UNIVERSITÄT BASEL

Investigation of the guanine quadruplex resolving activity of the DEAH-box RNA helicase RHAU

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

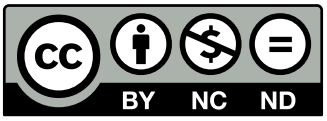
von

Simon Lattmann
aus Hütten (ZH), Schweiz

Leiter der Arbeit
Dr. Yoshikuni Nagamine

FMI
Friedrich Miescher Institute
for Biomedical Research

Basel, 2012



Genehmigt von der Philosophisch-Naturwissenschftlichen Fakultät auf Antrag von:

Prof. Dr. Christoph Moroni

Prof. Dr. Susan M. Gasser

Dr. Yoshikuni Nagamine

Basel, den 11. Dezember 2012 Prof. Dr. Jörg Schibler, Dekan

UNI
BASEL

«Felix, qui potuit rerum cognoscere causas»

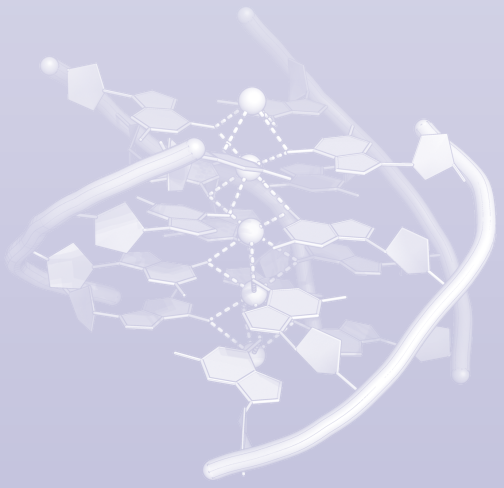
Virgile (29 BC)

Summary

Generally depicted as single- or double-stranded molecules, nucleic acids sequences can adopt various forms of stable secondary structures. In particular, guanine-rich sequence of DNA and RNA can form atypical four-stranded helical structures termed G-quadruplexes (G4). While the functional relevance of G4 structures is still a mater of debate, an increasing number of evidence suggests that these structures can form in various regions of the genome and may be implicated in a wide array of processes such as gene expression regulation and telomere protection. Owing to their high-thermodynamic stability, *in vivo* conversion of G4 structures to single-stranded nucleic acid requires specialised proteins with G4 destabilising/unwinding activity. RHAU is a human RNA helicase of the DEAH-box family that exhibits ATP-dependent G4 resolving activity with high affinity and specificity for its substrate *in vitro*. However, how RHAU recognises G4 and what are its substrates in cells are key questions that needed to be addressed.

In the first part of this research work, we undertook to address the molecular mechanisms underlying the specific recognition of G4 structures by RHAU. Through biochemical analysis of truncated and mutated recombinant forms of RHAU, we have uncovered the functional importance of the amino-terminal region for interaction with G4 structures and further identified within this region the evolutionary conserved RSM (RHAU-specific motif) domain as a major affinity and specificity determinant. We also show that the G4-RNA substrate specificity and resolving activity shown by RHAU is an evolutionary conserved attribute in higher eukaryotes, insofar as CG9323, the *Drosophila* orthologue of RHAU, binds and readily unwinds G4 structures.

In the second part of this work, we sought RNAs bound by RHAU in living cells. To this end, we employed high-throughput gene array technologies to identify RNAs associated with RHAU on a genome-wide scale. Approximately 100 RNAs were found to be significantly enriched with RHAU. Computational analysis of RNA sequences for potential intramolecular G4 structures revealed the preferential association of



RHAU with transcripts bearing G4-forming motifs, suggesting direct targeting of G4-RNAs by RHAU. Among the most abundant RNAs selectively enriched, we identified the human telomerase RNA template TERC as a *bona fide* target of RHAU. Remarkably, binding of RHAU to TERC depended on the presence of a stable G4 structure in the 5'-region of TERC, both *in vivo* and *in vitro*. In-depth studies further revealed that RHAU was also part of the telomerase holoenzyme through direct interaction with TERC G4 structure. Collectively, these data provide the first evidence of a specific and direct interaction between a G4 resolvase enzyme and a potentially relevant intramolecular G4-RNA substrate, and more generally support the idea that intramolecular G4-RNAs are naturally occurring substrates of RHAU. Furthermore, these results provide circumstantial evidence for the existence of a G4-RNA structure in a fraction of the telomerase holoenzyme.

Overall, the present work brings new insight into the mechanisms of G4 substrate recognition by RHAU and supports its potential role as a G4 resolvase enzyme *in vivo*.

Contents

CHAPTER 1

Introduction

RNA helicases

The RNA folding problem

RNAs constitute essential structural and multifunctional components of the gene expression machinery. RNAs are highly polymorphic molecules that can function as transitory carriers of genetic information (mRNAs), catalysts of biochemical reactions (23S rRNA, ribonuclease P RNA), adapter molecules (tRNAs, snoRNAs, miRNAs), or as structural molecules in ribonucleoproteins (rRNAs, ref. 1). In cells, most of RNAs have to fold into well-defined structures to be biologically active. However, due to the single-strandedness of nascent RNA, the relative simplicity of its molecular composition and its possibility to form non-Watson-Crick base-pairing, RNA molecules are prone to adopt a multitude of non-functional and thermodynamically stable conformations. Moreover, even during their normal metabolic process, many RNAs undergo conformational changes or are transiently base-paired with other RNA species. Rearrangement of intra- or intermolecular base-pairings *in vivo* is achieved by RNA chaperones that facilitate conformational changes of RNA to its active form (for reviews, see ref. 2,3). Among these chaperones are the RNA helicases, which couple the hydrolysis of nucleotide 5'-triphosphates (NTPs) with structural and functional rearrangement of the RNA (4). RNA helicases represent a large group of proteins that have been identified in all biological systems, viruses included. Functionally, they have been shown to disrupt RNA–RNA or RNA–DNA duplexes (5,6) and to dissociate proteins from RNA molecules (7,8). In eukaryotes, RNA helicases form by far the largest group of proteins dedicated to RNA metabolism (9).

RNA helicases and the superfamily of helicases

RNA helicases (EC 3.6.4.13) belongs to the helicase class of enzymes that also includes DNA helicases. Helicases are NTP-driven molecular motors that dissociate double-stranded nucleic acids or displace nucleic acid bound proteins. A large fraction of the eukaryotic and prokaryotic genomes encode helicases. In the yeast *Saccharomyces cerevisiae*, about two percents of the protein-encoding genes encode helicase-related proteins (10) and nearly one percent of the open reading frames of the human genome encodes putative DNA and RNA helicases (11). Helicases are characterised by the presence of several conserved and discrete motifs in the core region that are involved in NTP-binding/hydrolysis and nucleic acid binding. These motifs essentially form the motor which converts the chemical energy derived from NTP hydrolysis into mechanical force and drives helicase movements leading to the disruption of DNA or RNA base pairs. These conserved motifs provide also the basis for classifying helicases into six phylogenetically distinct superfamilies, designated SF1 to SF6 (12,13). These superfamilies presumably represent evolutionary relationships and may have evolved from a common ancestor. Although this classification is currently still employed, it was established before the availability of structural and functional information. It is now clear that these conserved helicases motifs are present in a wide range of NTP-dependent nucleic acid enzymes (14), many of which are not *bona fide* helicases (15) and some of which do not even seem to translocate along nucleic acids (6). Therefore, in many respects, these motifs more generally denote nucleic acid-stimulated NTPases rather than purely and simply helicases.

Superfamilies 1 and 2 form the largest and most closely related groups of helicases. They include both DNA and RNA helicases and appear to be active essentially as monomers or dimers (16,17). Structurally, the helicase core regions of SF1 and SF2 helicases manifest the same overall folding patterns and consist of two abutting RecA-like domains (18). Additionally, SF1 and SF2 proteins display similarities in both occurrence and sequence composition of their helicase motifs. Both superfamilies are indeed characterised by the presence of at least seven conserved helicase motifs. In contrast, superfamilies 3 to 6 are essentially ring-shaped DNA helicases (13). Most of them are involved in DNA replication by unwinding dsDNA ahead of the DNA polymerase (for reviews, see ref. 19,20). The ring-shape of these helicases results from the assembly of six individual RecA-like domains into a hexameric structure with ssDNA in the center of the ring. Each subunit of hexameric helicases contains only two to five conserved helicases motifs, and ATP-hydrolysis occurs at every interface of the six subunits (13).

Main features of superfamily 2 (SF2) helicases

Most RNA helicases are members of the SF2, a large and versatile superfamily that includes functionally diverse RNA and DNA helicases. Actually, only a few putative RNA helicases, the Upf1-like family as well as some viral RNA helicases belong to the SF1. The SF2 constitutes by far the largest superfamily of helicases. For instance, the human genome encodes 112 SF2 proteins, out of which 65 seem to be dedicated to RNA metabolism (Figure 1). Based on comparative genomic approaches, the SF2 can be further divided into seven phylogenetically distinct families: DEAD-box, DEAH-box, Ski2, RIG-I, RecQ, Rad3 and Swi2/Snf2. Each family is characterised by the presence of nine conserved and family-specific motifs (Q, I, Ia, Ib, and II–VI) that delineate the helicase core region (Figure 2A). Among these seven families, four consist essentially of RNA helicases (DEAD-box, DEAH-box, Ski2 and RIG-I) and three of DNA helicases (RecQ, Rad3 and Swi2/Snf2). The RecQ-

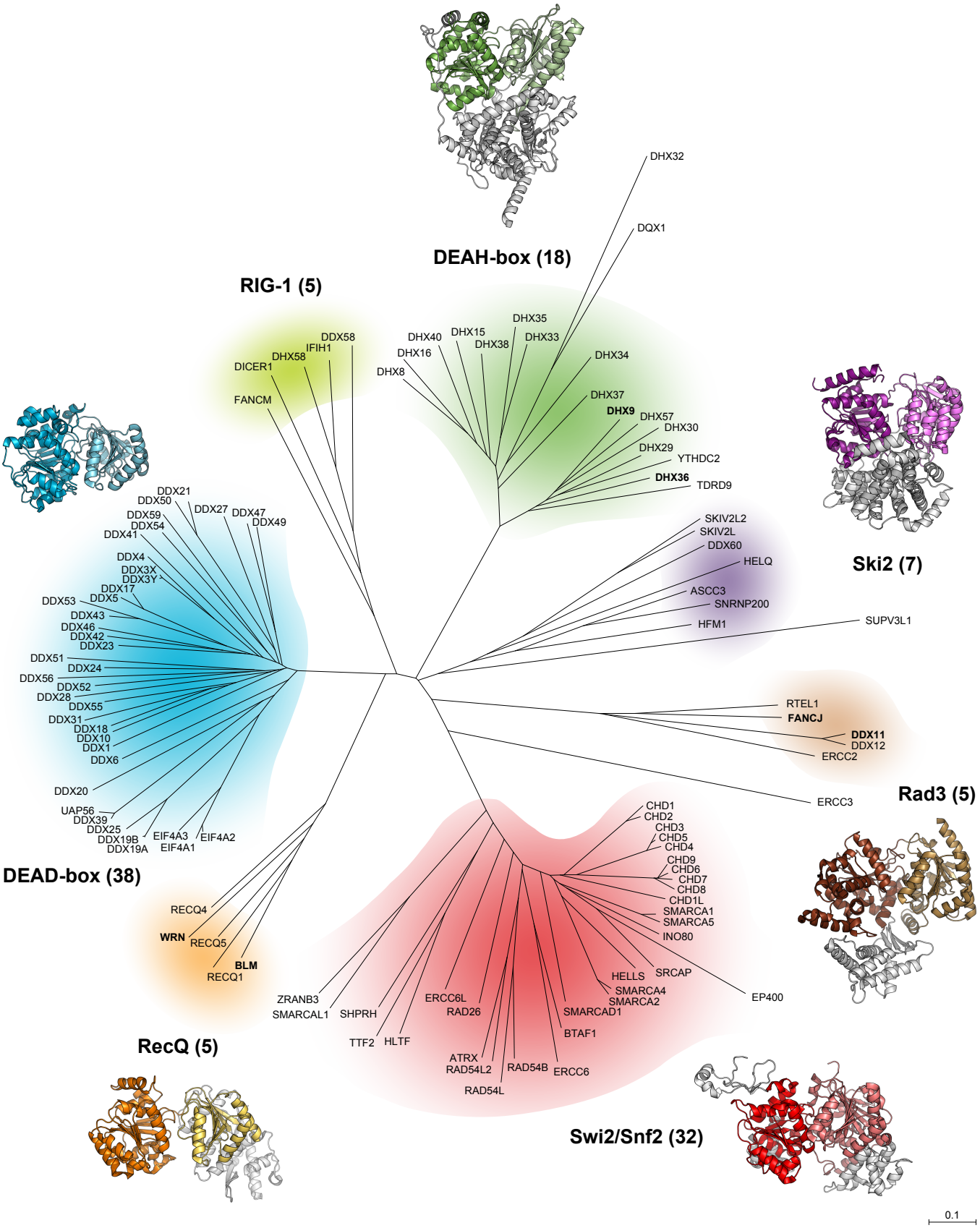


Figure 1 | **Phylogeny of the human superfamily 2 helicases**. The unrooted phyologenetic tree was compute by MAFFT using the neighbor-joining method on 157 ungapped sites (substitution model = JTT; heterogeneity among sites (α) = ∞; bootstrap resampling = 100). When available, a representative X-ray structure is depicted per family (DEAD-box: *Drosophila* Vasa (DDX4), 2DB3; DEAH-box: yeast Prp43 (DHX15), 3KX2; Ski2: *Archaeoglobus fulgidus* Hel308 (HELQ), 2P6R; Rad3: *Thermoplasma acidophilum* XPD (ERCC2), 2VSF; Swi2/Snf2: zebrafish Rad54, 1Z3I; RecQ: human RecQ1 (RECQ1), 2WWY). For each structure, the helicase core region is shown in colour. Boldfaced protein names denote helicases with reported *in vitro* G-quadruplexe resolving activity (WRN, BLM, DDX11, FANCJ, DHX9 and DHX36).

and Rad3-related proteins are *bona fide* helicases which function in replication, recombination and repair of DNA, and are essential for the maintenance of genomic stability (21-23). In contrast, the Swi2/Snf2-related proteins are chromatin-remodelling enzymes that translocate along dsDNA and catalyse chromatin rearrangements important for transcription (15,24).

The SF2 RNA helicases fall into four phylogenetically distinct groups: the DEAD-box, the DEAH-box, the Ski2 and the RIG-I helicases (Figure 1). In almost all eukaryotes, the DEAD-box and DEAH-box proteins constitute by far the two largest families; they usually represent more than half of all RNA helicases. Comparatively, the smaller Ski2-like and RIG-I-like families include only a few members. Besides, although a majority of the DEAD-box proteins take part in ribosome biogenesis (25,26), none of these four helicase families are functionally specialised in a particular domain of RNA metabolism. In fact, the various families of RNA helicases stand out rather by the mechanisms whereby they unwind or translocate along nucleic acids. Notably, DEAD-box proteins are non-processive helicases that catalyse unwinding of weak RNA duplexes by local bending of the nucleic acid (6,27). They may as well facilitate displacement of RNA-bound proteins or function as ATP-dependent clamps to ensure the unidirectionality of reactions (28). In particular, DEAD-box proteins stand out from other RNA helicases in that they do not show any strict unwinding polarity. In contrast, DEAH-box, Ski2-like and RIG-I-like proteins translocate along single-stranded RNAs with a 3'-to-5' polarity (Table III and references therein). They usually manifest as well a better processivity and can resolve longer RNA duplexes than DEAD-box proteins (29,30).

