steps except for the last one were done in the presence of a protease inhibitor cocktail (Roche). The cleavage and molecular weight were checked for each protein by N-terminal sequencing and mass spectrometry, respectively. The final buffer used for all nuclear magnetic resonance (NMR) experiments was 20 mM sodium phosphate and 4 mM TCEP at pH 6.5 for the RRM domain and at pH 7 supplemented with 100 mM NaCl for the other proteins.

Dynamic Light Scattering (DLS). The oligomeric state was evaluated by DLS using a 1 cm quartz cell in a DynaPro MS800 (Proterion) instrument with a laser at 824 nm. Protein samples were centrifuged for 30 min at 15000*g* to remove possible particles that would interfere with DLS measurements.

Circular Dichroism (CD) Spectroscopy. Far-UV CD spectroscopy was performed at room temperature on a Jobin Yvon CD6 spectrometer with a cylindric cell of 0.02 cm path length. The concentration of proteins was adjusted to 0.5 mg/mL in 20 mM phosphate buffer at pH 7.4. Each spectrum was the average of three successive individual scans from 180 to 260 nm by a step of 0.5 nm with 1−5 s of integration time per nanometer. The spectrum of the buffer alone was recorded under identical conditions and subtracted from the sample spectra. The secondary-structure content was estimated using CONTINLL (*10*, *11*) from the CDPro software package (*12*).

NMR Spectroscopy. For the SF domain, ^{1}H experiments were recorded at 298 K on Bruker 500 and 600 MHz spectrometers at 1.15 mM in H_2O at pH from 3.9 to 7.2. Total correlation spectroscopy (TOCSY) (*13*), nuclear Overhauser effect spectroscopy (NOESY) (*14*), and correlation spectroscopy (COSY) (*15*) experiments were recorded at pH 4.4.

For the G-patch domain, ^{1}H experiments were recorded at 298 K on a Varian Inova 600 MHz spectrometer at 1 mM,

in several solvents: water, 20 mM sodium phosphate buffer with 0−150 mM NaCl at pH 3.7 and 6.2. TOCSY, NOESY, and COSY experiments were recorded in 20 mM sodium phosphate buffer at pH 6.2. The SF and G-patch sequence-specific assignments were achieved according to the standard method developed by Wüthrich (*16*).

The RRM, Gp + RRM, and SF + Gp + RRM samples were concentrated to 0.3−1 mM, with 15% D_2O. ^1H-^{15}N heteronuclear single-quantum correlation (HSQC)−TROSY spectra of the three proteins were recorded at 298 K on a Varian Inova 600 MHz spectrometer equipped with a cryoprobe. For structure determination of the RRM domain, triple-resonance experiments were recorded at 288 and 298 K, processed with NMRPipe and NMRDraw (*17*) and analyzed with XEASY (*18*). Backbone HN, Hα, Cα, C*β*, CO, and N assignments of the RRM were obtained using standard triple-resonance HNCA, HNCO, HNCACB, and CBCA(CO)NH experiments (*19*). Nuclear Overhauser effect (NOE) interactions characteristic of the secondary structures were obtained from ^{15}N-edited NOESY−HSQC (*20*) recorded at 288 and 298 K with a mixing time of 100 ms. $^3J(H^NH\alpha)$ coupling constants were calculated from the HNHA experiment (*21*). For the evaluation of the amide proton solvent-exchange rates, 74 ^1H-^{15}N HSQC spectra were collected over 5 days using a freshly prepared lyophilized sample of protein dissolved in 99.9% D_2O. $^1H \rightarrow {}^{15}N$ NOE experiments (*22*) were acquired at 288 K on a Varian Inova 500 MHz spectrometer. Proton saturation was achieved by application of ^1H 120° pulses every 5 ms (during 3 s), and the overall delay between scans was 5 s in NOE experiments with and without ^1H saturation.

Molecular Modeling. A total of 20 3D models of TgDRE−RRM (residues 356−466) were built with Modeller 6v2 (*23*). To find the structures of proteins suitable for modeling, the NCBI PSI-BLAST (*24*) was launched against the Protein Data Bank (*25*) with TgDRE−RRM as query sequence. As the result, the atomic structures of three following proteins containing the RRM domain were used as templates: 1JMT ($U2AF^{35}$), 1O0P ($U2AF^{65}$), and 1P1T (CstF-64). First, a structural alignment of the templates was performed on CE (*26*). Then, the sequence of TgDRE−RRM was added to this alignment, taking into account its secondary structures determined on the basis of our NMR data. Furthermore, information about regular secondary-structure elements was added to the input of Modeller to constrain the conformation of the regions corresponding to the helices and the four-stranded antiparallel sheet. This information is taken into account by a special routine of the program that makes additional restraints on φ,ψ angles, distances between some atoms, and hydrogen bonds enforcing the conformation of the segment in question. The quality of models was assessed using PROCHECK (*27*), and the structures were analyzed and visualized with MOLMOL (*28*).

