PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Solution structure of the GUCT domain from human RNA helicase II/Guβ reveals the RRM fold, but implausible RNA interactions

Satoshi Ohnishi,[1] Kimmo Pääkkönen,[1] Seizo Koshiba,[1] Naoya Tochio,[1] Manami Sato,[1] Naohiro Kobayashi,[1] Takushi Harada,[1] Satoru Watanabe,[1] Yutaka Muto,[1] Peter Güntert,[1] Akiko Tanaka,[1] Takanori Kigawa,[1,2] and Shigeyuki Yokoyama[1,3]*

[1] Systems and Structural Biology Center, RIKEN, Tsurumi, Yokohama 230-0045, Japan
[2] Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Midori-ku, Yokohama 226-8503, Japan
[3] Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

ABSTRACT

Human RNA helicase II/Guα (RH-II/Guα) and RNA helicase II/Guβ (RH-II/Guβ) are paralogues that share the same domain structure, consisting of the DEAD box helicase domain (DEAD), the helicase conserved C-terminal domain (helicase_C), and the GUCT domain. The N-terminal regions of the RH-II/Gu proteins, including the DEAD domain and the helicase_C domain, unwind double-stranded RNAs. The C-terminal tail of RH-II/Guα, which follows the GUCT domain, folds a single RNA strand, while that of RH-II/Guβ does not, and the GUCT domain is not essential for either the RNA helicase or foldase activity. Thus, little is known about the GUCT domain. In this study, we have determined the solution structure of the RH-II/Guβ GUCT domain. Structural calculations using NOE-based distance restraints and residual dipolar coupling-based angular restraints yielded a well-defined structure with β-α-α-β-β-α-β topology in the region for K585-A659, while the Pfam HMM algorithm defined the GUCT domain as G571-E666. This structure-based domain boundary revealed false positives in the sequence homologue search using the HMM definition. A structural homology search revealed that the GUCT domain has the RRM fold, which is typically found in RNA-interacting proteins. However, it lacks the surface-exposed aromatic residues and basic residues on the β-sheet that are important for the RNA recognition in the canonical RRM domains. In addition, the overall surface of the GUCT domain is fairly acidic, and thus the GUCT domain is unlikely to interact with RNA molecules. Instead, it may interact with proteins via its hydrophobic surface around the surface-exposed tryptophan.

Proteins 2009; 74:133–144.

Key words: RNA helicase; GUCT domain; NMR; residual dipolar couplings; RRM domain superfamily; structural genomics.

INTRODUCTION

Human RNA helicase II/Guα (RH-II/Guα or Deadbox protein 21) is a multifunctional enzyme that unwinds double-stranded RNA in the 5′ to 3′ direction and folds single-stranded RNA in an ATP-dependent manner.[1–5] These RNA-unwinding and RNA-folding activities are independent, and they reside in distinct regions of the protein. The RNA helicase activity is catalyzed by the N-terminal three-quarters of the molecule in the presence of Mg^{2+}, where as the RNA-foldase activity is located in the C-terminal region and functions in a Mg^{2+} independent manner.[2] As shown in Figure 1(A), RH-II/Guα consists of three conserved structural domains: the DEAD box helicase domain (DEAD domain, Pfam code: PF00270), the helicase conserved C-terminal domain (Helicase_C domain, Pfam code: PF00271), and the GUCT (NUC152) domain (Gu C-terminal domain, Pfam code: PF08152) according to the Pfam HMM algorithm (http://www.sanger.ac.uk/Software/Pfam/). The structures of the DEAD and Helicase_C domains are known (e.g., PDB codes: 1T6N,[6] 1VEC DOI: 10.2210/pdb1vec/pdb, and 3B6E DOI: 10.2210/pdb3b6e/pdb). These domains play critical roles in the RNA-helicase activity of RH-II/Guα.[2,5] In contrast, the structure of the GUCT domain has not been solved. The GUCT domain is followed by three FRGQR repeats and one PRGQR sequence, which are separated by five amino acid residues. The ~70 residue C-termi-

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: The RIKEN Structural Genomics/Proteomics Initiative (RSGI) of the National Project on Protein Structural; Grant sponsor: Functional Analyses, Ministry of Education, Culture, Sports, Science, and Technology of Japan.
*Correspondence to: Shigeyuki Yokoyama, Systems and Structural Biology Center, RIKEN, 1-7-22 Suehiro-Cho, Tsurumi, Yokohama 230-0045, Japan.
E-mail: yokoyama@biochem.s.u-tokyo.ac.jp
Received 29 February 2008; Revised 24 April 2008; Accepted 30 April 2008
Published online 9 July 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22138