Structurally speaking, while showing low sequence similarities, the helicase core regions of SF2 helicases present a remarkably conserved folding pattern. It consists of two tandemly repeated RecA-like domains (hereafter referred to as RecA1 and RecA2) coupled by a short linker (Figures 1 and 2). The amino-terminal RecA1 domain contains the composite helicase motifs Q, I, Ia, Ib, II and III, while motifs IV, V and VI are located on the carboxy-terminal RecA2 domain. All SF2 helicases bind and hydrolyse NTP at the interface cleft of the two RecA-like domains. Motifs Q, I (aka Walker-A or P-loop), II (aka Walker-B) and VI are directly implicated in NTP-binding/hydrolysis. The other conserved helicases motifs (Ia, Ib, III, IV and V) are less well studied, but evidence suggests that they are involved either in nucleic acid binding or in coupling of the NTP hydrolytic state to protein conformational transitions (31). Besides the helicase core region, most SF2 helicases harbour ancillary N- and C-terminal flanking regions that are essential for the various functions of these enzymes.

SF2 helicase classification and nomenclature

The SF2 helicases are often referred to as the DExD/H-box proteins with regard to the consensus amino acid sequence of their helicase motif II (Figure 2A). Originally, helicases of the SF2 superfamily were classified into various groups termed DEAD-, DEAH- and DExH-box proteins. This classification was done prior to the availability of structural and functional data and the nomenclature of these groups depended essentially on amino acid variations in the helicase motif II. However, phylogenetic, functional and structural studies of human SF2 proteins pointed out that the superfamily should instead be divided into seven distinct families (Figure 1). Examination of motif II consensus sequences within the different families (Figure 2A) reveals that the amino acid composition of motif II alone is not sufficient to assign a given SF2 protein to one of the seven families. The sequence heterogeneity of the Walker-B motif within a certain family of helicases renders the initial Walker-B-based nomenclature somewhat inaccurate. This has led to generate a lot of confusion in classification. To further illustrate

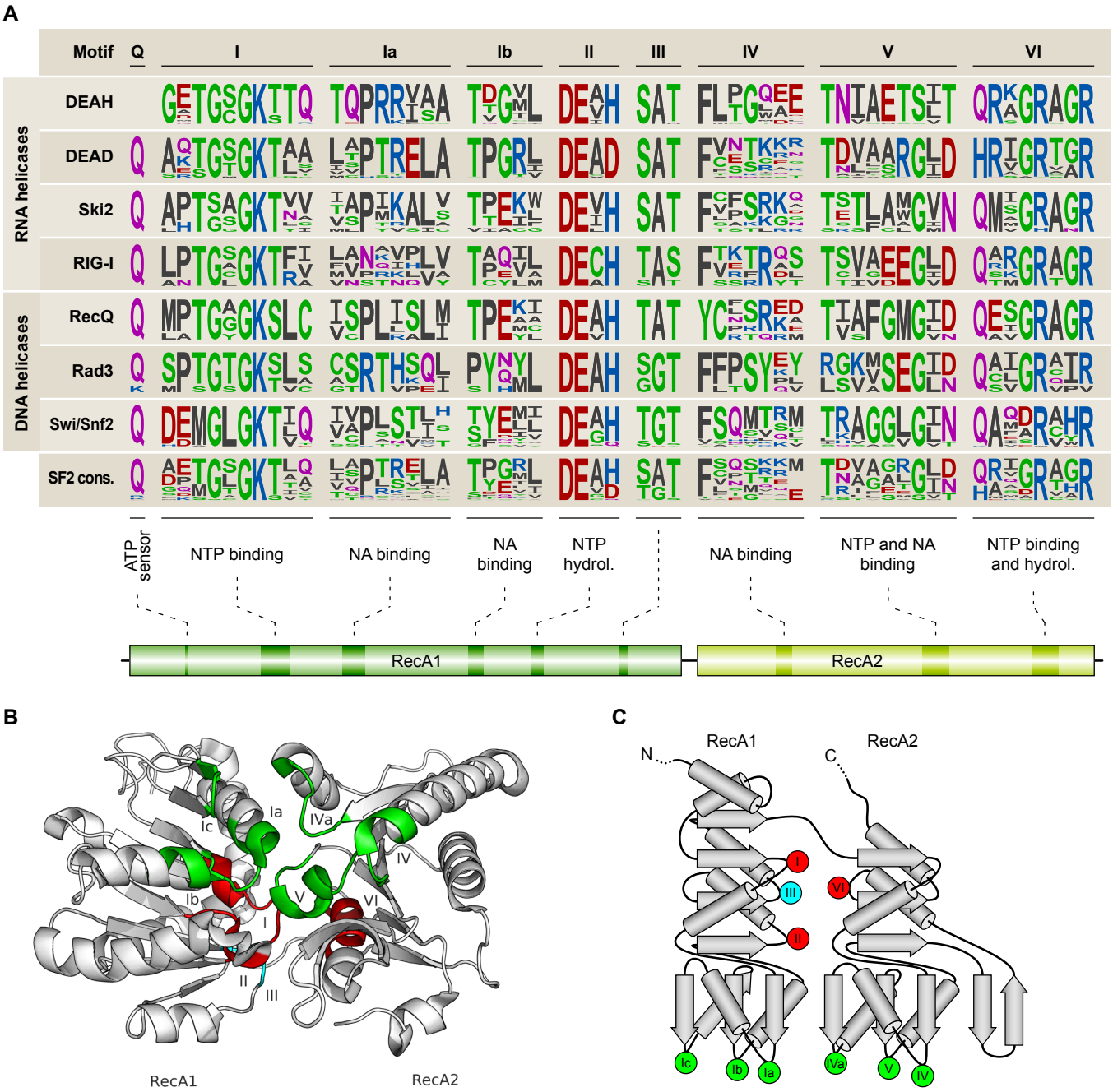


Figure 2 | **Conserved ATPase/helicase motifs in superfamily 2 (SF2) helicases**. (**A**) Sequence logos of the conserved and family-specfic helicase motifs involved in NTP binding/hydrolysis and nucleic acid binding for human SF2 helicases. Amino acids are coloured according to their biochemical properties: green for polar, blue for basic, red for acidic and black for hydrophobic. SF2 cons., SF2 helicase consensus sequence. (**B**) Structure of the helicase core domain of the yeast DEAH-box helicase Prp43 (2XAU, aa. 69–451). The conserved helicase motifs implicated in NTP binding/hydrolysis and nucleic acid binding are shown in red and green, respectively. (**C**) Schematic secondary structure topology of the helicase core of DEAH-box helicases. The position of the conserved helicase motifs is indicated using the same colour scheme as for panel (B).

this aspect, several RNA helicases like RHAU or the RNA helicase A (alias RHA or DHX9), both of which harbour an Asp–Glu–Ile–His motif II, were (and are still) erroneously referred to as DExH-box proteins while their helicase core domains showed extensive similarities with those of DEAH-box proteins like the prototypical yeast Prp2, Prp16, Prp22 and Prp43 RNA helicases (32-34). On the other hand, some RecQ-like and Rad3-like DNA helicases are sometimes incorrectly referred as DEAH-box proteins (35), because most of them hold the Asp–Glu–Ala–His motif within their Walker-B site (Figure 2A). However, both RecQ-like and Rad3-like proteins are structurally and functionally different from DEAH-box RNA helicases (Figure 1). Thus, a given SF2 helicase cannot be

classified exclusively by its sequence of motif II, but by the sequence identity of all characteristic helicases motifs. The terms 'DEAD-box', 'DEAH-box' and 'DExH-box' are somehow misnomers and should clearly be defined once and for all to avoid ambiguities. However, the terms 'DEAD-box' and 'DEAH-box' will likely be kept for historical reasons, although not all the members of these families of RNA helicases harbour the expected Asp–Glu–Ala–Asp or Asp–Glu–Ala–His signatures within their Walker B site.

The term 'DEAH-box' hereafter will be employed exclusively to refer to RNA helicases of the 'DEAH-box' group/family as depicted in Figures 1 and 3. As for 'DExD/H-box', it will be employed as a synonym of SF2 helicases.

The DEAH-box protein family of RNA helicases

The DEAH-box protein family of RNA helicases was first recognised in 1991 following the discovery of three novel putative SF2 helicases in *S. cerevisiae* (36). The examination of the conserved helicase motifs of the yeast Prp2, Pr16 and Prp22 proteins revealed that they could neither be classified as members of the DEAD-box family nor as viral RNA helicase-like proteins. Thus, it was proposed that these proteins constituted a novel family of RNA-helicase-like proteins, which was termed 'DEAH-box' on the basis of the sequence (Asp–Glu–Ala–His) of their conserved motif II. Following this finding, four additional DEAH-box proteins have been identified in *S. cerevisiae* and many others in various organisms. In fact, DEAH-box proteins constitute a widely spread family of RNA helicases that have been identified in almost all species, from bacteria to mammals.

Biological functions of DEAH-box proteins

DEAH-box proteins play an essential role in RNA biology. Indeed, they harbour a wide range of biological functions and have been shown to take part in nearly all aspects of the RNA metabolism, ranging from transcriptions to RNA decay (Table I and references therein). Although RNA helicase activity has been confirmed for a few of them, the precise function of DEAH-box proteins in these contexts remains mostly elusive. As for DEAD-box proteins, they are thought to function in remodelling the structure and/or composition of ribonucleoprotein (RNP) complexes by locally melting RNA duplexes or displacing proteins from RNA molecules. It should also be mentioned that while DEAH-box proteins in yeast have been extensively investigated on functional and biochemical aspects, the biological significance of a majority of their counterparts in higher eukaryotes is largely unknown (Figure 3).

Genetic studies in yeast have demonstrated that DEAH-box proteins achieve highly specific tasks. In most cases, they are required at a specific stage of RNA metabolism and a majority of them are highly specific for their substrates. As revealed by the lethality of null mutants in yeast, most DEAH-box proteins are essential, suggesting tight target specificity for each protein (26,37). Nevertheless, in rare occasions, some RNA helicases have emerged to harbour multiple facets acting at different steps of the RNA metabolism or even showing activity on DNA (33,38,39). For instance, the yeast DEAH-box protein Prp43 was initially identified as a pre-mRNA splicing factor acting in the release of the intron lariat from the spliceosome (40-42). However, it was subsequently shown to function also in ribosome biogenesis (43-45). Otherwise, as shown in Table I, RNA helicase A is a multifunctional mammalian helicase that was shown to function as a transcriptional activator and to take part as well in the translation

Table I | **Biological functions of DEAH-box RNA helicases** from *E. coli* (Ec), *S. cerevisiae* (Sc), *C. elegans* (Ce), *D. melanogaster* (Dm) and *H. sapiens* (Hs).

Symbol	Org.	Protein	Function	Remarks
hrpA	Ec	hrpA	Takes part in the ribosome-mediated cleavage of the *daa* mRNA (46).	
DHX8	Sc	Prp22	Takes part in pre-mRNA splicing. Mediates the release of the spliced mRNA from spliceosome (36,47-50). Was also shown to repress the splicing of aberrant splicing intermediates (51).	Essential
	Ce	mog-5	Takes part in the post-transcriptional control of the switch from spermatogenesis to oogenesis (52,53).	Embryonic lethal
	Hs	DHX8	Takes part in pre-mRNA splicing. Mediates the release of the spliced mRNA from spliceosome (109,110).	
DHX9	Ce	RHA	Required for germ cell proliferation, normal germ cell nuclear morphology, RNA-mediated interference of germline-expressed genes, and silencing of germline-expressed transgenes (54,55).	Not essential
	Dm	Mle	Is an essential component of the dosage compensation machinery required for increased transcription of X-chromosome linked genes in males (56-60). Is also involved in RNA editing process (61).	Essential for males
	Hs	RHA / NDH-II	Multifunctional helicase (62). Functions as a transcriptional activator (63-67). Is involved in the expression and nuclear export of retroviral RNAs (68-71). Takes also part in translation of selected mRNAs through interaction with their 5'-untranslated region (72). Was identified as a RISC component and shown to functions in RISC as an siRNA-loading factor (73). Was shown to be associated with SMN protein and possible involvement for RHA in pre-mRNA processing (74).	Early embryonic (E7.0) lethality for *Rha*$^{-/-}$ knock-out mice (237)
DHX15	Sc	Prp43	Takes part in pre-mRNA splicing. Mediates the disassembly of spliceosome after the release of mature mRNA (40,41). Takes also part in the processing of 35S rRNA precursor (43-45,75).	Essential
	Hs	DHX15	Pre-mRNA processing factor involved in disassembly of spliceosomes after the release of mature mRNA (76-78).	
DHX16	Sc	Prp2	Takes part in pre-mRNA splicing. Mediate the activation of the spliceosome before the first transesterification step (79-81).	Essential
	Ce	mog-4	Take part in the post-transcriptional control of the switch from spermatogenesis to oogenesis (52,53).	Embryonic lethal
	Hs	DHX16	Likely involved in pre-mRNA splicing since expression of *DHX16* gene in *S. pombe* partially rescued the temperature-sensitive phenotype of *dhx16* null mutant cells (82).	
DHX29	Hs	DHX29	Takes part in translation initiation. Required for efficient initiation on mammalian mRNAs with structured 5'-UTRs by promoting efficient NTPase-dependent 48S complex formation (83,84).	
DHX30	Hs	DHX30	Required for optimal function of the zinc-finger antiviral protein (ZAP, ref. 85).	
DHX32	Hs	DHX32	Might be involved in regulating T-cell response to certain apoptotic stimuli (86).	Not essential in mouse*
DHX34	Ce	SMGL-2	Involved in non-sense-mediated mRNA decay process (87).	Essential (87)
	Hs	DHX34	Involved in non-sense-mediated mRNA decay process (87).	
DHX36	Dm	CG9323	Possesses *in vitro* G4-RNA-resolvase activity (88).	
	Hs	RHAU	Involved in ARE-mediated decay of uPA mRNA (32). Possesses *in vitro* G4 re-solvase activity and was identified as the major source of G4 resolving activity in HeLa cell lysates (33,89).	Early embryonic (E7.5) lethality for *Rhau*$^{-/-}$ knockout mice (90)
DHX37	Sc	Dhr1	Takes part in the processing of 18S rRNA (91-93).	Essential
DHX38	Sc	Prp16	Takes part in pre-mRNA splicing. Required during the second catalytic step and promotes 3' splice site cleavage, exon ligation as well as conformational change in the spliceosome (94-97).	Essential
	Ce	mog-1	Takes part in the post-transcriptional control of the switch from spermatogenesis to oogenesis (98).	Embryonic lethal
	Hs	DHX38	Takes part in pre-mRNA splicing. Required during the second catalytic step (99).	
TDRD9	Dm	Spindle-E	Component of the piRNA pathway. Plays a central role during meiosis by forming complexes composed of piRNAs and Piwi and govern the methylation and subse-quent repression of transposable elements (100-103). Takes also part in the control of telomere maintenance in the germline (104).	Essential
	Hs	TDRD9	Takes part to the repression of transposable elements during spermatogenesis. Acts via the piRNA metabolic process (105).	

*Abdelhaleem, M. unpublished data.

of specific mRNAs. Therefore, RNA helicases may also achieve more than one task acting on different RNAs depending on interactions with protein cofactors.

Evolutionary aspects of the DEAH-box family of RNA helicases

Comparative genomic analyses have revealed that SF2 RNA helicases are ubiquitously distributed over a wide range of organisms, viruses included (9). While DEAD-box, RIG-I and Ski2-relatated RNA helicases are widespread in all three domains of life (archaea, bacteria and eukarya), DEAH-box proteins are in contrast only found in bacteria and in eukaryotes. The *hrpA* gene product is the only bacterial representative of the DEAH-box protein family (Figure 3). True orthologues of this gene are also found in proteobacteria and spirochaetes phyla as well as in *Deinococcus* genus (deinococcus-thermus phylum). This suggests dissemination via horizontal gene transfer among bacteria, although the initial direction of horizontal transfer responsible for the bacterio-eukaryotic distribution remains obscure. HrpA is implicated in mRNA processing (46), and like all DEAD-box proteins in *Escherichia coli*, its function is not essential for cell viability under standard culture conditions (46,106). A paralogue of *hrpA* gene, named *hrpB*, is also found in *E. coli*. It may have appeared following gene duplication. However, as shown in Figure 3, the *hrpB* gene product is distantly related to *hrpA* and to other DEAH-box proteins found in eukaryotes.