RNA-Binding Experiments. The titration of 0.25 μM synthetic RNA oligonucleotides, seven-nucleotide oligo(U) or oligo(A) labeled with fluorescein (Dharmacon) by the G-patch and the RRM domain, respectively, was followed by fluorescence anisotropy. Anisotropy was measured with a PTI Quantamaster fluorometer equipped with polarizers for the excitation and emission beams using a photomultiplier tube in the L configuration. All experiments were carried out in a 1 cm path-length cell at room temperature with excitation and emission wavelengths of 490 and 518 nm, respectively. The band-passes of the excitation and emission monochromators were set at 6 and 9 nm, respectively. Steady-state fluorescence anisotropy was calculated according to the equation, $A = (I_{VV} - GI_{VH})/(I_{VV} + 2GI_{VH})$, where A is anisotropy, $G = I_{HV}/I_{HH}$ is a correction factor for the wavelength-dependent distortion, and I is the fluorescence intensity component. The fluorescence anisotropy is expressed in arbitrary units. Each point is the result of 20 recordings taken during a 2 min period. All measurements were carried out in 25 mM Tris/HCl buffer at pH 8 containing 50 mM NaCl. The dissociation constant was estimated by fitting the anisotropy data to the binding isotherm equation.

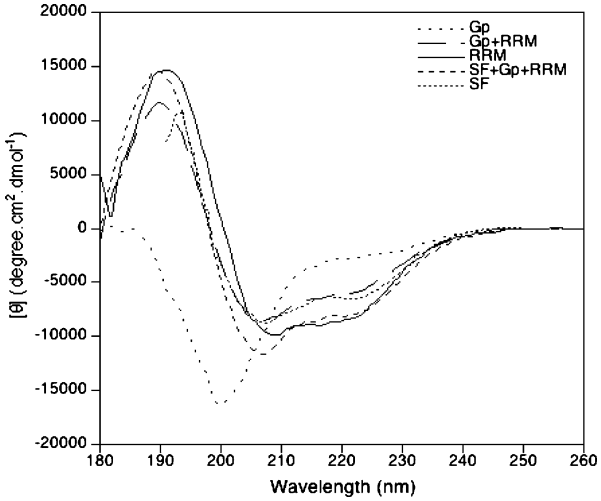

FIGURE 1: Far-UV CD spectra for Gp, RRM, Gp + RRM, and SF + Gp + RRM. The secondary-structure content estimated by CONTINLL gave the following values: RRM (27% α, 23% *β*, 21% turn, and 28% irregular), Gp + RRM (20% α, 26% *β*, 23% turn, and 31% irregular), and SF + Gp + RRM (25% α, 21% *β*, 23% turn, and 31% irregular).

RESULTS

Folding Analysis. The SF domain presents poor solubility. Far-UV CD data show a maximum ellipticity at 195 nm, a minimum at 210, and an inflection at 225 nm testifying for a major helical content of SF (Figure 1). In agreement with the CD data, about 25 NOE cross-peaks are present in the NH−NH region of the NOESY spectrum, indicating that the SF domain is structured, at least partially, as a helix. Only a partial assignment was obtained because of the low dispersion of the NMR signals. The NOESY spectrum analysis does not provide clear evidence for the SF domain to adopt a stable tertiary fold.

Under our experimental conditions, we were unable to express the G-patch domain. Therefore, we decided to synthesize the corresponding peptide. This G-patch peptide is highly soluble up to 1.5 mM at least and is always found in a monomeric state independently of the solvent conditions tested (ionic strength, pH, and organic solvent). Far-UV CD (Figure 1) analysis shows an absence of secondary structure. The ^1H NMR spectra do not indicate additional secondary and tertiary folds. The different buffer conditions tested had no effect in promoting the G-patch folding.

The RRM domain was abundantly expressed in *E. coli* and remained soluble and stable at low concentrations,

whereas the protein precipitated at concentrations higher than 0.8 mM in a sodium phosphate buffer containing 50 mM NaCl at pH 7. Solution conditions (buffer, pH, and added stabilizers) were optimized using the microdrop screening technique (*29*) to improve the RRM solubility and stability at high concentrations. Finally, 20 mM sodium phosphate buffer at pH 6.5 containing the reducing agent TCEP but no NaCl was used for the biophysical characterization. The RRM domain remains stable and in a monomeric state in these conditions. The CD spectrum decomposition of RRM estimates the α and β content as 27 and 23%, respectively (Figure 1). The well dispersion of peaks in the ^{1}H-^{15}N HSQC spectrum (Figure 2A) indicates that this domain adopts a stable tertiary structure.

Because the folding analysis of the three single domains showed that only RRM was able to adopt a stable tertiary fold, the multidomain proteins Gp + RRM and SF + Gp + RRM were used to study their influence on the folding.

The Gp + RRM and SF + Gp + RRM proteins were well-expressed in *E. coli*, but they were less stable than RRM alone during purification. DLS spectrophotometry showed that freshly prepared samples of Gp + RRM and SF + Gp + RRM appear respectively in a monomeric and dimeric state in solution. Nevertheless, mass spectrometry and sequencing indicated a high susceptibility to protease degradation of N-terminal parts that led to the aggregation of the Gp + RRM and SF + Gp + RRM samples with time.