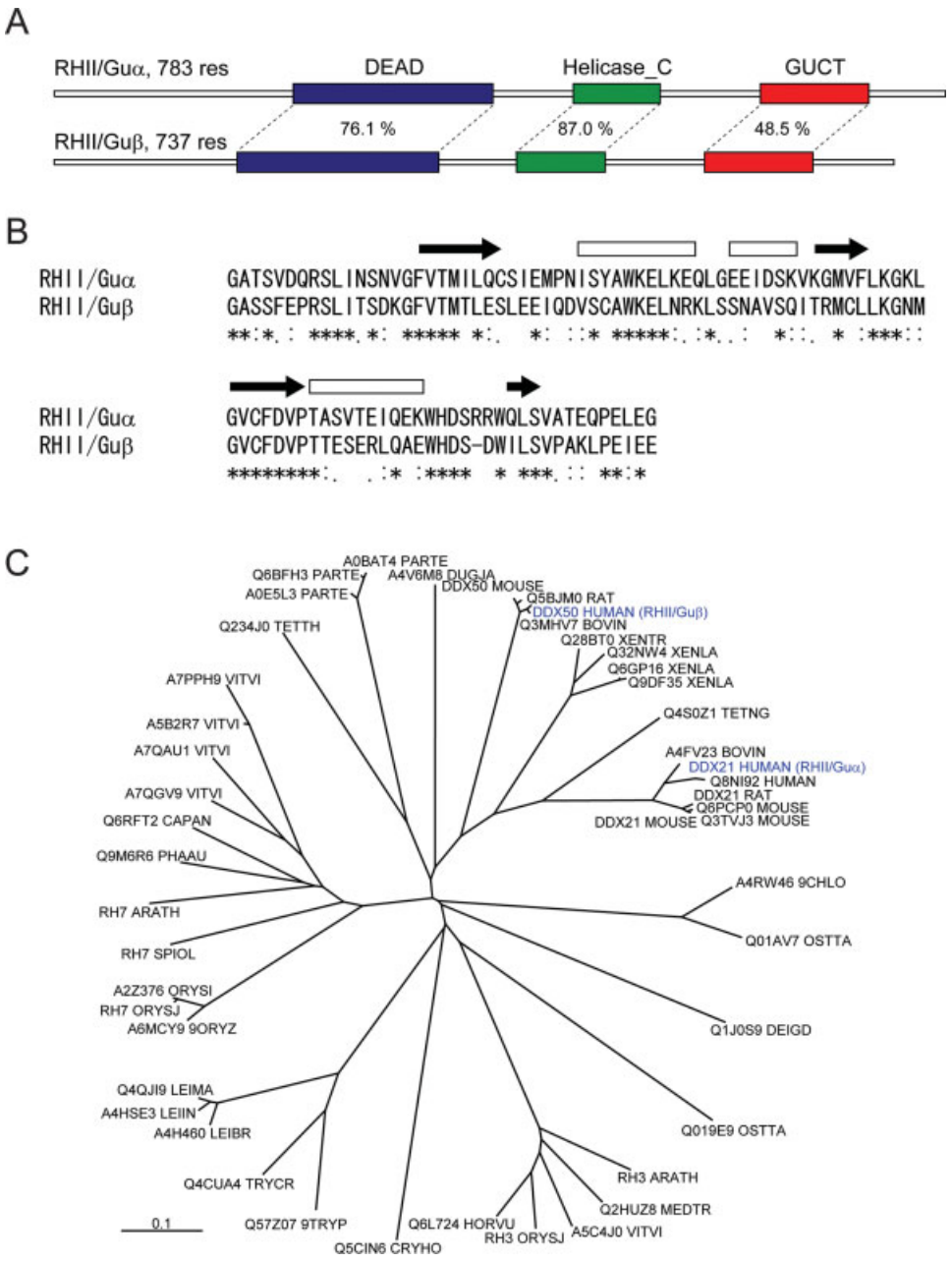


Figure 1

Comparative paralogues of human RH-II/Gu. **A:** Comparison of the domain structures between RH-II/Guα and RH-II/Guβ. The regions shown in blue, green, and red, respectively, correspond to the DEAD domain (Pfam code: PF00270), the helicase conserved C-terminal domain (Helicase_C, Pfam code: PF00270), and the GUCT domain (Pfam code: PF08152), according to the prediction by the Pfam HMM definition. **B:** Sequence comparison between the human RH-II/Guα and RH-II/Guβ GUCT domains. The rectangular bars and arrows represent the α-helices and β-strands, conformed to the present NMR structure of the RH-II/Guβ GUCT domain. **C**: Dendrogram of the UniProt entries found by the homologue search using the Pfam HMM profile for the GUCT domain. The human RH-II/Guα and RH-II/Guβ GUCT domains are highlighted in blue.

nal region, including these repeats, is necessary for the RNA-foldase activity, where as the GUCT domain is not, as demonstrated by mutagenesis studies.[1,2] Thus, the functional role of the GUCT domain in this bifunctional enzyme, RH-II/Guα, remains unclear.

Recently, a paralogue of RH-II/Guα, referred to as RH-II/Guβ and encoded by the human DDX50 gene, has been identified.[7,8] Both proteins localize to nucleoli, suggesting roles in ribosomal RNA production, but RH-II/Guβ also localizes to nuclear speckles containing the

splicing factor SC35, suggesting its possible involvement in pre-mRNA splicing.[8] These paralogues share a similar domain structure constituted by the DEAD domain, the Helicase_C domain, and the GUCT domain [Fig. 1(A)]. The overall sequence identity is 55.6%, and those for the DEAD domain, the Helicase_C domain, and the GUCT domain are 76.1%, 87.0%, and 48.5%, respectively [Fig. 1(B)]. Remarkably, the characteristic (F/P)RGQR repeats in the C-terminal tail of RH-II/Guα are replaced by an arginine-serine-rich sequence in the C-terminal tail of RH-II/Guβ. As predicted from the sequence comparison, *in vitro* assays demonstrated that RH-II/Guβ has RNA-unwinding activity, although it is weaker than that of RH-II/Guα, but no RNA-folding activity.[8]

A sequence homologue search in the uniprot database (http://www.pir.uniprot.org/) revealed that the GUCT domain is unique to the eukaryotic RH-II/Gu proteins. A classification analysis using CLUSTAL W[9] established three primitive classes in the sequence space, with the RH-II/Guα GUCT domain mapped relatively distant from the RH-IIβ GUCT domain in one of the classes [Fig. 1(C)].

Little is known about the structural and functional aspects of the GUCT domain, which was only defined by the characteristic sequence pattern detected by the Pfam HMM algorithm. In this study, we have determined the solution structure of the human RH-II/Guβ GUCT domain using NOE-based distance restraints and residual dipolar coupling-based angular restraints, which allowed redefinition of the boundary of the GUCT domain with the structured region. The present structure provides information about homologue classification as well as a possible functional role.

MATERIALS AND METHODS

Construct design, expression, and purification

The DNA encoding the human DDX50 gene was subcloned by PCR from the human cDNA clone locus NM_024045 (Ultimate™ ORF Clones, Invitrogen, ID: IOH4504). The regions between G571-K660, G571-Y667, G557-K660, G557-Y667, T589-V657, T589-A659, F587-A659, F575-V657, and F575-A659 were each cloned into the expression vector pCR2.1 (Invitrogen), as fusions with an N-terminal His6 affinity tag and a tobacco etch virus (TEV) protease cleavage site. The actual constructs contain seven extra residues (GSSGSSG) after the TEV-cleavage site that is derived from the PCR linker regions. The obtained PCR fragments were directly used as templates for cell-free protein synthesis using ^{15}N-labeled amino acids for the ^{1}H-^{15}N HSQC screening test.

The DNA fragment encoding the screened construct, F575-A659, was incorporated into a plasmid and used in the cell-free protein synthesis system with ^{13}C/^{15}N-labeled amino acids, with subsequent protein purification as described elsewhere.[10] The purified protein was concentrated to 0.6 m*M* in 20 m*M* d_{11}-Tris–HCl buffer (pH 7.0) containing 100 m*M* NaCl, 1 m*M* 1,4-DL-dithiothreitol-d_{10} (d_{10}-DTT), 10% $^{2}H_2O$, and 0.02% NaN_3.

NMR spectroscopy

All spectra were recorded on Bruker Avance 600 and 800 spectrometers at 298 K, unless otherwise mentioned. As described previously, resonance assignments were carried out using a conventional set of triple resonance spectra.[10] Interproton distance restraints were obtained from ^{15}N- and ^{13}C-edited NOESY spectra recorded with a mixing time of 80 ms. All spectra were processed using NMRPipe,[11] and the programs KUJIRA[12] and NMRView[13] were used for optimal visualization and spectral analyses. Complete resonance assignments have been deposited in the BioMagResBank (accession code: 10213).

RDC measurements

An aliquot of a Pf1 phage pellet, purchased from ASLA Biotech (Latvia), was washed with water through repeated ultracentrifugation steps at 90,000 rpm (450,000*g*) and 4°C for 90 min, using a Hitachi himac CS120GX ultracentrifuge with an S100AT5 angle rotor. The pellet was then suspended in, approximately, twice the volume of 30 m*M* *d*-Tris–HCl buffer (pH 7.0) containing 150 m*M* NaCl, 1.5 m*M* d_{10}-DTT, and 15% $^{2}H_2O$. The suspension was mixed with a freshly prepared ^{13}C/^{15}N-labeled protein solution in the same buffer used for the structural calculation with estimated Pf1 phage and protein concentrations of ~10 mg/mL and ~0.4 m*M*, respectively. This protein-liquid crystal sample showed ~8.5 Hz $^{2}H_2O$ quadrupole splitting at 25°C, which is indicative of ~9 mg/mL Pf1 concentration.[14] Two-dimensional ^{1}H-^{15}N IPAP HSQC spectra[15] with 2.48 Hz/point in the indirect dimension and three-dimensional HNCO spectra[16] with 16.50 Hz/point in the carbon dimension were recorded to yield a set of HN-N and Cα-C' residual dipolar couplings.