DExD/H-box RNA helicases are more widespread in eukaryotes than in prokaryotes and most of them are essential (26). The yeast *S. cerevisiae* genome encodes seven DEAH-box proteins out of which six are required for cell viability. Four of these helicase (Prp2, Prp16, Prp22 and Prp43) have been shown to take part in pre-mRNA splicing (36,40,79,94,107) and Prp43 was later demonstrated to be as well required during the processing of the 35S rRNA precursor (43-45). Two extra DEAH-box RNA helicases, Dhr1 and Dhr2, have been shown to take part in the processing of the 18S rRNA (91). Finally, the open reading frame Ylr419w is the only known DEAH-box helicase in yeast that was found to be dispensable for cell viability and whose function is not yet determined (91).

As shown in Figure 3, five yeast DEAH-box proteins followed a conservative pattern of evolution, insofar as true orthologues of Prp2, Prp16, Prp22, Prp43 and Dhr1 proteins are found in metazoan species. In several cases, direct data support similar roles of these proteins in higher eukaryotes (76-78,82,99,108-111). However, one apparent exception is mog-1, the *C. elegans* orthologue of Prp16, which is not explicitly required for pre-mRNA splicing (112). *Mog* genes in nematodes take part in the sex determination process of the germline. Two other DEAH-box proteins, mog-4 and mog-5, corresponding to Prp2 and Prp22, respectively, are also required for the posttranscriptional regulation of the switch from spermatogenesis to oogenesis (53). Although it is still possible that *mog-1*, *mog-4* and *mog-5* have switched to completely new functions, one cannot exclude the possibility that these three DEAH-box proteins are still involved in splicing and that the Mog (masculinisation of germline) phenotype results indirectly from abnormal splicing of factors implicated in the sex determination process (112).

Finally, although the yeast spliceosomal DEAH-box proteins manifest a conservative evolutionary pattern, the genes related to *ylr419w* have largely diversified in metazoans. *Ylr419w* has no clear orthologue in higher eukaryotes, which suggest that the paralogue forms of this gene found in metazoans may have arisen from several gene duplication events. The large number of biological functions assigned to this group of proteins (sometimes referred to as the 'RHA protein group'; ref. 14,30) indicates that DEAH-box RNA helicases have gained and diversified their functions to fit with some aspects of RNA metabolism in metazoans with respect to yeasts.

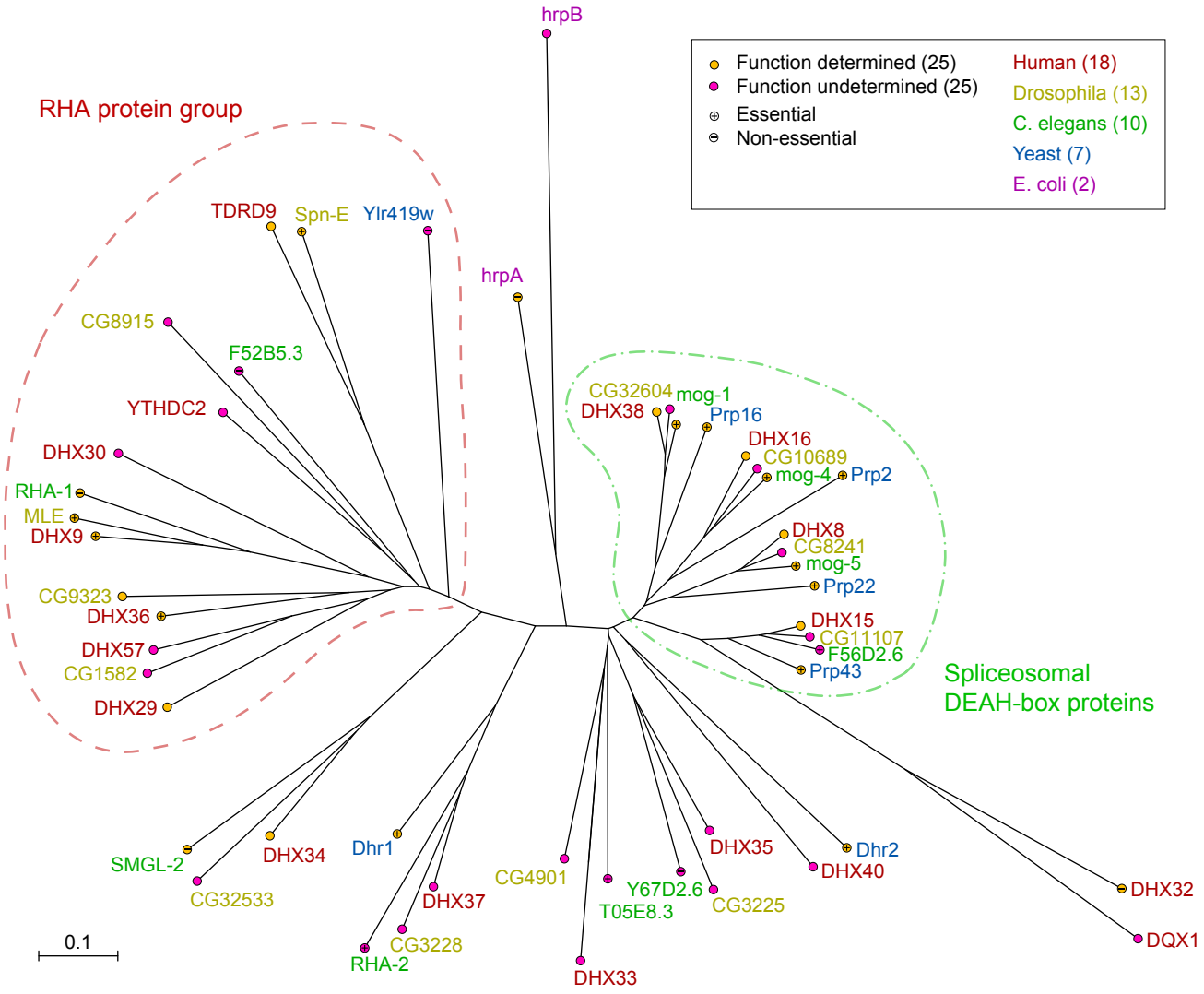


Figure 3 | **Phylogeny of DEAH-box proteins** from *E. coli*, *S. cerevisiae*, *C. elegans*, *D. melanogaster* and *H. sapiens*. The unrooted phylogenetic tree was computed by MAFFT from the multiple sequence alignment shown in Supplementary Figure 1 using the neighbor-joining method on 241 conserved sites (substitution model = JTT; heterogeneity among sites (α) = ∞; bootstrap resampling = 100). The coloured dots at the branch extremities indicate whether the protein is required for the organism viability and its current investigtion status.

Modular architecture of DEAH-box proteins

Structurally, DExD/H-box proteins contain of a highly conserved catalytic core composed of two RecA-like domains that couples NTP hydrolysis with the helicase activity. The helicase core domain is often flanked by N- and C-terminal regions of variable length and sequence (Figure 4). While the core domain of RNA helicases has been extensively investigated, much less is known about the biological role of these N- and C-terminal regions. The helicase core of DExD/H-box proteins is assumed not to contribute directly to the substrate specificity of the enzyme because in all crystal structures, the highly conserved helicase core region interacts only with the phosphoribose backbone of the bound single-stranded nucleic acid and not with the nucleobases. In contrast to the helicase core, the N- and C-terminal flanking regions are usually unique, with the exception of certain identifiable sequence features. On several occasions, these regions have been shown to provide substantial substrate specificity through their interaction with RNAs or with protein partners that modulate the activity and/or the specificity of the helicase (for reviews, see ref. 6,37). Thus, the helicase flanking regions can be regarded as ancillary domains that endow the

enzyme with specificity, thereby positioning the helicase core in close proximity to its substrate.

DEAH-box RNA helicases differ from their DEAD-box counterparts, insofar as their C-terminal region is conserved over 300 amino acids beyond the helicase core region. First evidence of this distinctive feature emerged from an early characterisation of the yeast spliceosomal DEAH-box proteins (36). Since deletions of the C-terminal region resulted in the loss of interaction with the spliceosome, it was initially assumed that this conserved region could mediate the interaction of the helicase with the spliceosome. However, such a scenario now seems less plausible, because all DEAH-box proteins, whether they are implicated in splicing or not, contain a conserved C-terminal region of 300 amino acid length [hereafter referred to as the 'helicase associated' (HA) region; Figure 4]. Compared to the helicase core, the HA region has so far received relatively little attention. Thus, its function is less well understood than that of the helicase core. However, several experimental observations strongly suggest its requirement for the proper function of DEAH-box helicase *in vivo* and *in vitro* (41,113-118). Notably, deletion experiments carried out on the prototypical Prp2, Prp16, Prp22 and Prp43 RNA helicases revealed that the most extreme and non-conserved part of the C-terminal region was not essential (Figure 5).

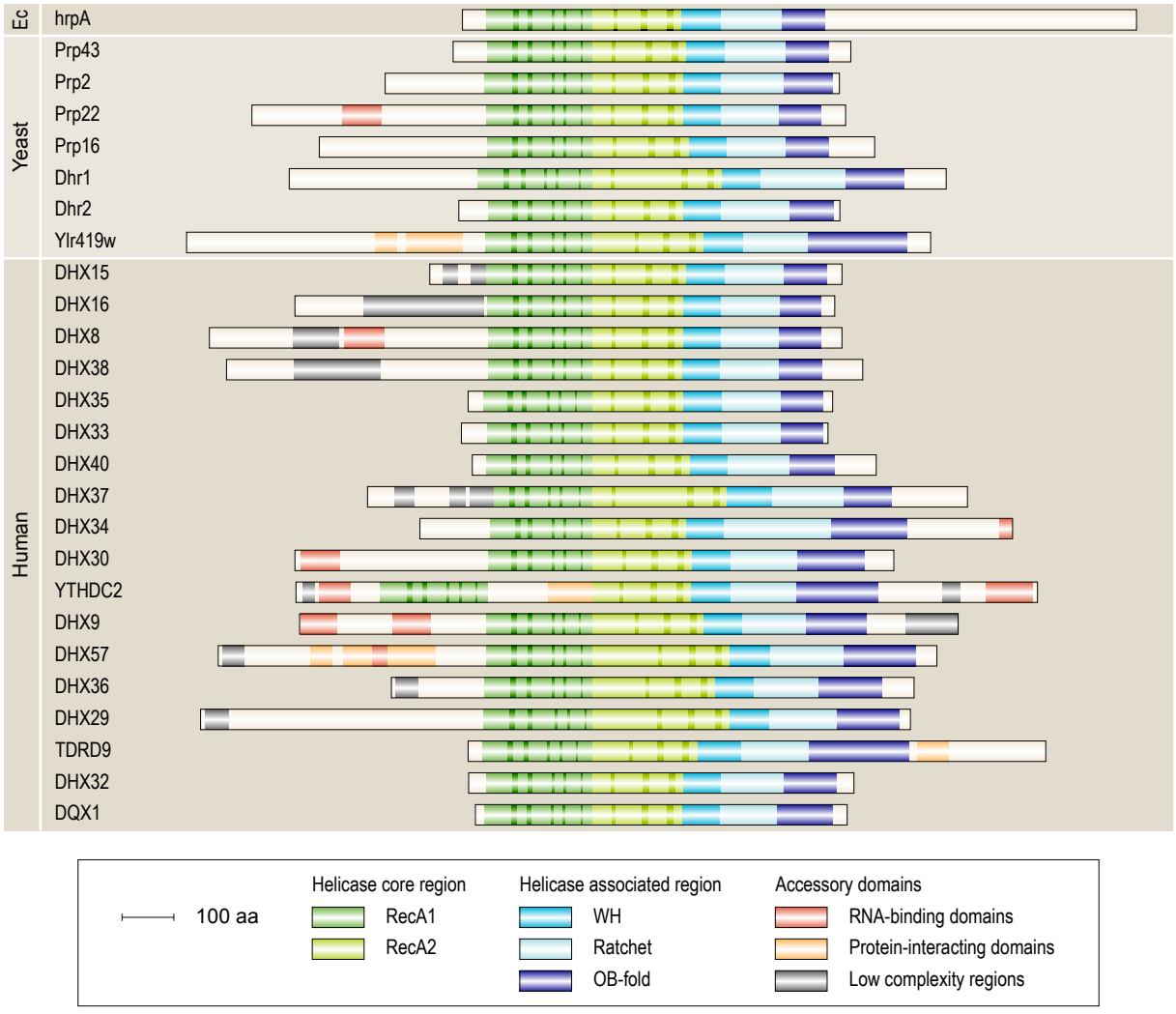


Figure 4 | **Domain architecture of DEAH-box RNA helicases from bacteria, budding yeast and human.** Proteins are aligned by the helicase core region, which is shown as two adjacent green boxes. The conserved ATPase/helicase motifs I–VI are indicated within the helicase core region by darker vertical bars. The conserved C-terminal helicase associated (HA) region is depicted as three adjacent blue boxes. All proteins and domains are shown to scale. Ec, *Escherichia coli*; WH, winged-helix domain; OB-fold, oligonucleotide/oligosaccharide-binding fold domain.

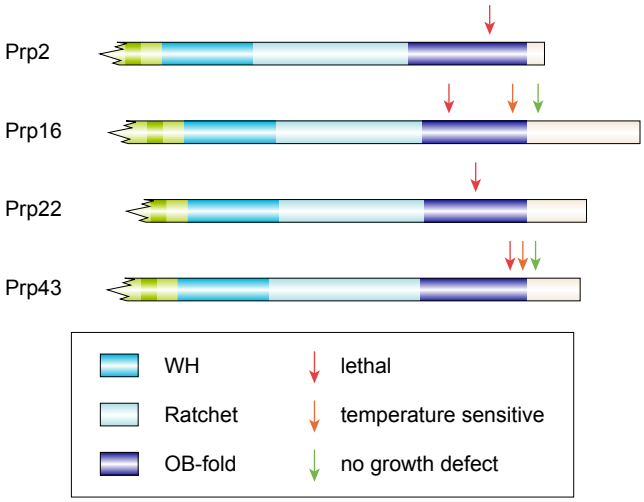


Figure 5 | **Requirement of the helicase associated (HA) region for DEAH-box protein function**. Phenotypic growth effects of C-terminal deletions on the *S. cerevisiae* spliceosomal DEAH-box proteins. The position and the severity of the deletions are indicated above the protein by coloured arrows. All proteins and domains are shown to scale.

Conversely, severe growth defect phenotypes were observed for mutants harbouring deletions in the HA region. The prevalence of the HA region among DEAH-box proteins and its apparent functional significance strongly support the idea that it constitutes an extension of the helicase core. Pattern and profile searches of the HA region in other proteins indeed have revealed that this domain is unique among DEAH-box proteins and is always associated with the helicase core region. Further evidence for the importance of this conserved sequence element came from the examination of the shortest DEAH-box proteins. Interestingly, the human DHX33 and DHX35 consist exclusively of the helicase core and the HA region and therefore represent minimal DEAH-box proteins (Figure 4). Taken together, these observations suggest that within the DEAH-box family of helicases, the two RecA domains together with the HA region define a minimal helicase enzymatic core region. Likewise, family-specific C-terminal conserved flanking regions of the viral NS3, the RecQ- and the Ski2-related helicases, being associated with the helicase core region, have already been shown to contribute to the unwinding process (29,119,120).