A comparison of the Gp, RRM, and Gp + RRM CD spectra shows that the G-patch domain is unstructured on its own as well as in the presence of the downstream RRM domain (Figure 1). Indeed, no increase of the secondary-structure content is observed from RRM to Gp + RRM. However, the addition of SF at the N terminus of Gp + RRM in the triple domain construct increases the overall helical secondary-structure content. This gain in helical conformation is of more than 25 amino acids and could be related to the SF module conformation observed for the single domain.

Among the five different constructs used to characterize the structure of the entire C terminus of TgDRE, three of them (RRM, Gp + RRM, and SF + Gp + RRM) contain the RRM domain. The comparative dispersion pattern of ^{1}H and ^{15}N chemical shifts in the HSQC spectra of this series of proteins indicates that the majority of well-dispersed peaks belongs to the RRM domain (Figure 2). The signals of SF and Gp are largely superimposed. This is indicative of their unfolded state. Only a few peaks undergo slight shifts in the HSQC spectra of the single, double, and triple constructs, showing that the RRM three-dimensional structure is not affected by the presence of the upstream SF and G-patch domains.

NMR and CD analysis of the single and multiple domain constructs allowed us to identify a core globular domain corresponding to the RRM domain. Residue assignments upstream of RRM were prevented because of severe NMR signal overlaps and a high susceptibility of Gp + RRM and SF + Gp + RRM to protease degradation leading to their aggregation. Therefore, only the NMR structure of the 12.9 kDa RRM domain was determined.

Structure Determination of the RRM Domain of TgDRE. Sequence-specific backbone assignments were determined using standard triple-resonance methods (*19*). All of the backbone amino acids were assigned with the exception of the first two, a Gly and an Asn. A large percentage of the resonances was assigned: 93% of HN and N, 91% of Cα, 86% of Hα, 91% of CO, and 88% of Cβ atoms. Three residues C444, E445, and E446 were not detected in the triple-resonance experiments because of their peculiar relaxation properties but were identified in the NOESY−HSQC spectrum.

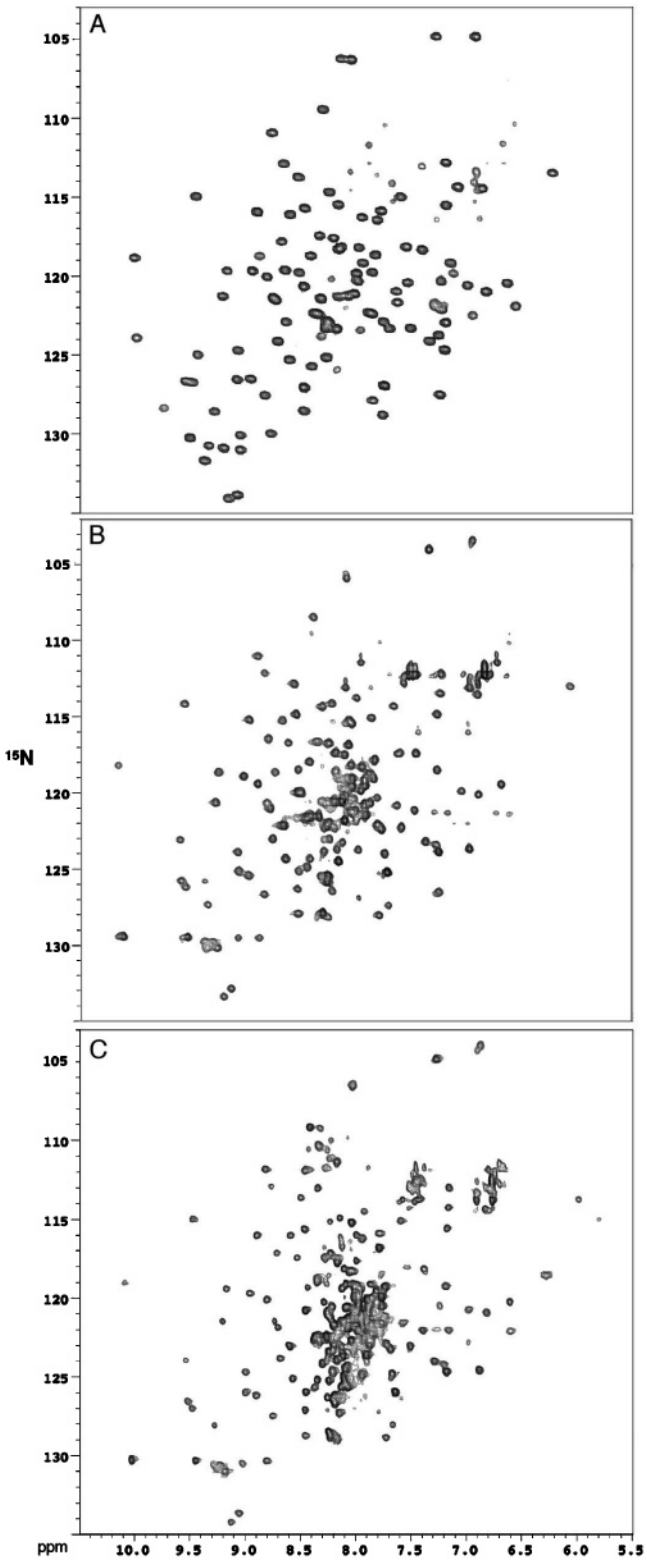

FIGURE 2: ^{1}H-^{15}N HSQC−TROSY spectrum of (A) RRM, (B) Gp + RRM, and (C) SF + Gp + RRM.