Restraint generation and structure calculation

Automated NOE cross-peak assignments[17] and structure calculations with torsion angle dynamics[18] were performed using the program CYANA 2.2,[19] which included a module for the incorporation of RDC restraints into the torsion angle dynamics calculations (K. Pääkkönen and P. Güntert, unpublished results). Dihedral angle restraints were obtained from TALOS[20] and applied during the structure calculations. The axial and rhombic components of the alignment tensor of the protein in the protein-liquid crystal sample were initially determined

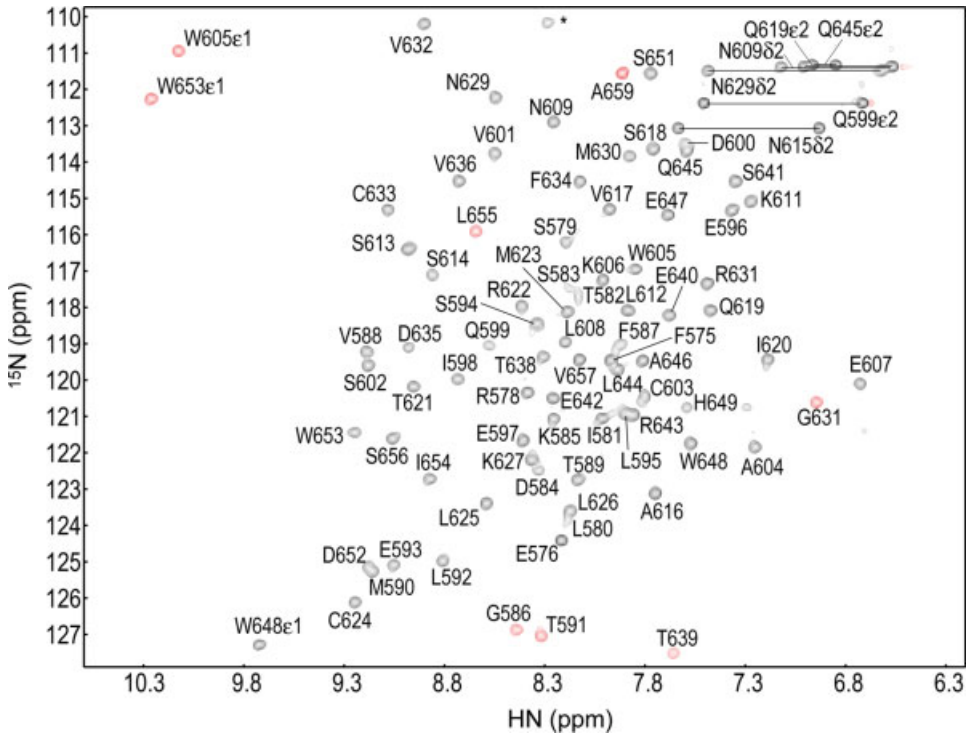


Figure 2

A ^{1}H-^{15}N HSQC spectrum of the human RH-II/Guβ GUCT domain (F575-A659). Red signals represent folded peaks due to a narrow spectrum width. Peaks marked with * are signals from the residues in the artificial tags of the construct.

from the measured HN-N and Cα-C' residual dipolar coupling values using the histogram method[21,22] and subsequently refined iteratively using SVD fitting to preliminary structures.[23] The orientation of the dipole tensor was represented in the structure calculations by a tensor "molecule" that was rotating according to the forces generated from violated residual dipolar coupling restraints.[24] The disordered tail residues were excluded from the analysis. From the total of 147 measured RDCs, 65 ^{1}H-^{15}N RDCs, and 64 Cα-C' RDCs were used in the calculations. The 20 conformers with the lowest CYANA target function values of the 100 calculated conformers were selected.

Energy refinement of the NMR structure

The 20 final CYANA conformers with the lowest target function values were subjected to restrained energy minimization in explicit solvent, using the program OPALp[25,26] with the AMBER force field.[27] The resulting 20 energy-minimized CYANA conformers that represent the solution structure of the GUCT domain from human RH-II/Guβ were evaluated with PROCHECK-NMR[28] and deposited in the Protein Data Bank (PDB code: 2E29).

RESULTS

Construct design

Because little was known about the GUCT domain, we first examined the structural properties of multiple constructs that were designed based on the Pfam HMM algorithm. Figure 1(B) shows the sequences of the GUCT domains from human RH-II/Guα and RH-II/Guβ with the Pfam definition, i.e., RH-II/Guα G620-G716 and RH-II/Guβ G571-E666. Accordingly, we cloned and purified multiple RH-II/Guβ constructs encoding G571-K660, G571-Y667, G557-K660, G557-Y667, T589-V657, T589-A659, F587-A659, F575-V657, and F575-A659 and applied them to the 2D ^{1}H-^{15}N HSQC screening tests. Each of these constructs was synthesized using the same amount of the source reagents in the standardized cell-free system, which was incorporated within the high-throughput NMR pipeline system. The two-dimensional ^{1}H-^{15}N HSQC spectra reflect the construct's behavior, such as protein expression efficiency, protein solubility, protein foldability, and protein stability. All these HSQC spectra are presented in Supplementary Figure S1. The constructs G557-K660, T589-V657, and T589-A659 generated significant amounts of precipitates and poor