Structural aspects of DEAH-box proteins

Until recently, a molecular understanding of DEAH-box helicase function was significantly limited due to the unavailability of structural models. However, two independent X-ray structures of the yeast spliceosomal Prp43 were released in 2010 (113,121). In both models (RCSB PDB ID: 3KX2, 2XAU), full length Prp43 was cocrystallised in the presence of ADP. The two studies have proposed essentially the same structure, with a root mean square deviation (RMSD) of 0.27 Å for all corresponding C_α atoms.

Prp43 provides a good model for DEAH-box protein. Its amino acid sequence displays explicit similarity with other members of the DEAH-box family over 86 % of its length (Figure 6A). Only the first 70 amino acids and the last 40 amino acids of its N- and C-terminal regions, respectively, are unique to Prp43. Thus, as all DEAH-box helicases share the same domain organisation (Figure 4), structural observations made on Prp43 helicase core and HA regions can be readily extrapolated to the other members of the family.

The crystal structure of Prp43 reveals that the helicase core and the HA region consist of five structural domains (Figure 6B and C). The helicase core region consists of two abutting RecA-fold domains connected by a short linker. The RecA domains 1 and 2 contain the composite helicase motifs involved in ATP binding, hydrolysis and nucleic acid binding. The overall structure of Prp43 helicase core region more resembles the core region of the viral helicases than that of DEAD-box proteins (Figure 7A and Supplementary Figure 2). Like the

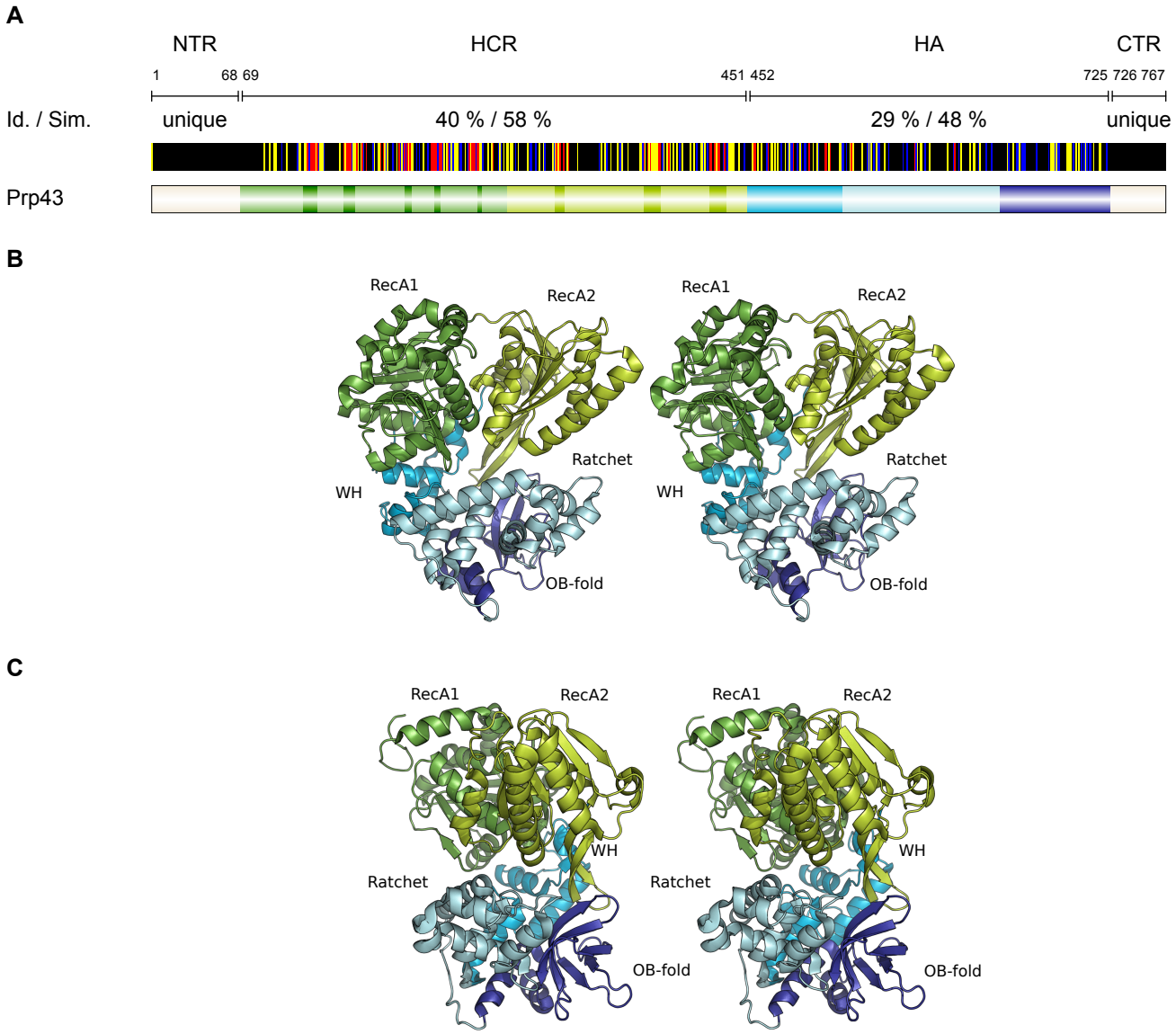


Figure 6 | **Structure of the yeast DEAH-box protein Prp43.** (**A**) Schematic representation of the domain organisation and amino acid conservation of *S. cerevisiae* Prp43 protein. The conserved ATPase/helicase motifs I–VI of the DEAH-box family are indicated within the helicase core region (HCR) by vertical bars. The HCR is flanked by the N-terminal (NTR) and C-terminal (HA–CTR) regions of 68 aa. and 316 aa., respectively. Each residue of Prp43 sequence is represented with a colour code that denotes its degree of conservation amongst various paralogous DEAH-box protein sequences (Supplementary Figure 1). Similarity is shown in red for 100 %, yellow for 99–80 % and blue for 79–60 %. Average values of identity (Id.) and similarity (Sim.) for NTR, HCR, HA and CTR regions are indicated. (**B**) Front and (**C**) side view of Prp43 structure (aa. 69–725). The unique NTR (aa. 1–68) and CTR (aa. 726–767) are omitted for clarity. The five structural domains are coloured as in panel (A).

flavivirin NS3 related helicases [hepatitis C virus (HCV, ref. 122), Murray Valley encephalitis virus (123) or yellow fever virus (124)], the RecA2 domain of Prp43 possesses a long and twisted antiparallel β-hairpin positioned between motifs V and VI that inserts into a cleft of the HA region (Figures 6B and 7A). In NS3 helicases, the corresponding β-hairpin is assumed to act as a pivot point that allows to lock the RecA2 domain orientation in the ADP state relative to the rest of the molecule (125). Both the β-hairpin and the residues it contacts are in overall highly conserved among various strains of virus.

The HA region of Prp43 consists of three distinct domains; two of which displays an unanticipated structural similarity with the C-terminal region of the HELQ-related archaeal Hel308 and Hjm DNA helicases (Ski2 family of helicases, ref. 29,126; Figure 7B). This was an unexpected finding since Prp43 displays very little sequence similarity (<10 % identity) with Ski2-related helicases (Supplementary Figure 3). The two first domains of the HA region fold respectively into of a

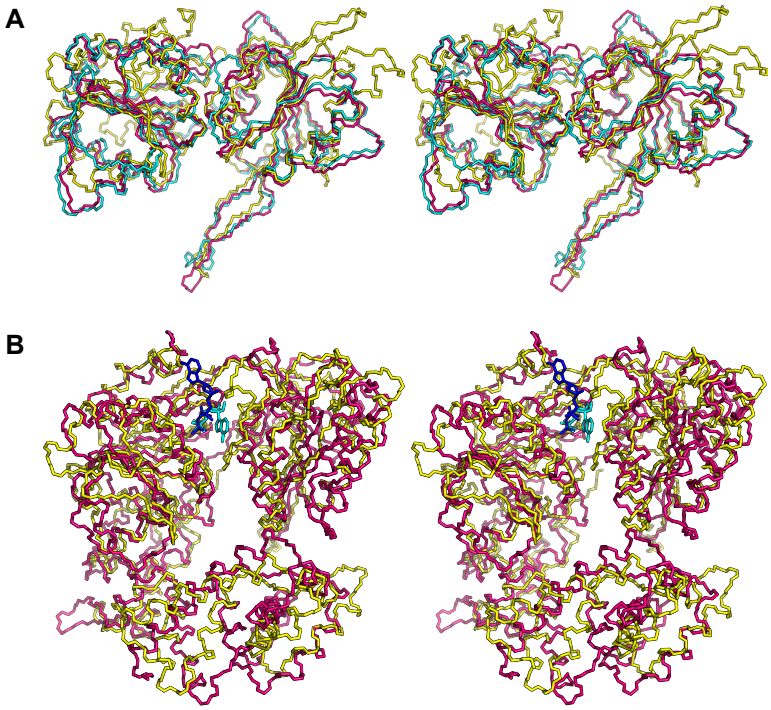


Figure 7 | **Structural similarities between Prp43 and the NS3 and Ski2-like helicases**. (**A**) Three-dimensional protein structure superposition of Prp43 helicase core (yellow, RCSB PDB ID: 2XAU, aa. 105–451) with two flavivirin NS3 related helicases [pink, 2WV9 (Murray Valley encephalitis virus); cyan, 1YKS (yellow fever virus)]. (**B**) Three-dimensional protein structure superposition of Prp43 helicase core and HA regions (yellow, 2XAU, aa. 90–634) with the Ski2-related Hjm helicase (pink, 2ZJA). The adenosine cofactor is shown in cyan for Prp43 and in blue for Hjm.

winged-helix (WH) motif and a seven-helix bundle. The WH motif is tightly packed against the RecA1 domain, while the seven-helix bundle, referred as the 'Ratchet' domain in Hel308 owing to its putative function as a ratchet for nucleic acid, binds across the two RecA-like domains. The third and last domain of the HA region consists of an oligonucleotide/oligosaccharide-binding (OB) motif arranged in a five-stranded β-barrel. The OB-fold domain is linked to the Ratchet domain by an α-helix, which extensively packs with the Ratchet domain.

Mechanism of rNTP binding and hydrolysis by DEAH-box proteins

As for the other SF2 proteins, the catalytic site for rNTP hydrolysis in Prp43 is located at the interface cleft of the two RecA-like domains. However, the orientation of the nucleoside moiety differs appreciably from that observed in other human SF2 helicases and more resembles that of the viral helicases. In Prp43, the pyrophosphate moiety of the nucleotide is likewise orientated as in DEAD-box proteins relative to motifs I and II. However, the adenine ring is rotated by about 140° around the triphosphate axis compared to that in the DEAD-box protein and points inward (Figure 8A). Unlike other families of SF2 helicases, DEAH-box proteins lack the Q-motif, which hydrogen-bonds to the N6 and N7 positions of the adenine base, thereby providing ATP-binding specificity (127,128). Instead, in Prp43, the adenine ring is stacked between the side chains of an Arg residue from the RecA1 domain and the phenyl ring of a Phe from RecA2 domain. The corresponding Arg and Phe residues are conserved in almost all DEAH-box proteins, suggesting the existence of a similar nucleotide binding site among members of the family (Supplementary Figure 1). By analogy to the Q-motif, these two residues and the few conserved surrounding amino acids will be referred to as the R- and the F-motifs, respectively (Figure 9). The R-motif is located on an α-helix positioned between domains Ia and Ib, while the F-motif is situated on another α-helix between domain IV and V. According to this binding mechanism, the base moiety of the NTP is exclusively maintained by hydrophobic effect. Hence the base is not specifically recognised by the helicase, indicating why DEAH-box proteins are so promiscuous in term of their rNTP/dNTP specificity (Table III and references therein).

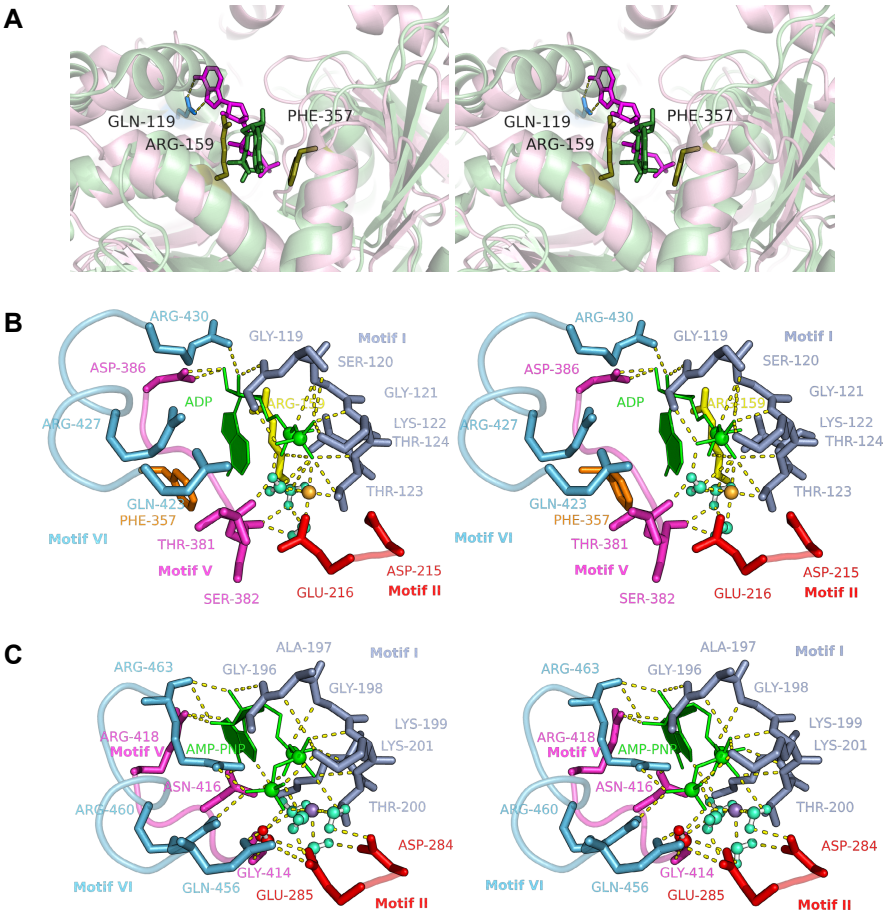


Figure 8 | **NTP binding by DEAH-box RNA helicases**. (**A**) ATP binding by DEAH-box proteins versus DEAD-box proteins. Three-dimensional protein structure superposition of the helicase core of Prp43 (dark green, RCSB PDB ID: 2XAU, aa. 105–451) with that of the human DDX19 (violet, 3G0H, 60–466). The adenosine cofactor is shown in green for Prp43 and in violet for DDX19. The side chains of Gln-119 (Q-motif, DDX19), Arg-159 (R-motif, Prp43) and Phe-357 (F-motif, Prp43) are depicted as coloured sticks. (**B**) Close-up view of the NTP binding and hydrolysis site of Prp43 in its open conformation state. (**C**) Close-up view of the NTP binding and hydrolysis site of the dengue virus DENV NS3 helicase (2JLV) in its closed conformation state. The catalytic water is shown in red.