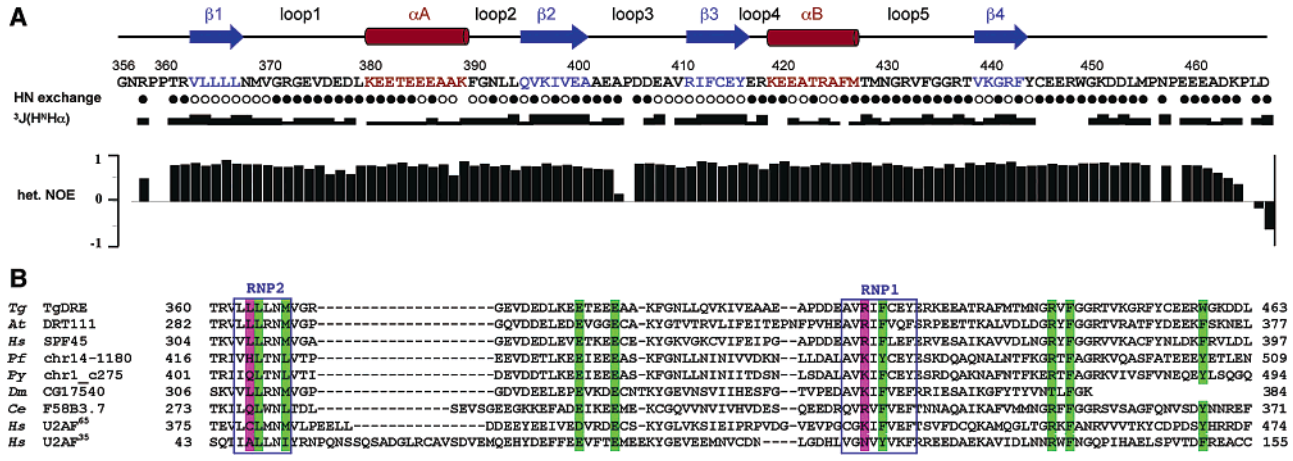

FIGURE 3: (A) Summary of NMR data for the RRM of TgDRE. Fast/slow exchange of amide proton in D_2O solution is shown by closed/open circles. $^3J(H^NH\alpha)$ coupling constants are indicated by small, medium, and large rectangles for values <5, 5 < − < 8, and >8 Hz, respectively. Secondary-structure elements from the CSI consensus resulting of the Ha, Ca, Cb, and CO secondary chemical shifts are indicated on top. (B) RRM sequence alignment of *T. gondii* TgDRE (gi 21666569), *A. thaliana* DRT111 (gi 20141383), *H. sapiens* SPF45 (gi 14249673), *P. falciparum* chr14−1180 (gi 23509735), *P. yoelii* ch1_c275 (gi 23482574), *D. melanogaster* CG17540 (gi 7289585), *C. elegans* F58B3.7 (gi 7504595), *H. sapiens* U2AF^{35}−UHM (gi 267187), and *H. sapiens* U2AF^{65}−UHM (gi 228543). Conserved residues that mediated recognition of SF1 by U2AF^{65}−UHM and U2AF^{65}−ligand by U2AF^{35}−UHM are shown in green. Nonconserved residues of the RNP motifs in a comparison with the consensus are colored in magenta.

The secondary structure of TgDRE−RRM was determined from analysis of secondary chemical shifts, $^3J_{NH,H\alpha}$ coupling constants, and amide proton exchange data (Figure 3). The localization of the secondary-structure elements was achieved using the CSI consensus resulting from the Hα, Cα, Cβ, and CO chemical shifts. The RRM domain of TgDRE presents four β strands (residues 362−366, 394−400, 410−415, and 438−442), two α helices (residues 379−388 and 418−426), and a long and poorly structured C terminus (residues 443−466). The resulting α and β content is consistent with the decomposition of the CD spectrum. A total of 93 $^3J_{NH,H\alpha}$ coupling constants were extracted from the HNHA experiment. A total of 12 of them lower than 5 Hz are found in the helices, and 18 higher than 8 Hz are localized in the β strands. Among the 103 exchangeable amide protons identified in HSQC experiments, 32 protons still giving rise to correlations after 50 h are considered to be slowly exchanging and therefore engaged in hydrogen bonds. A total of 7 of them are localized in the helices, and 22 are found in β strands. All of the amide protons of the N (Asn356−Arg361) and C (Cys444−Asp466) termini and almost all of the loops were quickly exchanged.

The structure of the antiparallel β sheet was fully confirmed by the presence of 13 NOE cross-peaks corresponding to the pairs of HN−HN and HN−Hα protons connecting the β4−β1−β3−β2 strands and by the hydrogen bonds identified from the HSQC and NOESY−HSQC spectra. They included the HN−HN interactions between Leu365−Arg452, Leu367−Lys450, Gln395−Glu415, Lys397−Phe413, Leu366−Ile412, and Leu364−Cys414 and the HN−Hα interactions between Leu367−Gly441, Leu365−Phe443, Val396−Glu415, Leu366−Phe413, Leu364−Glu415, Val363−Tyr416, and Cys414−Leu365.