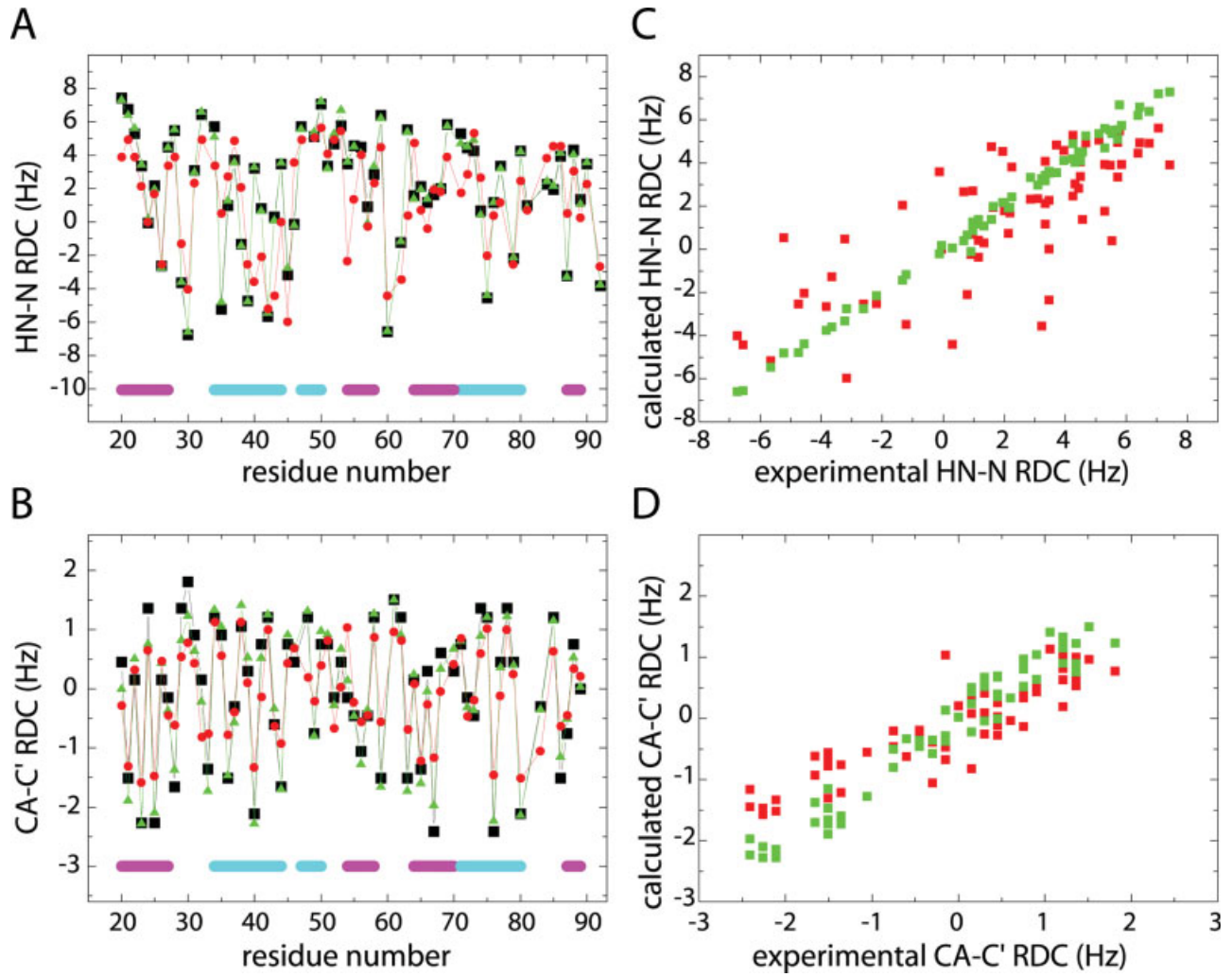


Figure 3

RDC study of the RH-II/Guβ GUCT domain. The two panels on the left show the superposition of the backbone HN-N (**A**) and the Cα-C' (**B**). RDC profiles obtained experimentally (black), those calculated from the NOE-based model (red), and those calculated from the NOE-RDC-refined model (green). RDC calculations from the structural models were performed using the program PALES.[29] Cyan and magenta bars represent α-helices and β-strands, respectively. The two panels on the right show the RDC correlations for the HN-N dipoles (**C**) and the Cα-C' dipoles (**D**) between the experimental values and those calculated from the NOE-based model (red) as well as those from the NOE-RDC-refined model (green).

NMR signals. The constructs G571-K660, G571-Y667, and G557-Y667 showed well-separated, but partially line-broadened HSQC spectra, whereas F587-A659, F575-V657, and F575-A659 showed better ^{1}H-^{15}N HSQC spectra with reasonable line widths. We chose the longest of the three good constructs, F575-A659, for the subsequent structural study. Figure 2 shows a ^{1}H-^{15}N HSQC spectrum of this construct with resonance assignments.

Structural determination by NMR

The standardized NMR data analyses and structural calculations, using the programs KUJIRA[12] and CYANA,[19] accomplished 92% complete assignments for all nonlabile and backbone amide ^{1}H chemical shifts, excluding the artificial tail region, and assignments for 99.5% of the NOE peaks, which yielded 1321 meaningful NOE distance restraints. The 20 structures calculated with CYANA were well defined for the residues from K585 to A659, and we calculated ^{1}H-^{15}N and Cα-C' RDC values for this ordered region using PALES.[29] Figure 3 shows a comparison between the calculated RDC values and the experimental RDC values. These RDC datasets correlated with Pearson correlation coefficients of $r = 0.76$ ($n = 65$, $P < 0.0001$) for the ^{1}H-^{15}N RDC set and $r = 0.89$ ($n = 64$, $P < 0.0001$) for the Cα-C' RDC set.

Table I
NMR Restraints and Structure Calculation Statistics

Quantity	Value
NOE distance restraints	
Total	1314
Shortrange ($\|i-j\| \leq 1$)	665
Mediumrange ($1 < \|i-j\| < 5$)	181
Longrange ($\|i-j\| \geq 5$)	468
Maximal violation (Å)	0.11
Torsion angle restraints	
ϕ	48
ψ	48
Maximal violation (°)	0.53
RDC restraints[a]	
Average correlation	0.983 ± 0.0005
Average Q	0.183 ± 0.003
Average axial component value	−5.81 ± 0.02
Average rhombicity value	0.185 ± 0.003
Maximum violation (Hz)	0.91 ± 0.06
CYANA target function value (Å^2)	0.16 ± 0.01
AMBER energies (kcal/mol)	
Total	−3114 ± 96
van der Waals	−243 ± 9
Electrostatic	−3472 ± 195
Ramachandran plot (%)[b]	
Residues in most favored regions	89.6
Residues in additional allowed regions	9.9
Residues in generously allowed regions	0.5
Residues in disallowed regions	0.0
RMSD from the averaged coordinates (Å)[b]	
Backbone atoms	0.38
Heavy atoms	0.78

Where applicable, the values given correspond to the average and standard deviation over the 20 conformers that represent the solution structure.
[a]Computed before energy refinement.
[b]For residues G586-A659.

Subsequently, we refined the NOE-based NMR structure by incorporating the experimental RDCs as restraints into the combined automated NOE assignment and structural calculation with CYANA, which yielded assignments for 99.7% of the NOE peaks and 1314 meaningful NOE distance restraints. This RDC-refined structure showed excellent agreement with the experimental RDC datasets (see Fig. 3); the Pearson correlation coefficients for the ^{1}H-^{15}N and CA-C' RDC datasets were $r = 0.99$ and $r = 0.98$, respectively. We further refined the structure by restrained energy minimization in explicit solvent using OPALp.[25,26] The resultant structures maintained excellent agreement with the experimental RDC datasets (see Fig. 3); the Pearson correlation coefficients for the ^{1}H-^{15}N and Cα-C' datasets were $r = 0.94$ and $r = 0.97$, respectively. The statistics for the structural calculations are summarized in Table I.

Structure of the GUCT domain

The solution structure of the human RH-II/Guβ GUCT domain construct, F575-A659, displays a disordered N-terminal region (F575-D584) with an artificial cloning tag and a well-converged region (K585-A659) with β-α-α-β-β-α-β topology (see Fig. 4). NOE peaks were not detected between the N-terminal disordered region and the folded region, suggesting that the boundary for the GUCT domain can be defined based on this folded region. The present structure revealed that the first β-strand starts with F587, which is consistent with the ^{1}H-^{15}N HSQC-screening result, where the construct F587-A659 showed a good number of HSQC peaks (Supplementary Fig. S1). In remarkable contrast, the construct T589-A659 showed poor signals due to sample precipitation, suggesting that the presence of the N-terminal two residues of the first β-strand is critical for the stability of the GUCT domain. Thus, we experimentally defined the domain boundary for the GUCT domain as K585-A659.

The GUCT domain displays a fairy acidic surface, particularly on the β-sheet face [Fig. 4(C)]. As discussed later, this acidic feature would be relevant to its functional role.

The structural neighbors of the GUCT domain

The present structure (see Fig. 4) is the first tertiary structure determined for the GUCT domain. A structural homologue search with the presently determined coordinates for K585-A659 in the PDB, using the Dali server (http://www.ebi.ac.uk/dali/), detected 28 entries with *Z*-scores above 5.0, most of which are typical RRM domain family members. RRM stands for RNA recognition motif, which is also known as the RBD or RNP domain. Table II lists the hits with high *Z*-scores (above 5.6).