Apart from the newly identified R- and F-motifs, the nucleotide binding pocket of DEAH-box proteins is formed by the helicases motifs I, II, V and VI (Figure 8B). Motif I contains the characteristic P-loop and makes several contacts with the triphosphate part of the nucleotide directly and through Mg^{2+} ion and water. The carboxyl groups of Asp-II* and Glu-II coordinate the Mg^{2+} ion of Mg·ADP/ATP through outer-sphere electrostatic interactions. The Glu-II residue is also thought to act as a catalytic base during ATP hydrolysis. In Prp43, several residues from motifs V and VI in the RecA2 domain also contact the nucleotide. However, in both available Prp43 structures, the helicase was crystallised with ADP in the open conformation state, which does not reflect the correct positioning of the residues necessary for NTP hydrolysis. Closure of the inter-domain cleft between the two RecA domains was shown to occur upon cooperative binding of NTP and RNA by the helicase and provides the cadre of the ATP binding and hydrolysis site (129,130). Structural basis of NTP hydrolysis by Prp43 can be inferred from the analysis of the structurally related dengue virus (DENV) NS3 helicase which was crystallised at several stages along its catalytic cycle (129). Unlike the open conformation of Prp43, the closed conformation state of the DENV NS3 helicase reveals that the RecA2 domain interacts with the NTP through several amino acids from motif V and the evolutionary conserved Gln-VI, Arg_2-VI and Arg_3-VI (Figure 8C). In both Prp43 and DENV NS3 structures, the 2'-hydroxyl group of the ribose is contacted by residues from motif V. However, this interaction does not appear to play a decisive role in discriminating rNTPs from dNTPs, since Prp43 along with

* Hereafter, the conserved amino acids of the helicase motifs will be designated as follow: the amino acid in question is referred using the three letter code; the subscript index (when necessary) corresponds to the rank of the amino acid when presents in multiple instances in a given motif; the roman numeral refers to the helicase motif. E.g. 'Arg_3-VI' denotes the third conserved arginine in helicase motif VI (QRxGRAG<u>R</u>).

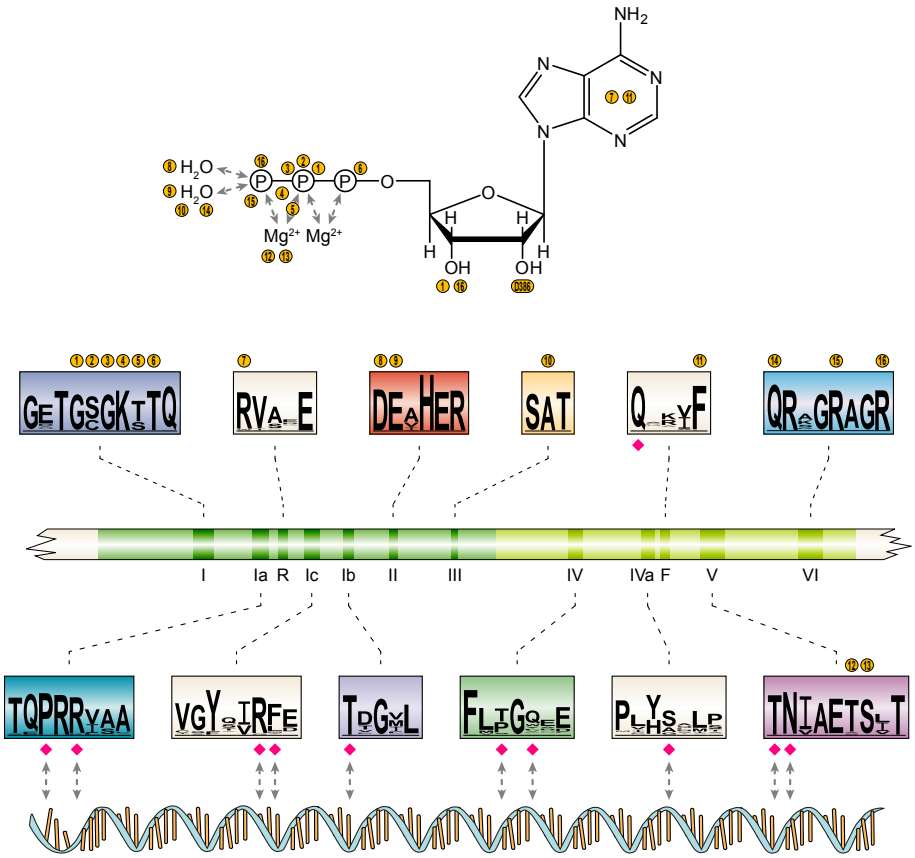


Figure 9 | **NTP and nucleic acid binding by conserved ATPase/helicase motifs in DEAH-box proteins**. In Prp43, the NTP cofactor is contacted directly or undirectly by several residues from motifs: I, R, II, III, F, V and VI. The numbered yellow dots above the sequence logos refer to the part of the NTP moiety they bind. Motifs Ia, Ic, Ib, IV, IVa and V are implicated in single-stranded nucleic acid binding. The pink diamonds beneath sequence logos denotes amino acids that may contact the translocating nucleic acid. These residues have been mapped based on the structural similarity between Prp43 and NS3 or Hel308 helicases which have been crystallised in complex with nucleic acids (Supplementary Figures 2 and 3). All the sequence logos were constructed from the multiple sequence aligment shown in Supplementary Figure 1.

many other DEAH-box proteins can irrespectively hydrolyse dNTPs besides rNTPs (96,107,118,131,132). In the closed conformation state, the Lys-I bridges the β- and γ-phosphates and is thought to coordinate the γ-phosphate during the hydrolysis step (133,134). On the RecA2 domain, Arg_2-VI was proposed to function like an 'arginine finger' (135). It contacts the γ-phosphate and might stabilise the transition state in the course of ATP hydrolysis (27,134,136). In addition, the conserved Glu-II and Gln-VI residues are in close proximity. Their side chain hydrogen-bond a water molecule which is positioned at right distance for an in-line nucleophilic attack of the γ-phosphate. Activation of the catalytic water was proposed to occur through proton transfer to the Glu-II and/or polarisation by Gln-VI (129). Following hydrolysis of the β–γ phosphoanhydride bond of the bound rNTP, both phosphate and rNDP moieties are released from the inter-domain cleft.

In addition to these structural data, several mutagenesis studies carried out essentially on the prototypical DEAH-box proteins Prp2, Prp16, Prp22 and Prp43 have confirmed the significance of the conserved amino acids in motifs I, II and VI for the ATPase activity of DEAH-box proteins (Table II and references therein). In particular, amino acid substitution of the conserved Lys-I, Asp-II, Glu-II and Gln-VI residues have been shown to reduce the ATPase activity of wild-type proteins by more than 90 %. Finally, the allosteric conformational changes of the catalytic site for ATP hydrolysis in response to RNA binding clarify the causes of the cooperativity between rNTPs and nucleic acid binding (137) and explains the substantial stimulation of the basal ATPase activity of DEAH-box proteins in the presence ssRNAs (32,41,48,83,94,107).

Mechanisms of ssRNA binding by DEAH-box proteins

Although Prp43 was not crystallised together with a bound oligoribonucleotide, insight into the bases of RNA recognition by DEAH-box proteins can be inferred from the study of the archaeal Hel308 (RCSB PDB

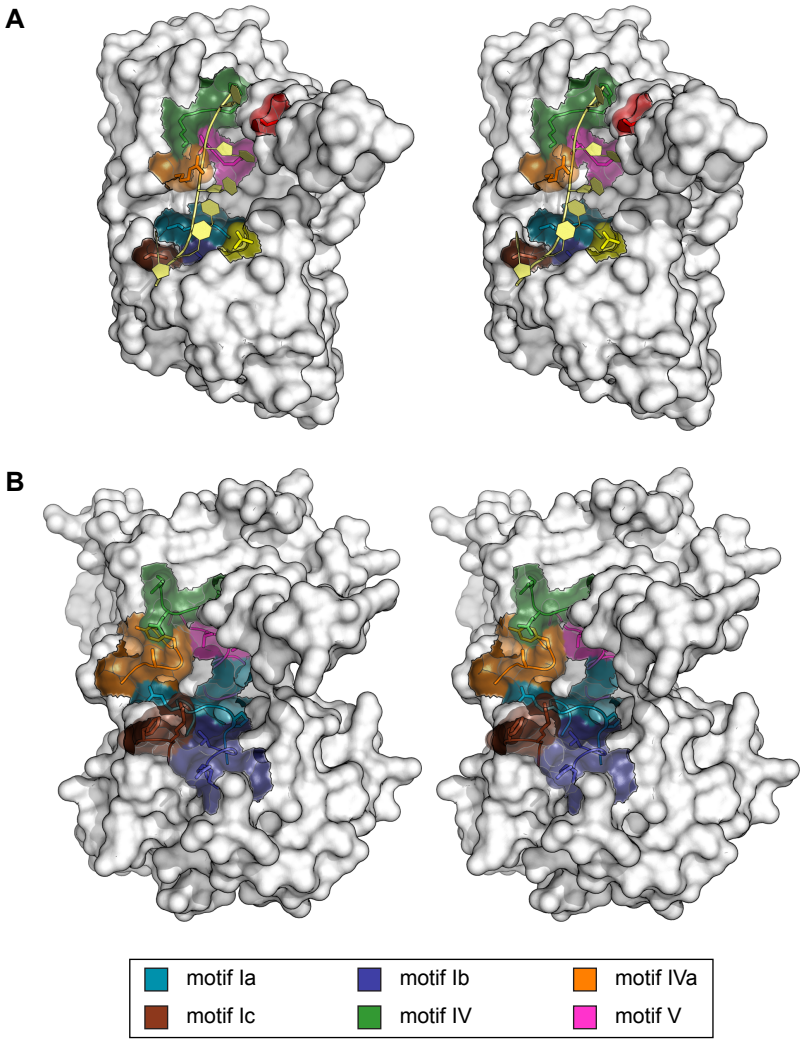


Figure 10 | **Nucleic acid binding motifs in helicase core region of HCV NS3 and Prp43 DEAH-box helicases**. (**A**) Surface representation of HCV NS3 helicase core region (RCSB PDB ID: 3KQL, aa. 189–481) in complex with ssDNA (dA_6, yellow). The conserved helicase motifs (Ia, Ic, Ib, IV, IVa and V) implicated in binding of single-stranded nucleic acid are depicted in colour with transparency surface. Amino acids that make direct contacts with the oligodoxynucleotide are shown as sticks. (**B**) Surface representation of Prp43 helicase core (2XAU, aa. 69–451) in the same orientation as NS3 in panel (A). Amino acids that may form direct contacts with the translocating nucleic acid are depicted as coloured sticks.

ID: 2P6R) and the viral NS3 (RCSB PDB ID: 2JLV, 3KQL) helicases which have been crystallised in complex with nucleic acids. In both Hel308 and NS3 structures, the single-stranded nucleic acid is located in the groove that separates the helicase core from the C-terminal region (29,122,129). The nucleic acid is extensively contacted by residues localised in the loops between secondary structure elements from the two RecA-like domains. Both the location and polarity of the bound nucleic acid are similar to that previously observed in DEAD-box protein–RNA complexes (27,133,138,139). The RecA1 domain binds the 3'-end of the nucleic acid substrate through motifs Ia, Ib and Ic while motifs IV, IVa and V from the RecA2 domain binds the 5'-extremity. Motif Ic is located between motifs Ia and Ib and is equivalent to the GG-doublet found in many DEAD-box helicases (31). As for motif IVa (140), which is situated between motifs IV and V, it corresponds to the QxxR motif of DEAD-box proteins. Nucleic acid binding by Hel308 and NS3 proteins is similar to that found in DEAD-box proteins, insofar as the helicase motifs Ia, Ib, Ic, IV, IVa and V interact essentially in a sequence-independent manner with the phosphoribose backbone of the single-stranded nucleic acid. No interaction occurs between the helicase core and the bases of the nucleic acid, explaining the lack of sequence specificity. However, direct contacts to the 2'-hydroxyl groups of the ribose moieties leads to discrimination against DNA as a substrate. Finally, several contacts to the nucleic acid involve only the peptide backbone, explaining the occasionally low sequence conservation of the nucleic acid binding motifs (Figure 2A).

Superposition of Prp43 with either Hel308 or NS3 revealed that Prp43 contains a similar single-stranded nucleic acid-binding cavity, formed by the two

RecA-like domains and the three domains of the HA region. As for Hel308 and NS3, the conserved motifs Ia, Ib, Ic, IV, IVa and V of Prp43 are positioned at the interface of the cavity and superpose remarkably well on those of Hel308 and NS3 (Figure 10 and Supplementary Figures 2 and 3). Actually, little is know regarding the functional contribution of motifs Ia, Ib, Ic, IV, IVa and V in single-stranded nucleic acid binding by DEAH-box proteins. In Prp43, motif Ia may contact RNA through conserved Pro-Ia and Arg_2-Ia residues. The corresponding amino acids in DEAD-box, Hel308 and NS3 helicases make contact with nucleic acid, and mutation of Arg_2-Ia residue in Prp22 and Prp43 proteins causes a cold-sensitive phenotype (118,141). Motifs Ib, Ic and IVa have not been investigated by mutagenesis in DEAH-box proteins and putative nucleic-acid binding residues depicted in Figure 9 were proposed solely based on structural similarity between Prp43 and NS3 or Hel308 helicases. The conserved Thr_1-V residue in RecA2 domain likely contacts RNA in DEAH-box proteins. Mutation of the equivalent residues in Prp22 and Prp43 affects both helicase and RNA-dependent ATPase activities and prevents the mutated proteins to complement the corresponding Δ*prp22* and Δ*prp43* yeast strains (118,141). Substitution of Thr_1-V residue, in contrast, does not affect the RNA-independent ATPase activity, indicating that this mutation in motif V does not impair the ability to hydrolyze ATP but rather affect the responsiveness to RNA cofactors. In the crystal structures of DENV NS3 or Hel308, the Thr_1-V contacts a phosphate in the nucleic acid (29,129). Similarly to Prp22 and Prp43, replacing the corresponding threonine by alanine abolished NS3 helicase activity while mutated NS3 protein retained basal ATPase activity, which could not be stimulated by nucleic acid (142).

Besides the two RecA-like domains, RNA is also likely contacted by the HA region in DEAH-box proteins. As mentioned earlier, the two first domains of the Prp43 HA region are structurally homologous to the WH and Ratchet domains of HELQ-related DNA helicases. Both WH and Ratchet domains have been shown to contact nucleic acid in Hel308 structure (ref. 29; Supplementary Figure 3). Opposing interactions between residues of the Ratchet domain and the base moieties of the single-stranded nucleic acid were proposed to provide further stabilisation of the unwound nucleic acid and are probably important for translocation processivity. In particular, the last 25 amino acid-long α-helix of the Hel308 Ratchet domain constitutes an ideal hook for directional transport of ssDNA across the two RecA-like domains. The stacking of Arg-592 and Trp-599 of this helix with DNA base moieties is assumed to function as a ratchet on the nucleic acid strand, which would allow Hel308 to processively translocate along the single-strand nucleic acid to unwind the DNA duplex (29). Prp43 and other DEAH-box proteins contain a similar long α-helix at the end of their Ratchet domain (Figure 6B and Supplementary Figure 3). Although the Hel308 Arg-592 and Trp-599 residues are not conserved among DEAH-box proteins, others residues from the ratchet helix may form stacking interactions with the base moieties of the bound nucleic acid. Hence, the corresponding ratchet helix might also function similarly in DEAH-box proteins.