Model of the RRM domain of TgDRE was built by homology modeling with the addition of restraints reinforcing the localization of the secondary-structure elements, obtained from the experimental NMR data (Figure 4). The homologous proteins used as templates were the RRM domain of the CstF-64 the structure of which was determined in

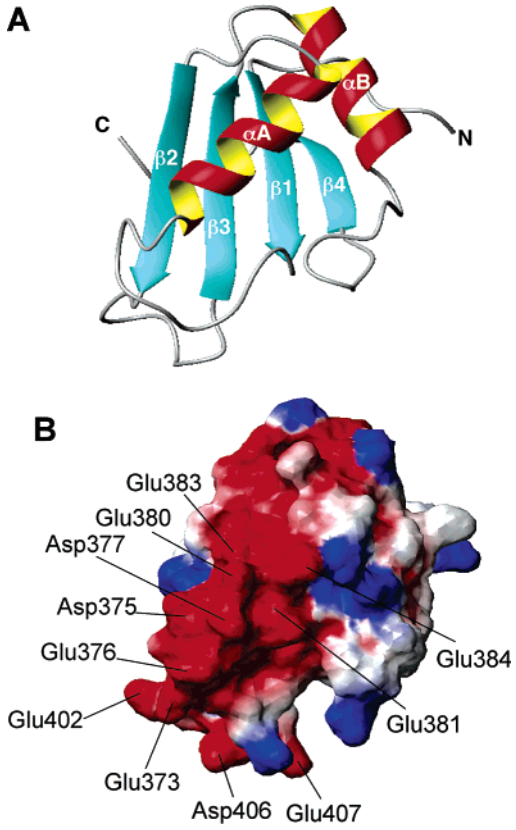

FIGURE 4: (A) Ribbon representation of the RRM of TgDRE. (B) Surface representation of the RRM of TgDRE in the same orientation as A calculated using MOLMOL with the positive charge in blue and the negative charge in red.

complex with a RNA fragment, and the two RRMs involved in protein recognition, U2AF^{65} (RRM3) and U2AF^{35}. Because the NMR data indicated that the C-terminal part of RRM is poorly structured and flexible, we carried out the construction of models from residues 356−449 of the protein. Structural alignment of this part of the RRM domain of TgDRE with the RRMs used for homology modeling gives a sequence similarity/identity of 31/21% with CstF-64, 32/

21% with U2AF65, and 26/16% with U2AF35. The generated models are consistent with the intra- and interstrands NOE cross-peaks identified in the NOESY−HSQC spectra and the hydrogen-bond donors resulting from the exchange experiments. Moreover, they display a good overall quality with 83.7% of backbone φ and ψ dihedral angles in the core, 13% in the allowed, 2.3% in the generously allowed, and 1.1% in the disallowed region of the Ramachandran plot.

The RRM of TgDRE has the same three-dimensional scaffold as the other RRM containing proteins characterized by a four-stranded β sheet packed against two perpendicular α helices (Figure 4A). The central pair of β strands presents two short sequence motifs, ribonucleoprotein (RNP)1 and RNP2, of eight and six residues long, respectively. The core of RRM would be stabilized by hydrophobic contacts through residues Leu364, Leu366, Val396, Ile398, Ile412, Tyr416, and Met429. The electrostatic potential surface shows that the positively charged residues are scattered on the surface of the molecule, whereas those charged negatively are clustered along the helix A and are well-exposed to the solvent (Figure 4B). The β-sheet surface presents an equivalent repartition of the positive and negative charges.

To probe the dynamics on the fast NMR time scale of the RRM domain, ^{1}H-^{15}N heteronuclear NOEs were measured (Figure 3A). They confirmed the presence of a well-defined structure with a mean root-mean-square deviation (rmsd) value of 0.76 ± 0.03 (on residues 360−399 and 403−454). Higher rmsd values are observed in the secondary-structure elements, and low rmsd values (from 0.59 ± 0.29 to 0.75 ± 0.03) are found in the loops with the exception of the fourth that is only two residues long. Loop 3 appears particularly flexible based on the low values of heteronuclear NOEs (mean value of 0.59 ± 0.29). The secondary structures of the RRM models are well-defined, whereas their loops adopt various conformations and orientations. The corresponding loops in the template structures also present a flexibility that is in agreement with our NMR dynamics analysis. From the last strand to residue Asn456, the heteronuclear NOE values remain high although no evidence of secondary-structure element is detected. Finally, the heteronuclear NOE values decrease from residue Asn456 toward the C terminus.

RNA-Binding Assays. In addition to structural information, we used some of our constructs to investigate their RNA-binding properties by fluorescence anisotropy. First and because no specific target was known to interact with TgDRE, we used small RNA oligonucleotides labeled with fluorescein. The titration of seven-nucleotide oligo(U) and oligo(A) by a concentrated solution of the G-patch domain resulted in a gradual increase in fluorescence anisotropy, reaching 15 times the basal value at 80 μM (Figure 5). In contrast, the RRM domain fails to show detectable RNA binding with the same oligonucleotides. The weak affinity of the G-patch domain (∼100 μM) does not increase for the Gp + RRM construct (data not shown).