DISCUSSION

Updated domain boundary for the GUCT domain

The GUCT domain is found in the C terminal part of the RNA helicase II/Gu protein family members [Fig. 1(A)], and that of the human RH-II/Guβ protein was defined as G571-E666 by the Pfam HMM algorithm. In this study, we have determined the solution structure of this GUCT domain and found that only the region K585-A659 is well defined. Thus, we propose a new domain boundary for the GUCT domain with this structurally converged region. Using this new definition, we performed PSI- and PHI-BLAST searches against the UniProt database with an *E*-value cutoff of 0.1, which yielded 40 nonredundant entries. On the other hand, another BLAST search using the Pfam HMM definition for the GUCT domain, G571-E666, yielded 47 nonredundant entries. The difference in the number of detected entries can be attributed to the difference in the domain boundary definition. As summarized in Figure 5(A), seven entries (RH3_ORYSJ, RH3_ARATH,

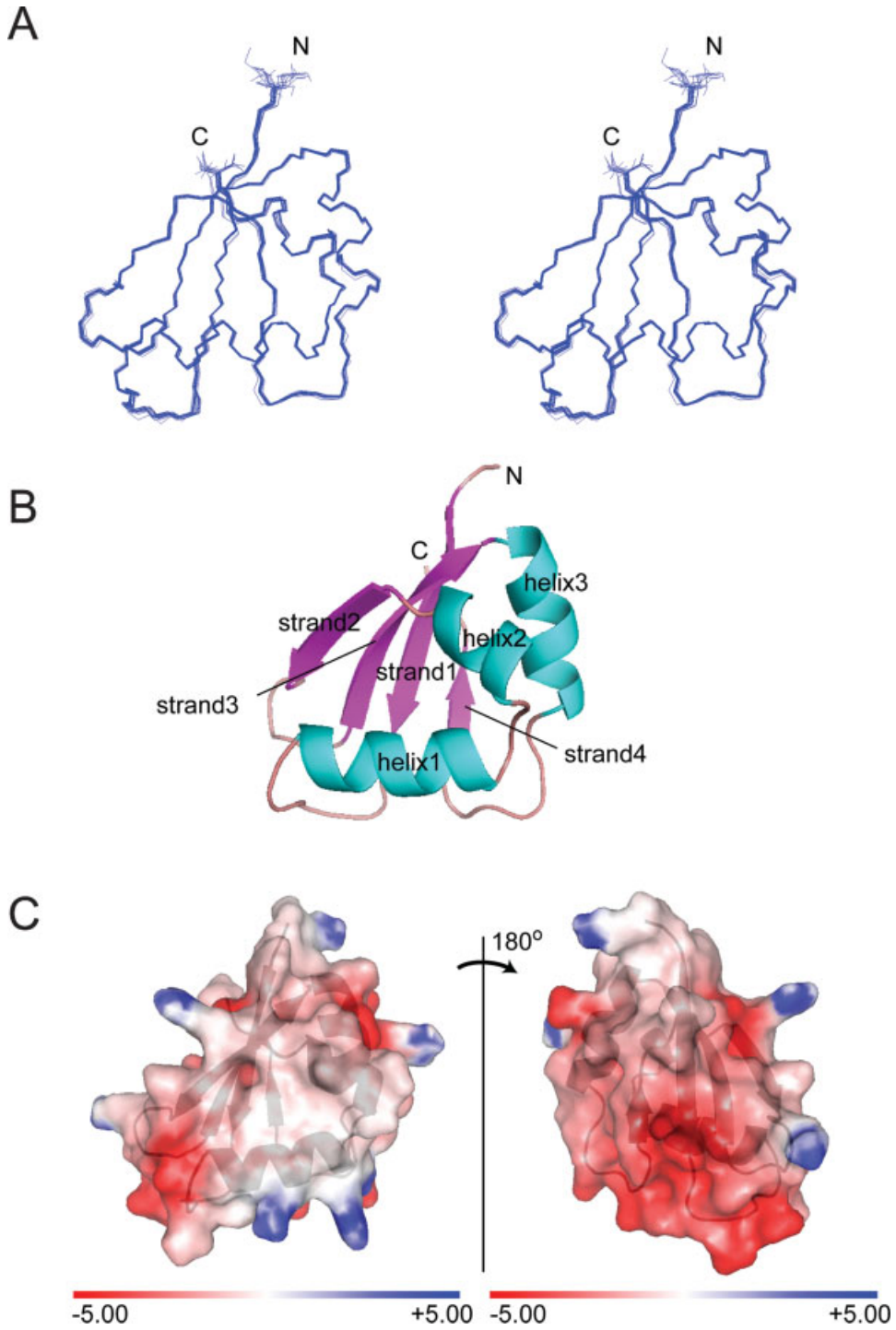


Figure 4

Solution structure of the human RH-II/Guβ GUCT domain. **A:** The stereoview shows the backbone structure of the region between K585 and A659. The 20 lowest-energy conformers were superimposed by fitting the backbone heavy atoms in this region. **B:** Ribbon representation of the minimized average structure of the GUCT domain, K585-A659. The secondary structural elements and the N- and C-termini are labeled. **C:** Electrostatic potential surface of the structured region (K585-A659) of the GUCT domain. The electrostatic potential distribution was calculated using GRASP,[30] and the negative and positive potential values (from −5.0 to 5.0 kT/e) are shown in red and blue, respectively. The left panel is drawn with the same view angle as those used in the panels A and B.

Table II
Structural Homologues of the Human RH-II/Guβ GUCT Domain

Protein name	PDB ID	Chain	Pfam entry name	*Z*-score[a]	No. of residues[b]	RMSD[c]
Aspartokinase	2HMF	A	ACT	6.9	73	2.8
Ribonuclease P protein component 2	2AV5	A	RNase_P_Rpp14	6.5	73	2.7
Nonsense transcripts 3B	1UW4	A	MIF4G	6.3	64	2.1
Conserved hypothetical protein	2FPH	X	null	6.1	68	2.5
Lupus La protein	1S79	A	RRM_1	5.9	68	2.7
ATP-dependent RNA helicase dbpA	2G0C	A	RRM_1	5.9	66	2.9
Heterogeneous nuclear ribonucleoprotein A1	1HA1	A	RRM_1	5.9	66	2.5
Conserved hypothetical protein	2F06	A	ACT	5.8	67	2.6
Thiamine biosynthesis protein thil	2C5S	A	THUMP	5.8	67	2.6
Nuclear cap-binding protein subunit 2	1H6K	Z	RRM_1	5.8	65	2.5
CG18350-PI	1B7F	A	RRM_1	5.7	67	2.7
Nucleolin (protein C23)	1FJ7	A	RRM_1	5.6	70	2.7

[a]Z-score computed by the dali server.
[b]Number of residues in the structural alignment computed in the Dali search.
[c]Positional root mean square deviation of superimposed Cα atoms in Ångstroms, based on the dali search.