Finally, the last domain of the HA region of Prp43 and other DEAH-box proteins consists of an OB-fold domain which is also supposed to contribute to single-stranded nucleic acid binding (113). OB-folds have been shown to bind various ligands, such as single-stranded nucleic acids, oligosaccharides, proteins, metal ions and catalytic substrates (143). Interestingly, a DALI search for structural homologues (144) reveals that Prp43 OB-fold domain is most closely related to OB-folds found in cold shock proteins (Csps). Csps comprise a family of small proteins (7–8 kDa) that are structurally highly conserved and bind to single-stranded nucleic acids through their five-stranded β-barrel OB-fold domain (145,146). Csps function as RNA chaperones by preventing formation

of stable secondary structures of RNA which occurs under suboptimal temperatures. Structural alignment of Prp43 OB-fold domain with the bacterial CspA and CspB proteins reveals only moderate degree of sequence similarity (Supplementary Figure 4A and B). However, residues from the two RNP motifs by which Csps bind single-stranded nucleic acids are also conserved in Prp43. The nucleic acid-binding interface in OB-fold domains is generally defined by the groove formed by the β2 and β3 strands, and the loops L_{12}, $L_{3\eta}$ and L_{45} (147). In Prp43, the OB-fold domain is optimally orientated to interact with RNA at the entrance of the nucleic acid cavity and may provide extra stabilisation of the unwound RNA strand (Supplementary Figure 4C). The substantial contribution of the OB-fold domain in RNA-binding by Prp43 was confirmed *in vitro* by deletion or mutation experiments which resulted in defective Prp43 proteins

Table II | **Phenotypes resulting from mutations of helicase motifs in five prototypical DEAH-box RNA helicases**. A, ATPase activity; B, nucleic acid binding; F, biological function; G, cell growth; H, helicase activity; −, reduced activity; n, normal activity; +, enhanced activity.

Motif	Resi.	hrpA			Prp2			Prp16			Prp22			Prp43		
		Mut	Ph	Ref	Mut	Ph	Ref	Mut	Ph	Ref	Mut	Ph	Ref	Mut	Ph	Ref
I	G							G373A	G^n	(115)						
	E							E374A	G^n	(115)						
	T							T375A	G^n	(115)						
	G							G376A	G^n	(115)						
	S							S377A	G^n	(115)						
	G	G105A	F^-	(46)				G378A	G^-, A^-	(115,149)						
	K	K106A	F^-	(46)				K379A	G^-, A^-	(115,149)	K512A	G^-, A^-, H^-, F^-	(47,48,	K122A	G^-, A^-	(41)
	T	T107A	F^-	(46)				T380A	G^-	(115)			149,150)	T123A	G^-, A^-	(41)
Ia	T										T534A	G^n	(141)	T146A	G^n	(118)
	Q										Q535A	G^n	(141)	Q147A	G^n	(118)
	P										P536A	G^n	(141)	P148A	G^n	(118)
	R										R537A	G^n	(141)	R149A	G^n	(118)
	R										R538A	G^-	(141)	R150A	G^-	(118)
II	D	D197A	F^-	(46)				D473A	G^-, A^-	(115,149)	D603A	G^-, A^-, F^-	(149,150)	D215A	G^-, A^-	(41)
	E	E198A	F^-	(46)				E474A	G^-, A^-	(115,149)	E604A	G^-, A^-, F^-	(149,150)	E216A	G^-, A^-	(41)
	A															
	H	H200A	F^n	(46)				H476A	G^-, A^-, F^n	(115,149)	H606A	G^-, A^+, F^-	(150,151)	H218A	G^n, A^-	(41)
III	S				S378L	G^-, A^-, F^-	(152)	S505A	G^n, A^-, F^n	(115,149)	S635A	G^-, A^n, H^-, F^-	(149,150)	S247A	G^-, A^-	(41)
	A															
	T							T507A	G^n, A^-, F^n	(115,149)	T637A	G^-, A^+, H^-, F^-	(150,151)	T249A	G^n, A^n	(41)
IV	F										F697A	G^-, A^-, B^-, H^-, F^-	(141)	F309A	G^-	(118)
	L															
	T										T699A	G^n	(141)	T311A	G^n	(118)
	G										G700A	G^n	(141)	G312A	G^n	(118)
V	T										T757A	G^-, A^-, B^n, H^-, F^-	(141)	T376A	G^-, A^-, B^-, H^-, F^n	(118)
	N										N758A	G^-	(141)	N377A	G^n	(118)
	I															
	A															
	E										E761A	G^-	(141)	E380A	G^n	(118)
	T										T762A	G^n	(141)	T381A	G^n	(118)
	S										S763A	G^-	(141)	S382A	G^n	(118)
	L										I764A	G^-, A^+, B^n, H^-, F^-	(141)	L383A	G^-	(118)
	T										T765A	G^-, A^+, B^n, H^-, F^-	(141)	T384A	G^-, A^n, B^n, H^n, F^-	(118)
VI	Q	Q395A	F^-	(46)	Q548H	G^-	(153)	Q685A	G^-, A^-	(115,149)	Q804A	G^-, A^-, F^-	(149,150)	Q423A	G^-, A^-, F^-	(41,45)
	R	R396A	F^-	(46)				R686A	G^-	(115)	R805A	G^-	(150,151)	R424A	G^-	(41)
	a							S687A	G^n	(115)						
	G	G398A	F^-	(46)	G551N	G^-	(153)	G688A	G^-	(115)	G807A	G^-	(150)	G426A	G^-	(41)
	R	R399A	F^-	(46)				R689A	G^-, A^-	(115,149)	R808A	G^-, A^-, F^-	(149,150)	R427A	G^-	(41)
	A															
	G							G691A	G^n	(115)	G810A	G^-	(150)	G429A	G^n	(41)
	R	R402A	F^-	(46)				R692A	G^-, A^-	(115,149)	R811A	G^-	(150)	R430A	G^-	(41)

In Prp16, the S505L substitution is lethal.

with reduced RNA-dependent ATPase and RNA-binding activities (113). Furthermore, the importance of the OB-fold domain for the proper function of DEAH-box proteins *in vivo* is indicated by the results of deletion experiments with Prp2, Prp16, Prp22 and Prp43, which revealed that DEAH-box proteins with truncated OB-fold domain could not rescue the corresponding null mutant allele in yeast (ref. 41,114,115-117; Figure 5).

In sum, it emerges that DEAH-box proteins bind to ssRNA using residues from both the helicase core and the adjacent C-terminal HA region. Nevertheless, due to the lack of structural and functional studies, it is still difficult to provide a more detailed overview of how these helicases grasp RNA. Besides, it should be kept in mind that many DEAH-box helicases harbour as well additional classical RNA-binding motifs and/or low complexity sequences in their N- and C-terminal flanking regions that can also confer extra affinity or specificity for some nucleic acid structures (ref. 148; Figure 4).

Insight into RNA unwinding by DEAH-box RNA helicases

With the exception of Prp2 and DHX29, all DEAH-box proteins whose unwinding activity has been investigated *in vitro* demonstrate rNTP-dependent nucleic acid strand separation activity (Table III and references therein). Although DEAH-box proteins are not as processive as the hexameric DNA helicases which unwind thousands of base pairs, most of the DEAH-box RNA helicases display higher processivity than DEAD-box proteins and can readily dissociate a 30-bp RNA duplex. Unfortunately, there is currently almost no structural and mechanistic data available on how DEAH-box proteins catalyse RNA strand separation.

Many structural and functional studies have evidenced that DExD/H-box proteins undergo conformational transitions in the course of their NTPase and helicase activities. However, the molecular mechanisms by which ATP hydrolysis is coupled to structural changes in the RNA substrates are somehow not fully understood. The greater part of the knowledge about nucleic acid unwinding by SF2 helicases follows from studies of DEAD-box and flavivirin NS3 helicases. However, because SF2 helicases are forming a large and versatile group of functionally and structurally diverse proteins, there is no emerging consensus and several different working models have been proposed. In particular, the recent observations made on the non-processive DEAD-box proteins are quite difficult to transpose on helicases of the DEAH-box family. The latter presents a much greater structural ressemblence to helicases of the Ski2-like family for which little information is currently available. Actually, from a structural perspective, DEAH-box proteins rather look like a sort of hybrid protein family. Their helicase core region shares a remarkable resemblance with that of viral NS3-like helicases whereas a major part of their HA region is structurally similar to the C-terminal WH and Ratchet domains of Ski2-like helicases (113,121). Therefore one of the main challenges lies in adapting the most recent functional models developed for NS3 or Ski2-related helicase to DEAH-box proteins.

Coupling of rNTP hydrolysis with nucleic acid strand separation

First evidence that DExD/H-box proteins convert chemical energy released by rNTP hydrolysis into mechanical force leading to the disruption of nucleic acid structures was obtained by mutagenesis studies. Substitution of a few amino acids in the helicase motifs III, IV and V has indeed been reported

to uncouple NTP-hydrolysis from duplex unwinding without severely affecting binding of rNTP or nucleic acid (41,118,141,150,154-156).

Motif III was initially suggested to couple ATP binding and hydrolysis to helicase activity, since mutation in this motif drastically interfered with substrate unwinding without affecting ATP binding and hydrolysis *in vitro* (150,154-156). However, there is a notable divergence of observations for the effects of the mutations of motif III on different SF2 proteins *in vitro* and *in vivo*. In some helicases, mutations in motif III cause a loss of rNTPase activity (157). Furthermore, the role of this motif remains somewhat ambiguous, as mutations do not always impair protein function *in vivo* (41,115,158,159). Recently, motif III was proposed to facilitate the proper alignment of the two RecA-like domains in response to rNTP binding and to create a high-affinity binding site for RNA (160). In most of the DExD/H-box protein, motif III bridges residues from motifs II and VI as well as the γ-phosphate (27,133,138,139). Indeed, Ser-III and Thr-III residues form hydrogen bonds with His-II and Gln-VI, respectively (Asp_2-II and His-VI in DEAD-box proteins), and Ala-III contacts the γ-phosphate of rNTP through a water molecule (Figure 11A). Therefore, motif III is optimally positioned to coordinate the conformational reorganisation of the helicase core in response to rNTP binding.

Actually, structural studies of DExD/H-box proteins have revealed the existence of a vast network of non-covalent interactions between the various conserved helicase motifs that may convey the nucleotide-binding state to the RNA-binding motifs and contribute thereby to RNA unwinding (28). For instance, in DEAD-box proteins, the rNTP-binding motif VI contacts the RNA-binding motif IV through van der Waals and cation-π interactions between the side chains of the conserved Phe-IV and Arg_1-VI residues (161). Likewise, these two residues are highly conserved among DEAH-box proteins (Figures 2A and 9). In the reported two structures of Prp43, Phe-IV and Arg_1-VI residues are properly positioned to form energetically significant cation-π interactions (-2.49 ± 0.08 kcal·mol^{-1}; Figure 11B) as predicted by CaPTURE (162). Furthermore, as for DEAD-box proteins, mutations of either Phe-IV or Arg_1-VI residues in DEAH-box proteins entailed a loss-of-function phenotype *in vivo* and *in vitro* (Table II and references therein). As Phe-IV mutants are still ATPase-proficient but failed to unwind dsRNA (141), Phe-IV along with Arg_1-VI may couple rNTP binding and hydrolysis to a mechanical step that promotes duplex

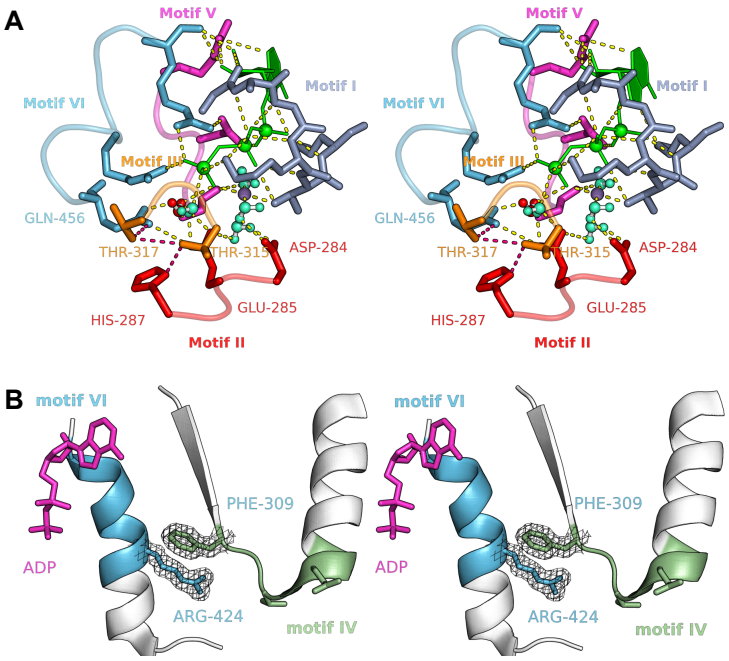


Figure 11 | **Inter-motif interaction network around the rNTP binding site**. (**A**) Close-up view of the NTP binding and hydrolysis site of the dengue virus DENV NS3 helicase (2JLV). In the closed conformation state, motif III (orange) is optimally positionned to bridge helicase motif II (red) in the RecA1 domain with motif VI (light blue) in the RecA2 domain. (**B**) Close-up representation of the inter-motif interaction between the nucleic acid binding motif IV and the rNTP binding motif VI in Prp43. Motifs IV (green) and VI (blue) interact through van der Waals and cation-π interactions through the side chains of Phe-309 (motif IV) and Arg-424 (motif VI). The mesh around residues Phe-309 and Arg-424 depicts the electron density map. In motif IV, amino acids that may form direct contacts with the translocating nucleic acid (Thr-311 and Glu-313) are depicted as sticks.

unwinding. Similarly, $(Ile/Leu)_2$-V and Thr_3-V have also been proposed to link rNTP hydrolysis to the RNA cofactor (118,141). Motif V is unusual with regard to the other helicase motifs, because it is the only motif that interacts with both nucleic acid and ATP (27,122,129). Although Prp22 and Prp43 proteins with mutations in $(Ile/Leu)_2$-V and Thr_3-V residues are still capable to bind RNA and to hydrolyse ATP, they failed to catalyse RNA unwinding and mRNA release from the spliceosome.

Translocation and unwinding mechanism

Almost all DEAH-box proteins tested *in vitro* unwind nucleic acid structures with a 3'-to-5' polarity (Table III and references therein). However, considering the relative paucity of experimental studies, the molecular details of the translocation and unwinding mechanisms for DEAH-box proteins remain purely speculative. Considering the structural resemblance between Prp43 and Hel308 helicases, Walbott et *al.* (113) proposed a mechanism for DEAH-box protein translocation and unwinding based on Hel308 model (29). It was suggested that rNTP binding entails the closure of the inter-domain cleft in the helicase core resulting in a 3'-to-5' translocation of the ssRNA overhang across the RecA1 domain by one nucleotide. This conformational transition is assumed to cause the displacement of the so-called 'ratchet helix' in the HA region and pull the ssRNA across the RecA1 domain. The RecA2 domain is supposed to regain its initial state upon rNTP hydrolysis and/or phosphate release. This would permit the antiparallel β-hairpin from RecA2 domain to disrupt one base-pair from the dsRNA structure (Figure 12). Actually, this model of translocation is reminiscent of the Brownian motor mechanism put forward to explain the 3'-to-5' translocation and nucleic acid unwinding by HCV NS3 helicase (163). According to this model, the helicase alternates between tight (ATP-free, open conformation) and weak (ATP-bound, closed conformation) RNA binding conformational states as a result of its ATPase activity. In the ATP-bound conformation, the helicase form fewer contacts with the ssRNA and could potentially slide either upstream or downstream of its current position. The net unidirectional translocation follows from the presence of residues which function as a ratchet and prevent the helicase to slide backward (122).