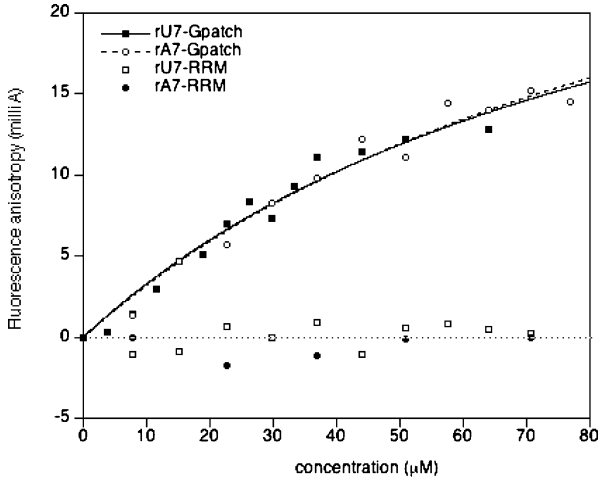

FIGURE 5: RNA-binding assays of seven-oligonucleotides oligo(U) and oligo(A) with RRM and Gp followed by fluorescence anisotropy.

DISCUSSION

TgDRE and its orthologues form a novel family of proteins characterized by three conserved motifs: SF, G-patch, and RRM. The presence of these domains suggests that TgDRE might be involved in other biological functions such as RNA metabolism in addition to its DNA repair activity, which has been previously described (*4*). In the present study, we report the expression, purification, and structural characterization of each domain isolated or combined. We also investigate their RNA-binding activity to further determine their putative biological function in the parasite TgDRE.

The folded part of the C-terminal moiety of TgDRE exhibits the classical $\beta\alpha\beta\beta\alpha\beta$ topology of the RRM domain composed of two α helices packed against a four-stranded antiparallel β sheet (*30*). The two RNP motifs lie in the center of the β sheet, with RNP1 in β3 and RNP2 in β1, but show a low sequence similarity with the consensus sequence. Because RRMs were essentially described as RNA-binding modules, we examined the ability of TgDRE−RRM to bind small RNA oligonucleotides. No interaction between the RRM domain of TgDRE and small oligo(U) or oligo(A) has been detected even though the binding platform common to all RRMs is not in essence sequence-specific (*31*). Recent structures of the heterodimeric splicing factor U2 snRNP auxiliary factor (U2AF) have revealed two domains presenting the RRM fold but unable to bind RNA (*6*, *7*). Instead, it has been demonstrated that these unusual RRMs have specialized features for protein recognition and therefore were called U2AF homology motifs or UHM to reflect their distinct role in the protein−protein interaction (*32*). Three characteristic features distinguish UHMs from classical RRMs: atypical RNP-like motifs, an Arg-X-Phe motif in the last loop, and an acidic character of the helix A. It was recently proposed that DRT111 and SPF45, two sequence homologues of TgDRE, belong to a group of 12 candidates closely related to UHMs (*32*). Therefore, we investigated the possibility that the RRM of TgDRE also belongs to this novel class of protein recognition motifs. Sequence and structure comparisons show that these three characteristics are also present in the RRM domain of TgDRE. First, the RNP1 and RNP2 motifs of TgDRE−RRM contain only one (Phe412) of the three highly conserved aromatic residues that have been shown in classical RRMs to be responsible for nonspecific ligand interactions via stacking with RNA bases (*33*−*35*). The corresponding amino acids are replaced by basic and aliphatic residues (Arg398 and Leu363). Then, from the structures of the complexes U2AF^{35}−UHM/U2AF^{65}−ligand and U2AF^{35}−UHM/SF1−ligand, details of the interaction have emerged. A tight hydrophobic pocket

involving the Arg-X-Phe motif of loop 5 and the acidic residues of the helix A was identified in UHMs. The residues of this pocket are conserved in the RRM of TgDRE: Arg431−Val432−Phe433 on one side and Glu380 and Glu384 on the other side. TgDRE is the only UHM candidate to possess a Val in the tripeptide motif. The electrostatic potential surface of the RRM of TgDRE reveals a negatively charged area involving four acidic residues on the helical side, whereas the conventional RRMs exhibit a marked basic character. These acidic residues are clustered along the helix A as in the U2AF^{65}−UHM, where they are solvent-exposed and directly engaged in electrostatic contacts with a stretch of positively charged residues of SF1 (*7*). Moreover, the low isoelectric point of TgDRE−RRM (around 4.5) does not favor interactions with the negatively charged RNA, whereas the conventional RRMs exhibit a pI often higher than 9. In addition, sequence alignment (Figure 3B) of the SF45 family shows that all of its members display unusual RNP motifs and that the residues involved in the hydrophobic pocket are conserved with the exception of the *D. melanogaster* CG17540 sequence, which is shorter than its counterparts.