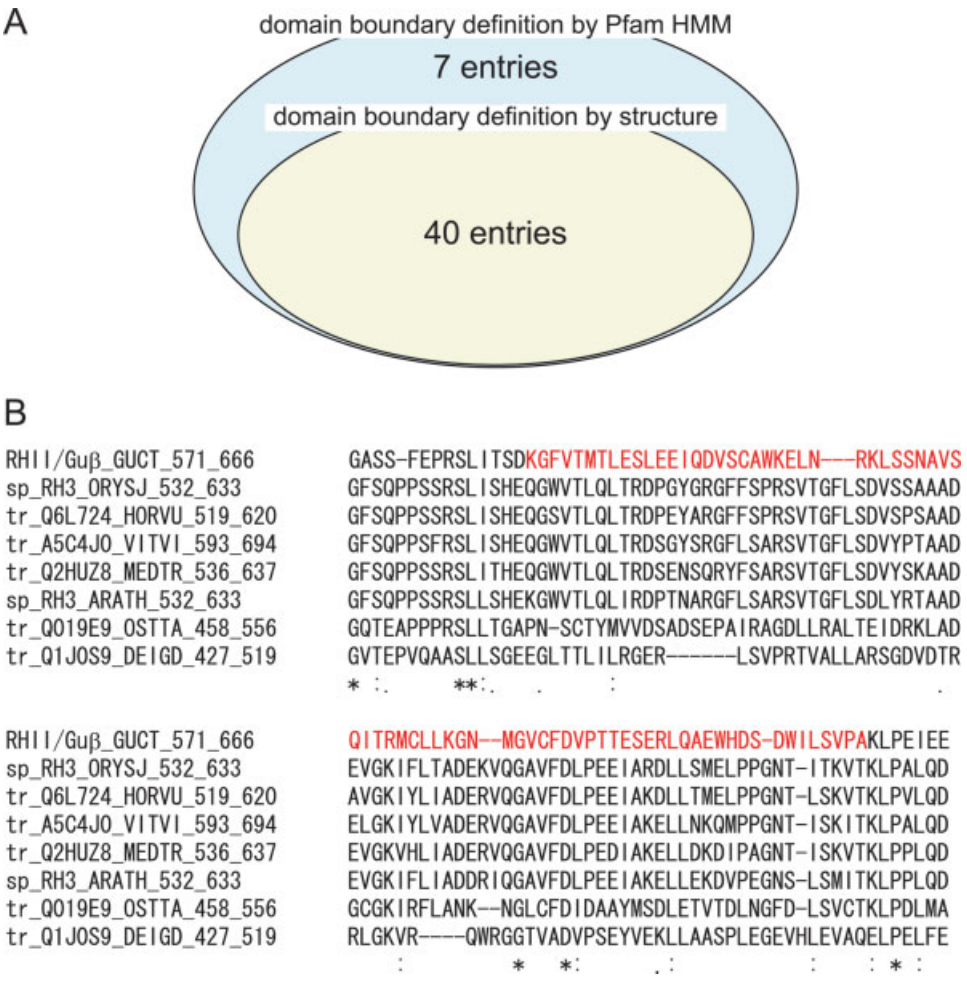


Figure 5
Sequence homologue search of the RHII/Guβ GUCT domain in the UniProt database. **A:** Venn diagram, representing the number of entries, was found by a PSI-BLAST search using the query of the GUCT domain region (G571-E666), as defined by the Pfam HMM profile and that using the query of the structured region (K585-A659). Redundant sequences were excluded. **B:** Alignment of the RHII/Guβ GUCT domain sequence and those of the entries that were found only when the Pfam HMM domain boundary definition was used. The structured region in the GUCT domain is highlighted in red.

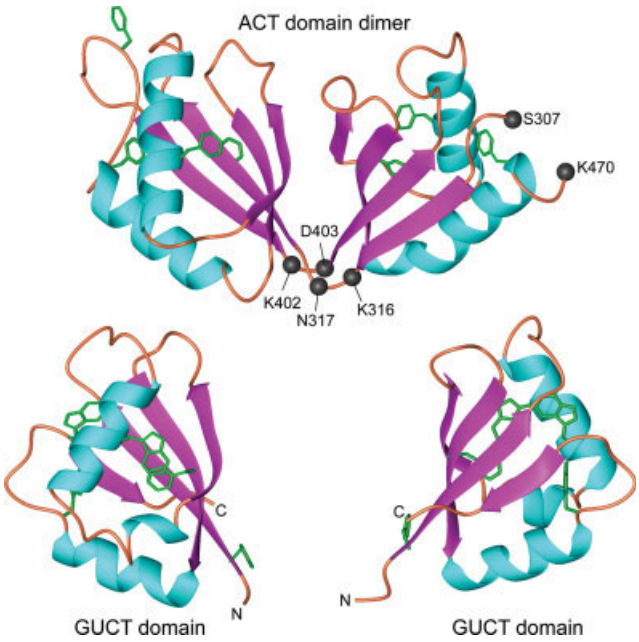


Figure 6

Structural comparison between the ACT domain from aspartokinase and the human RH-II/Guβ GUCT domain. Top panel shows the dimer of the ACT domain (PDB code: 2HMF chain A, S307-K470).[31] The bottom two panels show two GUCT domain molecules with the orientations comparable to those of the ACT domains in the top panel. Side chains of aromatic residues (W, Y, and F) are displayed.

A5C4J0_VITVI, Q019E9_OSTTA, Q1J0S9_DEIGD, Q2HUZ8_MEDTR, and Q6L724_HORVU) were found only in the search with the longer domain boundary definition. They all belong to one of the three primitive classes in the GUCT domain phylogenic tree [Fig. 1(C)]. Figure 5(B) shows the sequence comparison for these entries with respect to the RH-II/Guβ GUCT sequence. Interestingly, they share high sequence similarity in the linker regions outside of the structurally converged part and low similarity in the structured region. Therefore, these entries are potentially false positives detected by using the longer domain boundary definition.