Finally, it is currently impossible to predict whether RNA unwinding by DEAH-box proteins resort to a passive or an active mechanism. Recent structural work on HCV NS3 helicase indicates that the Brownian motor model could be compatible with a passive unwinding mechanism (164). According

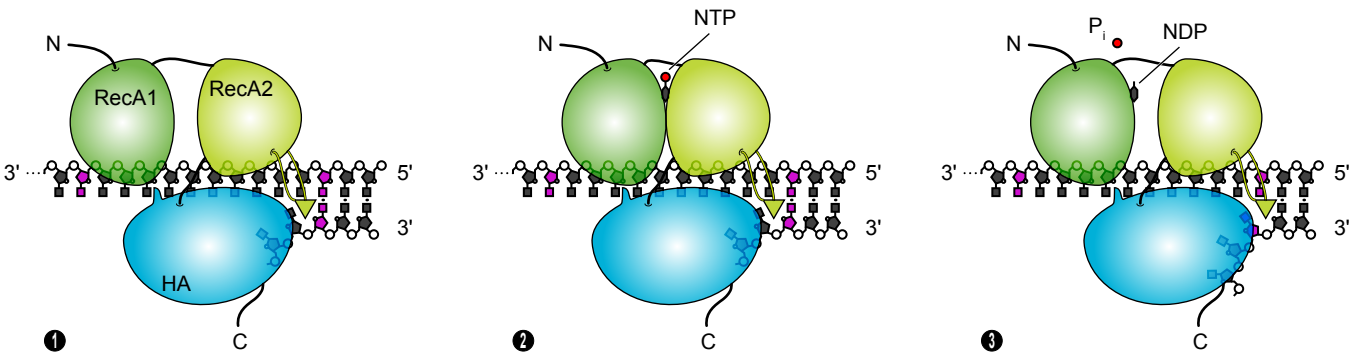


Figure 12 | **Model for the translocation and unwinding mechanism by DEAH-box proteins**. ❶ In the open (NTP-free) conformation state, the helicase is supposed to bind avidly to the translocating ssRNA sequence. ❷ The closure of the inter-domain cleft between the two RecA domains occurs upon cooparative binding of NTP and RNA. This conformational transition is assumed to be accompanied by a translational motion of the ratchet helix in the HA region. The ratchet helix functions as a hook and drags the ssRNA across the RecA1 domain. ❸ NTP hydrolysis and/or phosphate release may allow the RecA2 domain to get back to its inital state. During this last step, the protruding β-hairpin from RecA2 domain may work as a wedge that passively separates the two complementary strands by disrupting base stacking interactions.

to such a model, nucleic acid strand separation would simply result from the directional translocation of the helicase. Considering the structural resemblance between NS3 and the archaeal Hel308 helicase, it was hypotesised that the protruding antiparallel β-hairpin in the RecA2 domain may function as a wedge to physically separates the dsRNA substrate (165). However, one cannot exclude the possibility that NS3-related and DEAH-box proteins resort to a more active unwinding mechanism. As it stands, there is a clear need of more structural and functional studies in order to get a better understanding of the mechanism whereby DEAH-box proteins couple rNTP-hydrolysis to directional translocation and RNA unwinding.

Table III | **Nucleotide specificity, helicase polarity and nucleic acid substrate characteristics of some characterised DEAH-box RNA helicases.** rNTP, unspecific rNTPase activity; dNTP, unspecific dNTPase activity; ND, not determined.

Protein	rNTPase	dNTPase	Helicase activity	Polarity	RNA stem substrate	Ref.
Prp2	rNTP	ND	ND	ND	ND	(107)
Prp16	rNTP	dNTP	Yes	3′ → 5′	18 nt dsRNA	(96,166)
Prp22	rNTP	dNTP	Yes	3′ → 5′	29-40 nt dsRNA	(48,131,141)
Prp43	rNTP	dNTP	Yes	Bidirectional?	30 nt dsRNA	(118)
MLE	rNTP	ND	Yes	3′ → 5′	40 nt dsRNA	(57,148)
RHA	rNTP	ND	Yes	3′ → 5′	29 nt dsRNA	(132)
DHX29	rNTP	ND	Poor	ND	10 nt dsRNA	(83)
CG9323	ND	ND	Yes	ND	5 nt G4-RNA	(88)
RHAU	rNTP	dNTP	Yes	3′ → 5′	5 nt G4-RNA	(89)

Guanine-quadruplex nucleic acid structures

In cells, double-stranded nucleic acid is the most common and widespread form of substrate for helicases. Nonetheless, single strands of nucleic acids can also embrace a multitude of non-canonical and stable secondary structures that also need to be resolved (169-171). Notably, DNA and RNA sequences containing tandem repeats of guanine tracts can form G-quadruplexes (G4) motifs, a thermodynamically stable four-stranded helical arrangement (for reviews, see ref. 172,173). The building block of G4 structures is the G-quartet (aka G-tetrad), a square-planar assembly of four guanines held together by Hoogsteen hydrogen bonding (Figure 13A and B). G4 structures follow from the consecutive π-stacking of two or more G-quartets (Figure 13C) and are further stabilised by alkali metal cations such as Na^+ or K^+ that position along the helix axis and coordinate the O6 keto oxygens of the tetrad-forming guanines (Figure 13B and C; ref. 174,175). G4 scaffolds are extremely polymorphic regarding the relative orientation of strands (parallel, antiparallel or mixed configurations), the glycosidic conformation of guanine (*syn* or *anti*), as well as for the length, sequences and conformations of the loops connecting the G-tracts. In addition, *in vitro*, G4 can follow from the assembly of one (intramolecular G4) or more (bi- or tetramolecular-G4) nucleic acid strands (Figure 13E).

Historical aspect

The propensity of GMP or guanine-rich sequence of nucleic acids to self-assemble into G4 *in vitro* has now been recognised for over 50 years (176). As early as 1910, concentrated solution of guanylic acid were reported to be extremely viscous and, if cooled, to form a clear gel (177). The observation that guanosine 5'-monophosphate (GMP) forms helical aggregates with four units per helical step dates back to the pioneering work in the 1960s by Gellert and Davies (Figure 13B; ref. 167). Such an arrangement of guanines was proposed as an explanation for the remarkable stability of macrostructures resulting from the assembly of tri- and tetranucleotides of deoxyriboguanylic acid observed in the course of the same year by Ralph and co-workers (Figure 13D; ref. 168). Nevertheless, because of the lack of evidence that such structures really exist *in vivo*, G4 structures were considered for more than two decades as a structural curiosity without relevance for living organisms. In the late 1980s and early 1990s, nucleic acid was rediscovered to adopt a variety of atypical structures in addition to the canonical Watson-Crick conformation (178,179). This sudden increase of interest in these singular structures resulted from the discovery that guanine-rich sequences from a variety of functional genomic regions such as telomeres or immunoglobulin class switches could adopt G4 conformations under near physiological conditions (179-183) along with the finding that formation of G4 telomeric structures hindered the activity of telomerase *in vitro* (184).

Evidence for the existence of G4 nucleic acid structures *in vivo*

While G4 structures have been extensively investigated *in vitro*, thus far, direct experimental demonstration of their existence *in vivo* is sorely lacking. Nevertheless, a large number of observations provide circumstantial evidence for the existence and the physiological relevance of G4 structures in cells. Firstly, a large number of biochemical and biophysical studies have shown *in vitro* that stable G4 structures can form spontaneously from G-rich regions of single-stranded nucleic acid under near physiological conditions (173,185). Secondly, genome-wide computational surveys showed that the incidence of potential intramolecular G4-forming sequences (PQS) in the human genome

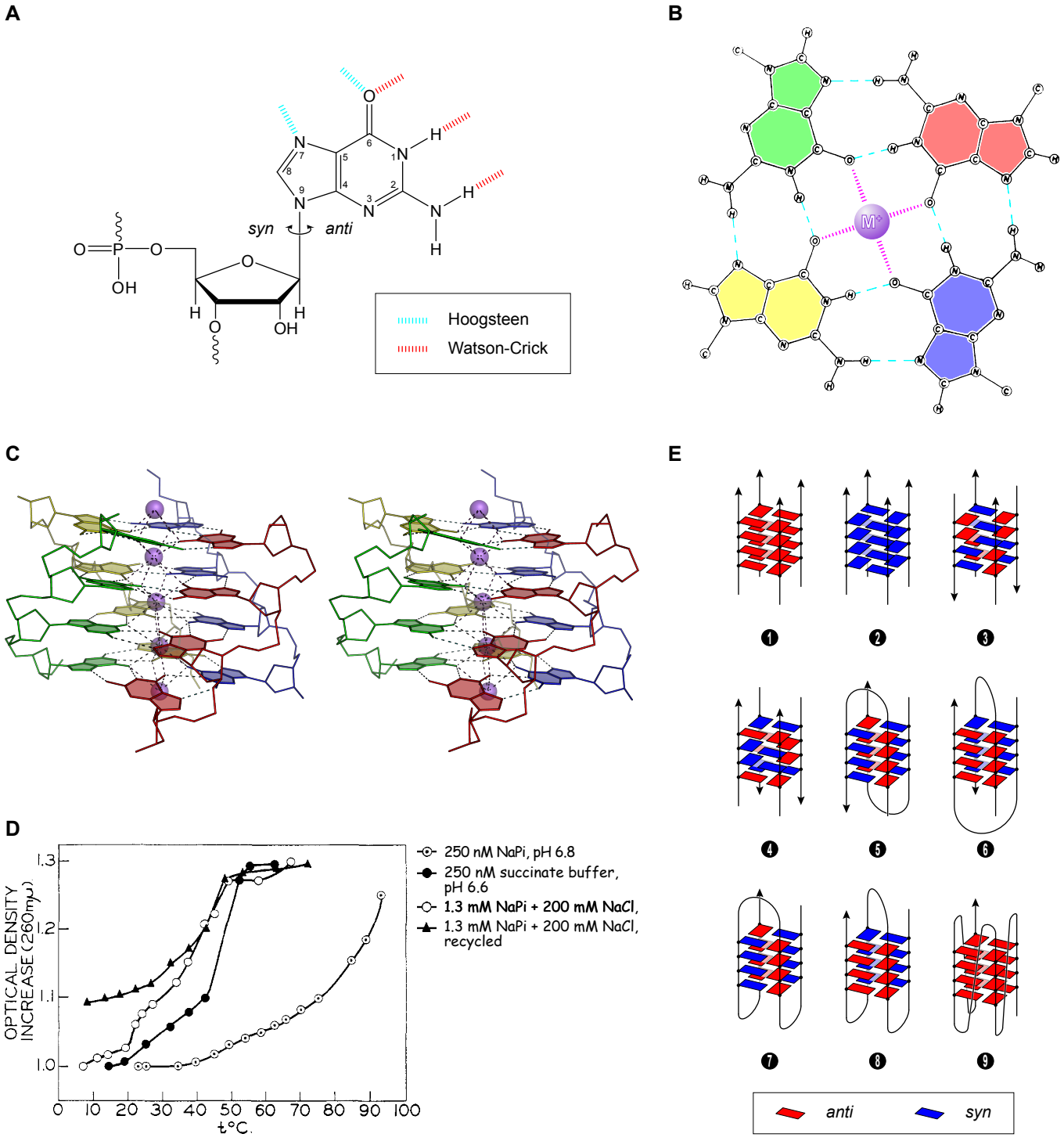


Figure 13 | **Guanine-quadruplex nucleic acid structures**. (**A**) Molecular diagram of the GMP unit in the *anti*-glycosidic conformation. Interconversion between the *anti*- and *syn*-conformations arises upon rotation of the base around the glycosidic bond. In the G-quartet motif, each guanine unit serves as both donor and acceptor of two hydrogen bonds (Hoogsteen bondings). (**B**) A colourised version of the original representation of the G-quartet motif as proposed by Davies and colleagues in 1962 (167). The G-quartet consists of a cyclic and planar array of four guanine bases linked by Hoogsteen hydrogen bonding (cyan dashes). In 1978, Miles and Fraser suggested that G-quadruplex structures could be further stabilised by axial ions. The monovalent or divalent cations (M^+) reside between two successive G-quartets and each ion coordinates (violet dashes) eight O6 keto oxygens. (**C**) Crystal structure of the potassium-containing antiparallel bimolecular G4 scaffold formed from the *Oxytricha nova* telomeric d(GGGTTTTGGG) sequence. The two diagonal d(TTTT) connecting loops are omitted for clarity. (**D**) Thermodynamic stability analysis of DNA macrostructures resulting from the assembly of tetra-deoxyribunucleotide d(GGGG) sequences by Ralph and colleagues in 1962 (168). (**E**) A non-exhaustive schematic representation of the structural diversity for G4 scaffolds. ❶–❹, tetramolecular G4; ❺&❻, bimolecular G4; ❼–❾, intramolecular G4; ❶,❷&❾, all-parallel strand orientation.

is about a third lower than would be expected by chance, suggesting that they were subjected to natural selection over time (186,187). Even so, these studies still identified more than 300'000 PQS within the human genome (186,188) and revealed a higher prevalence of these sequences in functional genomic regions such as telomeres (182,183), promoters (187,189), untranslated regions (UTRs, ref. 190,191), and first introns (192). Taken together, these observations show that PQS are not randomly distributed in the genome and that they may participate in regulating various nucleic acid processes, such as telomere homeostasis (193), the control of gene expression (194) or the initiation of DNA replication (195). In support of these observations, several recent findings concur with the existence of G4 structures in cells. Indeed, both a G4-specific dye and antibodies raised against telomeric G4-DNA specifically stained telomeres in human and ciliate cells, respectively (196-198). In addition, several PQS in promoters were shown to form stable intramolecular G4 structures *in vitro* and to affect gene expression *in vivo* (199,200). A possible contribution of G4 to regulating promoter activity was indicated by impairment of the transcriptional activity of several genes by G4-stabilising ligands (200) or a single-chain antibody specific for intramolecular G4-DNA (201), in a manner correlating with the occurrence of predicted G4 structures in the control regions (202).

Like DNA, RNA can also form G4 structures. Although, to date, G4-RNAs have not attracted as much attention as their DNA counterparts, the formation of G4 structures in RNA is emerging as a plausible regulatory factor in gene expression. RNA is more prone than DNA to form G4 structures due to its single-strandedness, and G4-RNAs have also proved to be generally more stable than their cognate G4-DNA under near physiological conditions (203-206). Bioinformatics analyses of human 5'-UTR sequences revealed potential G4-forming motifs in as many as 3'000 different RNAs (190,207). Moreover, the formation of G4 structures in 5'-UTR was shown to impede translation initiation (190,208-210). Given that PQS have also been identified within IRES (211,212) or near splicing (213-216) and polyadenylation sites (217-219), G4 formation may as well affect RNA metabolism at several different stages. Furthermore, formation of parallel G4-RNA structures has also been reported for telomeric RNA repeats (TERRA, ref. 220,221-223) and for the human telomerase template RNA (TERC/hTR, ref. 224), suggesting that G4-RNA may also occur in non-coding RNAs.

G4 binding and unwinding proteins

Although evidence is accumulating for the existence of G4 structures in cells, it is still unclear whether regulatory functions can be imputed to these structures or whether they should merely be regarded as non-functional misfolded conformations of nucleic acids. Nevertheless, G4 structure formation can constitute an obstacle for gene expression or DNA replication and, hence, be a possible source of nuisance for the cells if not resolved (239-242). However because of their elevated stability, *in vivo* conversion of G4 structures to single-stranded nucleic acid requires specialised proteins with G4 destabilising or unwinding activities. Among these proteins, several DNA helicases such as the RecQ family proteins BLM (227,243), WRN (226) and Sgs1 (244), the Rad3-like family proteins FANCJ/DOG-1 (228) and DDX11 (aka ChlR1, ref. 230), as well as the SF1 helicases Pif1 (Pif1-like helicase, ref. 233,245) and Dna2 (Upf1-like helicase, ref. 246) have been shown to manifest ATP-dependent G4 resolving activity *in vitro* (Table IV). Consistent with these observations, several functional studies underscored the fundamental role played by these helicases in the maintenance of G-rich genomic regions (22,23,228,239,240,247-249). Notably, several inherited human diseases, characterised by genomic instability, high cancer

predispositions or premature aging are caused by loss-of-function mutations in RecQ (Werner and Bloom syndromes), FANCJ (Fanconi anemia) and DDX11 (Warsaw breakage syndrome) helicases (231,250-253).