The RRM domain of TgDRE exhibits a long and poorly structured C-terminal extension after the last β strand. Several hydrophobic residues (Tyr443, Trp448, Leu453, Met454, and Leu465) localized in this C-terminal part may interfere with RNA binding as for U2AF^{65}−UHM (*7*) or contribute to homodimer formation in a manner similar to U1A (*33*), but we observed that TgDRE−RRM remains in a monomeric state during all of our biophysical studies. In several RRM or UHM structures, additional secondary-structure elements have been found after the last β strand as a C-terminal helix for CstF-64 (*36*) or U2AF^{65} (*7*). No evidence for the presence of a third helix in addition to the RRM fold has been found using the CSI consensus, in agreement with the absence of $^{3}J_{NH,H\alpha}$ small coupling constants, the lack of NOEs characteristic of helical conformation, and the presence of fast exchanging amide protons in this N-terminal region. The three residues following the last strand present peculiar relaxation properties, suggesting local conformational exchanges. We do not observe an enhanced flexibility on the fast time scale (pico−nanosecond) for this region, as monitored by high heteronuclear NOE values from the last strand to Asn456. A structural rearrangement of the unstructured C terminus upon the target binding could occur as it happens in hnRNPA1−RRM1 (*37*).

Many of the UHM candidates also contain motifs frequently observed in splicing factors such as canonical RRMs, arginine−serine (RS) domains, zinc fingers, and glycine-rich regions. Up to date, only 42 proteins containing a combination of a RRM domain and a G-patch domain have been identified in the Pfam database (*38*). The G-patch motif, characterized by its glycine-rich sequence signature, is present in a large number of eukaryotic RNA-binding proteins and therefore was suspected for a long time to be a new RNA-interacting module (*39*), although no structural data were available. The secondary-structure predictions of the overall G-patch domains suggested the presence of two α helices, but the CD and NMR data of the G-patch domain of TgDRE show that this single module is mainly unstructured and that the presence of the upstream SF and downstream RRM domains is not sufficient to promote its folding. We cannot rule out the possibility that the G-patch domain may require a specific target to be structured. Recent studies have shown that this domain could interact with RNA or DNA (*8*) as well as with a protein partner (*9*). Indeed, it has been experimentally confirmed that the G-patch domain of MIA-14 and MPMV proteinases are able to bind single-stranded nucleic acids, DNA and RNA, and that a highly conserved aromatic residue (Tyr126 of MIA-PR and Tyr121 of MPMV-PR) is critical for this interaction. A Trp is found at the equivalent position in TgDRE as well as in *P. falciparum*, *P. yoelii*, and *A. thaliana*, a Phe in *Homo sapiens* SPF45 and *D. melanogaster* CG17540, and a Tyr in *C. elegans* F58B3.7. We demonstrated that the G-patch domain of TgDRE can interact with RNA oligo(A) and oligo(U). Other targets should be tested to improve the interaction affinity and specificity and to examine their influence on the G-patch folding.

In addition to the G-patch and RRM domains, the SF45 family presents the specific SF45-like motif identified for the first time through the bioinformatics study of TgDRE (*4*). We report here the first biophysical data of the SF45-like motif of TgDRE. This domain does not adopt a stable tertiary fold, but CD and NMR data are in favor of a helical conformation for about half of the domain residues in the mono- and multidomain proteins. A comparison of the different constructs suggest that SF could to be involved in the dimeric association of the C-terminal moiety of TgDRE encompassing the three domains. This is in agreement with the secondary-structure predictions that showed a high propensity of SF to adopt a coiled-coil conformation (residues 230−257). The helical coiled-coil structural motif mediates subunit oligomerization of a large number of proteins, including many DNA-binding proteins such as the family of basic region leucine zipper (b/ZIP) proteins for which coiled-coil domains are responsible for specific recognition between molecules (*40*, *41*). It remains to be determined the impact of the possible oligomerization of TgDRE on its biological functions.

The DNA-repair activity of TgDRE demonstrated in vitro in a heterologous system was partial. This can be interpreted as the result of the phylogenetical distance between bacteria and protozoan parasites or as the involvement of TgDRE in other biological functions. Here, we have demonstrated the existence of an interaction between the G-patch and RNA. Additionally, the structural study of the RRM domain highlighted its strong probability to be involved in a protein−protein interaction. At present, the role of TgDRE in the parasite remains to be determined. In particular, it is still unknown whether TgDRE is a component of the splicing machinery or whether it is involved in other aspects of pre-mRNA processing. Our characterization provides a starting point for understanding the precise function of this protein in *T. gondii*. Further works are now in progress to identify the protein partners of TgDRE in the parasite.

ACKNOWLEDGMENT

We thank Dr. Françoise Baleux (Institut Pasteur, France) for the synthesis of G-patch, Dr. Véronique Arluison (IBPC, France) for the gift of the synthetic oligo(U) and oligo(A), Catherine Simenel for assistance with NMR experiments, and the Plate-forme de Biophysique (Institut Pasteur, France). K. F. was supported by a fellowship of the Ministère de l'Education Nationale, de la Recherche et de la Technologie

(MENRT) and the Fondation pour la Recherche Médicale (FRM).