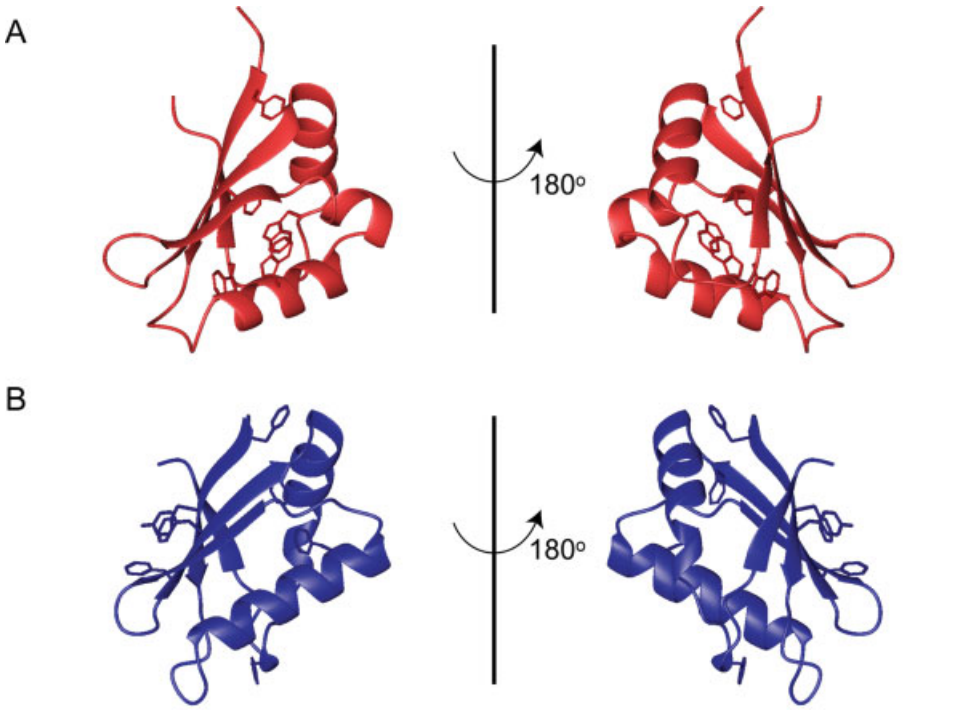
Structural homologue search

Because this is the first structural report of the GUCT domain, it is intriguing to examine the uniqueness of this protein fold. The structural homologue search using the Dali server provided a dozen hits with *Z*-scores above 5.0. As shown in Table II, a region of a threonine sensitive aspartokinase from *Methanococcus jannaschii* (PDB code: 2HMF),[31] referred to as the ACT domain in the Pfam definition, showed the best match with the present GUCT structure with the highest *Z*-score of 6.9. The ACT domain has the β-α-β-β-α-β topology and is proposed to be a conserved regulatory ligand-binding fold, which is found in a wide range of metabolic enzymes that are regulated by particular amino acids.[32] The aspartokinase consists of two ACT domains with an irregular orientation of the β-strand: the first domain consists of A319-K402, while the second domain is formed with residues S307-V318 and D403-K470 as shown in Figure 6. Although both of the ACT domains and the GUCT domain have a similar topology of the aromatic residues in the hydrophobic core, such an irregular orientation of the secondary structure may imply a distant evolutional relationship of these domains.

The majority of the structural homologues detected in the Dali search, including RNase_P_Rpp14,[33] MIF4G,[34] and THUMP[35] are known members of the RRM-domain family. Figure 7(A,B) shows the structural comparison between the GUCT domain and the RRM domain from nucleolin[36] (*Z*-score 5.6). The backbone traces of the RRM domain and the GUCT domain resemble well with deviations mainly around the last helix-loop-strand region. A comparison of the side-chain topology, however, revealed significant differences in the packing geometry of the aromatic side-chains that form the hydrophobic core. As shown in the sequence comparison between the GUCT domain and the canonical RRM_1 domain family members, the conservation of the aromatic residues is low [Fig. 7(C)]. Thus, the GUCT domain is an atypical member of the RRM fold superfamily.

Functional implications

The RH-II/Guα protein has been well characterized, in terms of its intriguing bifunctional abilities.[1–5] The DEAD domain and the Helicase_C domain are responsible for the RNA-unwinding activity, where as the ~70 residue C-terminal segment is responsible for the RNA-foldase activity. However, the region including the GUCT domain is not necessary for either the helicase or foldase activity, and the functional role of the GUCT domain has thus been unclear. Through the present structural determination, we found that the GUCT domain has the RRM fold, which is known as an RNA recognition motif. However, the RH-II/Guβ GUCT domain lacks the surface-exposed aromatic residues and basic residues on the second and third β-strands that are important for RNA recognition in the canonical RRM domains.[37] In addition, it has a fairly negative electrostatic potential map on the molecular surface, with a remarkably acidic β-sheet face [Fig. 4(C)]. A homology model for the RH-II/Guα GUCT domain, based on the present GUCT domain structure, presents a similar acidic molecular surface without the exposed aromatic residues on the β-sheet (data not shown). These findings suggest that the RH-II/Gu GUCT domains do not primarily recognize RNA molecules.



C

```
RHII/Guβ_GUCT/585-659     KGFVTMTLESLEEIQD---VSCAWKELNRKLSSNAVSQITRMCLLKGNM
NUCL_HUMAN/305-383        TAFNLF-VGNLNFNKSAPELKTGISDVF-AKND--LAVVDVR-IGMTRK
RBY1A_HUMAN/6-85          HPGKLF-IGGLNRETNEKMLKAVFGKHG-PISE--VLLIKDR-TSKSRG
HNRGT_HUMAN/6-86          RPGKLF-IGGLNLETDEKALEAEFGKYG-RIVE--VLLMKDRETNKSRG
TIA1_HUMAN/104-184        NHFHVF-VGDLSPQITTEDIKAAFAPFG-RISD--ARVVKDMATGKSKG
ELAV4_HUMAN/44-124        SKTNLI-VNYLPQNMTQEEFRSLFGSIG-EIES--CKLVRDKITGQSLG
ROAA_HUMAN/67-144         DAGKMF-VGGLSWDTSKKDLKDYFTKFG-EVVD--CTIKMDPNTGRSRG
ZCRB1_HUMAN/8-88          SKSTVY-VSNLPFSLTNNDLYRIFSKYG-KVVK--VTIMKDKDTRKSKG
SNRPA_HUMAN/206-282       PNHILF-LTNLPEETNELMLSMLFNQFP-GFKE--VRLVPGR-----HD
SFRS1_HUMAN/14-91         NDCRIY-VGNLPPDIRTKDIEDVFYKYG-AIRD--IDLKNRR---GGPP
RBM10_HUMAN/127-209       ASNIVM-LRMLPQAATEDDIRGQLQSHGVQARE--VRLMRNKSSGQSRG
                                +                                       +

RHII/Guβ_GUCT/585-659     GVCFDVP-TTESERLQAE-----WHDSD-WILSVPA-
NUCL_HUMAN/305-383        FGYVDFESAEDLEKAL-ELTGLKVFGN-EIKLEKPKG
RBY1A_HUMAN/6-85          FAFITFENPADAKNAAKDMNGKSLHGK-AIKVEQAKK
HNRGT_HUMAN/6-86          FAFVTFESPADAKAAARDMNGKSLDGK-AIKVAQATK
TIA1_HUMAN/104-184        YGFVSFFNKWDAENAIQQMGGQWLGGR-QIRTNWATR
ELAV4_HUMAN/44-124        YGFVNYIDPKDAEKAINTLNGLRLQTK-TIKVSYARP
ROAA_HUMAN/67-144         FGFILFKDAASVEKVLDQKE-HRLDGR-VIDPKKAMA
ZCRB1_HUMAN/8-88          VAFILFLDKDSAQNCTRAINNKQLFGR-VIKASIAID
SNRPA_HUMAN/206-282       IAFVEFDNEVQAGAARDALQGFKITQNNAMKISFAKK
SFRS1_HUMAN/14-91         FAFVEFEDPRDAEDAVYGRDGYDYDGY-RLRVEFPRS
RBM10_HUMAN/127-209       FAFVEFSHLQDATRWMEANQHSLNILGQKVSMHYSDP
                          + +
```

Figure 7

Comparison between the RH-II/Guβ GUCT domain and the typical RRM domain family members. **A:** Front view and back view of the GUCT domain (red, A). Side chains of aromatic residues (Tyr, Phe, and Trp) are displayed. **B:** Front view and back view of the nucleolin RRM domain (blue; B, PDB code: 1FJ7).[36] Side chains of aromatic residues (Tyr, Phe, and Trp) are displayed. **C:** Sequence comparison between the present GUCT domain and arbitrarily selected RRM domain family members from human origin, including the nucleolin RRM domain (NUCL_HUMAN). The sequence of the GUCT domain was manually aligned against the CLUSTAL W[9] alignment of the canonical members by the positions of the secondary structures. Cyan and magenta residues denote those in α-helices and β-sheets, respectively. Key residues in the RNA recognition by the RRM domains are marked with the plus symbols.

The theoretical pI values calculated from the sequences for the human RH-II/Guβ DEAD domain, the Helicase_C domain, and the C-terminal tail region are 9.0, 9.2, and 12.0, respectively, while that for the GUCT domain is 4.8. The same tendency is observed for RH-II/Guα. Thus, the GUCT domain may interact with basic pro-

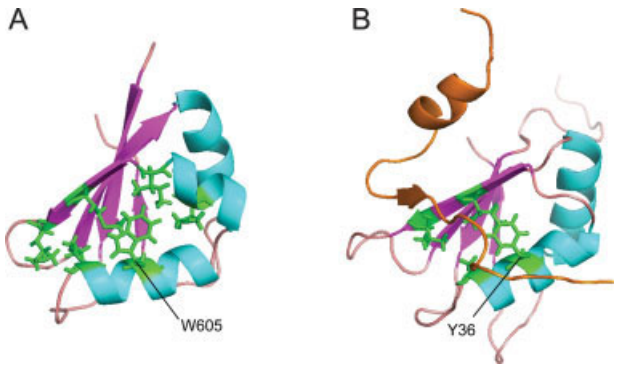


Figure 8

Comparison between the RH-II/Guβ GUCT domain and the p14 RRM domain. **A:** Characteristic hydrophobic residues (green) in the GUCT domain that are exposed to the solvent. **B:** The complex structure of the human p14 RRM domain and a peptide from the human spliceosomal protein SF3b155 (orange, PDB code: 2FHO).[38] Key hydrophobic residues used for the peptide binding are colored green.

teins, such as the DEAD domain, the Helicase_C domain, and/or the C-terminal tail.

We note that the RH-II/Guβ GUCT domain has the characteristic hydrophobic surface around the surface-exposed tryptophan and W605 [Fig. 8(A)]. Good similarity exists between this characteristic site and the peptide-binding interface of the RRM domain from the p14 protein,[38] in which the surface-exposed tyrosine, Y36 plays a key role [Fig. 8(B)]. Therefore, we propose that the GUCT domain functions in protein binding.

CONCLUSIONS

The GUCT domain was only defined with the characteristic sequence pattern detected by the Pfam HMM algorithm. In this study, we have determined the solution structure of the human RH-II/Guβ GUCT domain, which updated the domain boundary definition and provided some functional implications. The structure resembles the RRM domain fold, where as the electron potential distribution of the molecular surface suggests that the GUCT domain interacts with proteins, rather than nucleic acids. A potential candidate for its binding target is the DEAD domain, the Helicase_C domain, and/or the C-terminal tail of RH-II/Guβ. In addition, the structural determination allowed us to redefine the domain boundary, which revealed the false positives and negatives in the sequence search for the GUCT domain homologues.

ACKNOWLEDGMENTS

The authors are grateful to M. Sato, Y. Ide, and A. Hiyoshi for sample preparation. S.O. thanks K.B. Briggman for technical advice in RDC measurements.

REFERENCES

1. Valdez BC. Structural domains involved in the RNA folding activity of RNA helicase II/Gu protein. Eur J Biochem 2000;267:6395–6402.
2. Valdez BC, Henning D, Perumal K, Busch H. RNA-unwinding and RNA-folding activities of RNA helicase II/Gu—two activities in separate domains of the same protein. Eur J Biochem 1997;250:800–807.
3. Henning D, So RB, Jin R, Lau LF, Valdez BC. Silencing of RNA helicase II/Guα inhibits mammalian ribosomal RNA production. J Biol Chem 2003;278:52307–52314.
4. Valdez BC, Henning D, Busch RK, Woods K, Flores-Rozas H, Hurwitz J, Perlaky L, Busch H. A nucleolar RNA helicase recognized by autoimmune antibodies from a patient with watermelon stomach disease. Nucleic Acids Res 1996;24:1220–1224.
5. Ou Y, Fritzler MJ, Valdez BC, Rattner JB. Mapping and characterization of the functional domains of the nucleolar protein RNA helicase II/Gu. Exp Cell Res 1999;247:389–398.
6. Zhao R, Shen J, Green MR, MacMorris M, Blumenthal T. Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export. Structure 2004;12:1373–1381.
7. Valdez BC, Yang H, Hong E, Sequitin AM. Genomic structure of newly identified paralogue of RNA helicase II/Gu: detection of pseudogenes and multiple alternatively spliced mRNAs. Gene 2002;284:53–61.
8. Valdez BC, Perlaky L, Henning D. Expression, cellular localization, and enzymatic activities of RNA helicase II/Guβ. Exp Cell Res 2002;276:249–263.
9. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
10. Tochio N, Umehara T, Koshiba S, Inoue M, Yabuki T, Aoki M, Seki E, Watanabe S, Tomo Y, Hanada M, Ikari M, Sato M, Terada T, Nagase T, Ohara O, Shirouzu M, Tanaka A, Kigawa T, Yokoyama S. Solution structure of the SWIRM domain of human histone demethylase LSD1. Structure 2006;14:457–468.
11. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
12. Kobayashi N, Iwahara J, Koshiba S, Tomizawa T, Tochio N, Guntert P, Kigawa T, Yokoyama S. KUJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. J Biomol NMR 2007;39:31–52.
13. Johnson BA, Blevins RA. NMRVIEW: a computer program for the visualization and analysis of NMR data. J Biomol NMR 1994;4:603–614.
14. Hansen MR, Mueller L, Pardi A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 1998;5:1065–1074.
15. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson 1998;131:373–378.
16. Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 1990;89:496–514.
17. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 2002;24:171–189.
18. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
19. Guntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol 2004;278:353–378.
20. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–302.

21. Clore GM, Gronenborn AM, Bax A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. J Magn Reson 1998;133:216–221.
22. Skrynnikov NR, Kay LE. Assessment of molecular structure using frame-independent orientational restraints derived from residual dipolar couplings. J Biomol NMR 2000;18:239–252.
23. Losonczi JA, Andrec M, Fischer MW, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson 1999;138:334–342.
24. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 1997;4:732–738.
25. Koradi R, Billeter M, Güntert P. Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun 2000;124:139–147.
26. Luginbuhl P, Guntert P, Billeter M, Wuthrich K. The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 1996;8:136–146.
27. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
28. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
29. Zweckstetter M, Hummer G, Bax A. Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. Biophys J 2004;86:3444–3460.
30. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
31. Faehnle CR, Liu X, Pavlovsky A, Viola RE. The initial step in the archaeal aspartate biosynthetic pathway catalyzed by a monofunctional aspartokinase. Acta Crystallogr Sect F Struct Biol Cryst Commun 2006;62:962–966.
32. Chipman DM, Shaanan B. The ACT domain family. Curr Opin Struct Biol 2001;11:694–700.
33. Wilson RC, Bohlen CJ, Foster MP, Bell CE. Structure of Pfu Pop5, an archaeal RNase P protein. Proc Natl Acad Sci USA 2006;103:873–878.
34. Kadlec J, Izaurralde E, Cusack S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nat Struct Mol Biol 2004;11:330–337.
35. Waterman DG, Ortiz-Lombardia M, Fogg MJ, Koonin EV, Antson AA. Crystal structure of *Bacillus anthracis* ThiI, a tRNA-modifying enzyme containing the predicted RNA-binding THUMP domain. J Mol Biol 2006;356:97–110.
36. Allain FH, Gilbert DE, Bouvet P, Feigon J. 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 2000;303:227–241.
37. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 2005;272:2118–2131.
38. Kuwasako K, Dohmae N, Inoue M, Shirouzu M, Taguchi S, Guntert P, Seraphin B, Muto Y, Yokoyama S. Complex assembly mechanism and an RNA-binding mode of the human p14-SF3b155 spliceosomal protein complex identified by NMR solution structure and functional analyses. Proteins 2007;71:1617–1636; DOI: 10.1002/prot.21839.