REFERENCES

1. Tenter, A. M., Heckeroth, A. R., and Weiss, L. M. (2000) *Toxoplasma gondii*: From animals to humans, *Int. J. Parasitol. 30*, 1217−1258.
2. Yahiaoui, B., Dzierszinski, F., Bernigaud, A., Slomianny, C., Camus, D., and Tomavo, S. (1999) Isolation and characterization of a subtractive library enriched for developmentally regulated transcripts expressed during encystation of *Toxoplasma gondii*, *Mol. Biochem. Parasitol. 99*, 223−235.
3. Sampath, J., Long, P. R., Shepard, R. L., Xia, X., Devanarayan, V., Sandusky, G. E., Perry, W. L., III, Dantzig, A. H., Williamson, M., Rolfe, M., and Moore, R. E. (2003) Human SPF45, a splicing factor, has limited expression in normal tissues, is overexpressed in many tumors, and can confer a multidrug-resistant phenotype to cells, *Am. J. Pathol. 163*, 1781−1790.
4. Dendouga, N., Callebaut, I., and Tomavo, S. (2002) A novel DNA repair enzyme containing RNA recognition, G-patch and specific splicing factor 45-like motifs in the protozoan parasite *Toxoplasma gondii*, *Eur. J. Biochem. 269*, 3393−3401.
5. Pang, Q., Hays, J. B., and Rajagopal, I. (1993) Two cDNAs from the plant *Arabidopsis thaliana* that partially restore recombination proficiency and DNA-damage resistance to *E. coli* mutants lacking recombination-intermediate-resolution activities, *Nucleic Acids Res. 21*, 1647−1653.
6. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer, *Cell 106*, 595−605.
7. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP, *Mol. Cell. 11*, 965−976.
8. Svec, M., Bauerova, H., Pichova, I., Konvalinka, J., and Strisovsky, K. (2004) Proteinases of betaretroviruses bind single-stranded nucleic acids through a novel interaction module, the G-patch, *FEBS Lett. 576*, 271−276.
9. Silverman, E. J., Maeda, A., Wei, J., Smith, P., Beggs, J. D., and Lin, R. J. (2004) Interaction between a G-patch protein and a spliceosomal DEXD/H-box ATPase that is critical for splicing, *Mol. Cell. Biol. 24*, 10101−10110.
10. Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism, *Biochemistry 6*, 33−37.
11. van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra, *Anal. Biochem. 15*, 110−118.
12. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem. 287*, 243−251.
13. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotopic mixing: Application to proton correlation spectroscopy, *J. Magn. Reson. 53*, 521−528.
14. Kumar, A., Ernst, R. R., and Wuthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton−proton cross-relaxation networks in biological macromolecules, *Biochem. Biophys. Res. Commun. 95*, 1−6.
15. Aue, W. P., Bartholi, E., and Ernst, R. R. (1976) Two-dimensional spectroscopy: Application to nuclear magnetic resonance, *J. Chem. Phys. 64*, 2229−2246.
16. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
17. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
18. Bartels, C., Xia, T. H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, *J. Biomol. NMR 5*, 1−10.
19. Sattler, M., Schleucher, J., and Griesinger, C. (1998) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. NMR Spectrosc. 34*, 93−158.
20. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques, *J. Biomol. NMR 4*, 845−858.
21. Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) Measurement of HN−H α J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods, *J. Biomol. NMR 4*, 871−878.
22. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium, *J. Biomol. NMR 4*, 727−734.
23. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct. 29*, 291−325.
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
25. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures, *Eur. J. Biochem. 80*, 319−324.
26. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, *Protein Eng. 11*, 739−747.
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Cryst. 26*, 283−291.
28. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
29. Lepre, C. A., and Moore, J. M. (1998) Microdrop screening: A rapid method to optimize solvent conditions for NMR spectroscopy of proteins, *J. Biomol. NMR 12*, 493−499.
30. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A, *Nature 348*, 515−520.
31. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J. 272*, 2118−2131.
32. Kielkopf, C. L., Lucke, S., and Green, M. R. (2004) U2AF homology motifs: Protein recognition in the RRM world, *Genes Dev. 18*, 1513−1526.
33. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and 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.
34. Allain, F. H., Howe, P. W., Neuhaus, D., and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein, *EMBO J. 16*, 5764−5772.
35. Price, S. R., Evans, P. R., and Nagai, K. (1998) Crystal structure of the spliceosomal U2B″−U2A′ protein complex bound to a fragment of U2 small nuclear RNA, *Nature 394*, 645−650.
36. Perez Canadillas, J. M., and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein, *EMBO J. 22*, 2821−2830.
37. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA, *Genes Dev. 13*, 1102−1115.
38. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database, *Nucleic Acids Res. 30*, 276−280.
39. Aravind, L., and Koonin, E. V. (1999) G-patch: A new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins, *Trends Biochem. Sci. 24*, 342−344.
40. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled coils: A highly versatile protein folding motif, *Trends Cell. Biol. 11*, 82−88.
41. Lupas, A. (1996) Coiled coils: New structures and new functions, *Trends Biochem. Sci. 21*, 375−382.

BI051948E