Table IV | **G4 unwinding polarity, G4 substrate specificity and loss of function phenotype for mammalian G4 helicases.** ND, not determined. Adapted from Wu and Brosh (225).

Protein	Polarity	G4 substrate	Loss of function phenotype	Ref.
WRN	3′ → 5′	G4-DNA	Werner syndrome: premature aging, genomic instability.	(22,226)
BLM	3′ → 5′	G4-DNA	Bloom syndrom: cancer, elevated sister-chromatid exchanges, genomic instability.	(22,227)
FANCJ	5′ → 3′	G4-DNA	Fanconi anemia, breast cancer, defective interstrand cross-link repair and slow S phase progression.	(23,228, 229)
DDX11	5′ → 3′	G4-DNA	Warsaw breakage syndrome: abnormal sister chromatid cohesion, chromosomal breakages.	(230-232)
Pif1	5′ → 3′	G4-DNA	Possible telomere defects.	(233,234)
Dna2	5′ → 3′	G4-DNA	Cell-cycle delay and aberrant cell division, genomic and mitochondrial DNA instability.	(235,236)
RHA	ND	G4-RNA, G4-DNA	Early embryonic (E7.0) lethality for *Rha*$^{-/-}$ knockout mice.	(237,238)
RHAU	3′ → 5′	G4-RNA, G4-DNA	Early embryonic (E7.5) lethality for *Rhau*$^{-/-}$ knockout mice.	(33,89,90)

In the last few years, a growing number of proteins have also been discovered to interact with various synthetic and *bona fide* G4-RNA structures (Table V). Among these G4-RNA binding proteins, two DEAH-box RNA helicases (RHA/DHX9 and RHAU/DHX36) have been shown to unwind synthetic tetramolecular G4-RNA structures *in vitro* (89,238). However little is known regarding the *in vivo* substrate of these helicases as their actual roles as G4 resolving enzymes in cells. Actually, the only reliable evidence of proteins functioning as potential regulators of G4 structures in RNA biology emanate from studies of the fragile X mental retardation proteins FMRP and FMR2P. FMRP is an RNA-binding protein showing high affinity for G4-RNA structures *in vitro* and preferentially binding to RNAs bearing PQS *in vivo* (254). FMRP was proposed to function as a translational repressor through its interaction with G4 structures in RNA. Protein expression of several FMRP target RNAs was notably found to be inversely correlated with levels of FMRP in cells (255-258). Alternatively, FMRP as well as the functionally related FMR2P

Table V | **Mammalian G4-RNA binding proteins and their identified G4-containing target RNAs.** ND, not determined. Adapted from Millevoi et *al.* (260).

Protein	Description	G4-RNA related activity	*In vivo* G4-RNA targets	Ref.
AFF4	AF4/FMR2 family member	G4 binding protein	ND	(261)
CBF-A	hnRNP family member	G4 stabilising protein	*FMR1* 5′-UTR $(CGG)_n$ repeats	(262)
EWS	Ewing sarcoma breakpoint region 1	G4 binding protein	ND	(263)
FMRP	Fragile X mental retardation protein	G4 stabilising protein	*FMR1* exon 15, *TCF19*, *MAP1B*	(254)
FMR2P	AF4/FMR2 family member	G4 binding protein	*FMR1* exon 15	(259)
hnRNP A2	hnRNP family member	G4 destabilising	*FMR1* 5′-UTR $(CGG)_n$ repeats	(262)
hnRNP D0	hnRNP family member	G4 binding protein	ND	(264)
hnRNP H/F	hnRNP family member	G4 binding protein	*TP53* 3′-UTR	(219)
LAF-4	AF4/FMR2 family member	G4 binding protein	ND	(261)
MSP58	58-kDa microspherule protein	G4 binding protein	ND	(265)
RHA	DEAH-box family helicase	G4 unwinding	ND	(238)
RHAU	DEAH-box family helicase	G4 unwinding	ND	(89)
TRF2	Shelterin complex component	G4 binding protein	TERRA	(266)
XRN1	5′ → 3′ exoribonuclease	G4 ribonuclease	ND	(267)

were both shown to act as a splicing regulator of various genes harbouring G4 motifs close to alternative splicing sites (215,259). As for the majority of recently identified G4-RNA binding proteins, there is still a relative lack of functional studies demonstrating causal links between their *in vitro* G4-related biochemical properties and known biological processes. However, this situation will no doubt change considering the abundance of PQS in RNAs and the recent gain of interest in this research field.

The RNA helicase RHAU

RHAU (alias DHX36 or G4R1) is a human RNA helicase of the DEAH-box family. RHAU acronym stands for 'RNA helicase associated with AU-rich element' reflecting the fact the protein was initially identified as a factor associated with the 3'-UTR AU-rich element (ARE) of the urokinase-type plasminogen activator (uPA) mRNA (32).

Structural aspects

RHAU is a 1008 amino acid-long modular protein (Figure 14). It consists of a ~440-amino-acid helicase core comprising all signature motifs of the DEAH-box family of helicases as well as N- and C-terminal flanking regions of about 180 and 380 amino acids, respectively. Like all the DEAH-box proteins, RHAU harbour a HA domain adjacent to the helicase core region that occupies 75 % of the C-terminal region. Based on comparative sequence analysis, RHAU has clear orthologues in almost all species of the animal kingdom ranging from choanoflagellates to humans (Supplementary Figure 5). A multiple sequence alignment between eight of these orthologues showed the helicase core (aa. 183–625) together with the C-terminal region (aa. 626–1008) of RHAU to be evolutionary conserved (Figure 14). In contrast, sequence in the N-terminal region (aa. 1–182) show little similarity with the exception of a cluster of 13 highly conserved amino acids, termed RSM (RHAU specific motif, aa. 54–66; ref. 268). The RSM domain does not share any similarity with any known functional sequence elements and proved to be highly specific to RHAU, hence its name. Vertebrate forms of RHAU harbour also a ~40 amino acid-long low-complexity Gly-rich domain immediately upstream of the RSM domain.

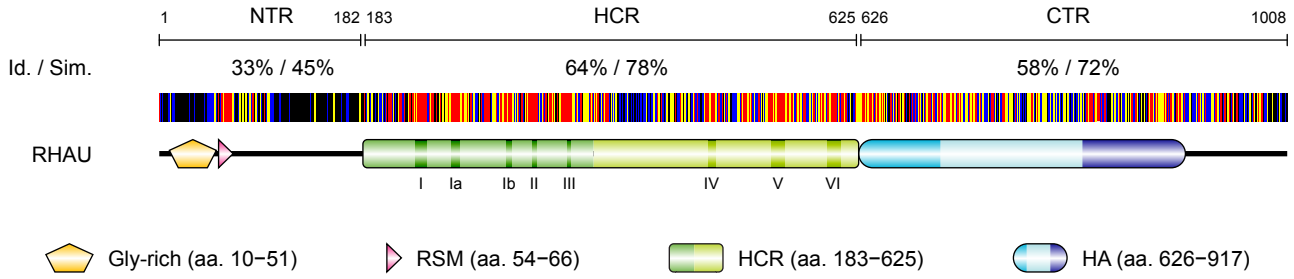


Figure 14 | **Domain architecture of human RHAU.** Schematic representation of the domain organisation and amino acid conservation of the 1008-amino acid RHAU protein. The conserved ATPase/helicase motifs I–VI of the DEAH-box family are indicated within the helicase core region (HCR) by vertical bars. The HCR is flanked by the N-terminal (NTR) and C-terminal (CTR) regions of 182 aa. and 383 aa., respectively. The N-terminal Gly-rich (aa. 10–51) and RSM (RHAU-specific motif, aa. 54–66) domains are indicated. Each residue of human RHAU sequence is represented with a colour code that denotes its degree of conservation amongst eight orthologous sequences (Supplementary Figure 5). Similarity is shown in red for 100 %, yellow for 99–80 % and blue for 79–60 %. Average values of identity (Id.) and similarity (Sim.) for NTR, HCR and CTR regions are indicated.

Subcellular localisation

RHAU is a nucleocytoplasmic shuttling protein found predominantly in the nucleus and to a lesser extent in the cytoplasm (32,269). Further characterisation revealed that RHAU does not localise uniformly in the nucleus, but accumulates in subnuclear structures, termed nuclear speckles, that are enriched in pre-mRNA splicing factors (269). The molecular bases of the translocation of RHAU to the nucleus remain obscure, insofar as it does not harbour an apparent nuclear localisation signal and ATPase-deficient forms of RHAU localise exclusively in the cytoplasm (268,269). The nuclear exclusion of the ATPase-deficient forms of RHAU is not a consequence of protein misfolding since

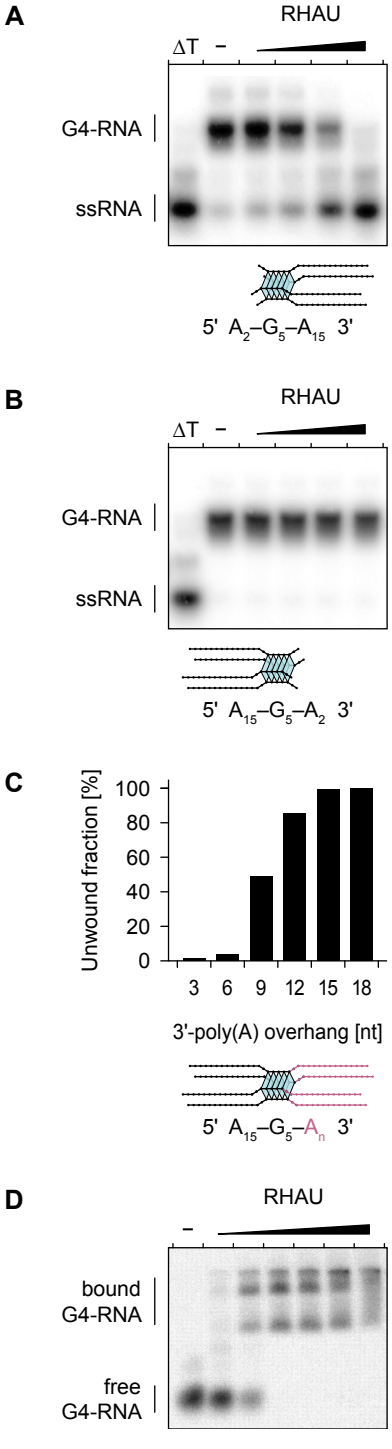


Figure 15 | **G4-RNA unwinding and binding activities of RHAU**. (**A**) 3'-tailed G4-RNA substrate unwinding by RHAU. (**B**) 5'-tailed G4-RNA substrate unwinding by RHAU. (**C**) Analysis of the dependency of the 3'-overhang length on the G4-RNA substrate unwinding activity of RHAU (Vaughn, J.P. and Akman, S.A., unpublished data). (**D**) Tetramolecular G4-RNA binding by RHAU.

ATPase deficient forms of RHAU are stable and still bind to RNA avidly (268). Moreover, ATPase deficient forms of RHAU are not trapped in the cytoplasm by strong and irreversible interaction with RNA since a double mutated form of RHAU showing markedly reduced RNA binding and ATPase activity remained essentially in the cytoplasm (268). Therefore, the translocation of RHAU to the nucleus is probably regulated by its ATPase activity. Likewise, comparable ATPase-dependent nuclear localisation was already reported for the *Xenopus laevis* DEAD-box helicase DDX3 (270).

In addition to its nuclear localisation, RHAU was recently shown to accumulate in stress granules upon translational arrest induced by various environmental stresses (268). Stress granules are temporary sites of accumulation of stress-induced stalled translation initiation RNP complexes (271,272). The recruitment of RHAU to stress granules is mediated by RNA interaction, and deletion experiments revealed that a region of the first 105 amino acids including Gly-rich and RSM domains was critical for RNA binding and relocalisation of RHAU to stress granules (268). Importantly this 105 amino acid-long region alone had the ability *in vivo* to bind RNA and relocalise to stress granules. Furthermore, the ATPase activity of RHAU was shown to play a determining role in regulating *in vivo* RNA interaction and the retention of RHAU in stress granules.

Role of RHAU in uPA mRNA decay

RHAU was initially identified as an ARE-associated factor of uPA mRNA (32). Further characterisation of RHAU revealed that a cytoplasmic isoform that lacks a 14 amino acid stretch between helicase motifs IV and V promoted mRNA deadenylation and decay by recruiting the poly(A) ribonuclease PARN and the exosome to uPA transcripts. In agreement with other known DExD/H-box RNA helicases, the RHAU-mediated decay of ARE^{uPA}-containing transcripts was shown to be dependent on its ATPase activity.

However, despite the apparent role of RHAU in regulating uPA mRNA stability, further investigations showed that its influence on gene expression was essentially not linked with the regulation of mRNA decay process. Indeed, analysis of global gene expression at different time points revealed that only one percent of the mRNAs manifesting differential expression levels upon RHAU knockdown showed corresponding variations in mRNA stability (269). Therefore, RHAU is not really considered as a common regulator of ARE-mediated decay process.

RHAU is a G4 resolvase enzyme with 3′-to-5′ polarity

At about the same time of the finding RHAU facilitates deadenylation and decay of uPA mRNA, RHAU was discovered to harbour *in vitro* ATP-dependant G4 resolvase activity on tetramolecular G4 nucleic substrates (Figure 15A; ref. 33,89). Like other DEAH-box proteins, RHAU is a unidirectional 3'-to-5' helicase and requires a minimal 9-nt 3'-overhang for efficient unwinding activity (Figure 15B and C). RHAU displays a clear preference for G4 substrates over classical double-stranded nucleic acids (ref. 33,89 and Tran, H., unpublished data) and binds various types of G4 structures with sub-nanomolar affinity (Figure 15D; ref. 89). Consistent with these biochemical observations, RHAU was also identified as the major source of tetramolecular RNA-resolving activity in HeLa cell lysates (89). Despite these advances, little is known regarding the fundamental mechanism by which RHAU recognises and resolves G4 structures. Besides, there is a clear lack of information on the RNAs that may serve as physiologically relevant targets for RHAU and about its role as a G4 resolvase in cells.

Murine RHAU is essential for embryogenesis and haematopoiesis

According to BioGPS database (273-275), RHAU is ubiquitously expressed in all human and mouse tissues (Supplementary Figure 6). Relatively high mRNA expression levels are detectable in human haematopoietic cells and mouse brain tissues. Besides, several standard cell lines like HeLa, HEK293, PC-3 and mouse undifferentiated embryonic stem cells (CCE) express high levels of RHAU. Despite its abundance, down-regulation of RHAU by RNA interference (RNAi) did not affect cell growth and viability of many common cell lines, suggesting that RHAU is not an essential protein for cell proliferation. However, to clarify the requirement of RHAU protein *in vivo*, mouse models were developed in which RHAU could be knocked-out (90). As mentioned earlier, RHAU is evolutionary highly conserved and the murine orthologue is nearly identical to the human form. Both proteins display 91 % identity and 95 % similarity and differ only by seven amino acids in length (Supplementary Figure 7). Thus, the mouse form of RHAU is expected to possess the same biochemical properties than its human orthologue. A mouse model in which RHAU can be knocked-out constitutes therefore an excellent tool to investigate *in vivo* the biological role of this protein.

Homozygosity for the null mutant allele of RHAU led to early embryonic lethality with no effect on the implantation efficiency of the embryo (90). Knockout embryos degenerated around seven days post-fertilisation (Figure 16A), revealing the requirement of RHAU for the normal progression of gastrulation. Unfortunately, molecular basis of this requirement has not been further explored. However, similar critical roles of RNA helicases during mammalian embryogenesis have already been reported for the DEAD-box protein p68 (alias DDX5) and the DEAH-box RNA helicase A (237,276). Furthermore, severe developmental defects have also been observed in *C. elegans*, *Drosophila* and zebrafish as a result of abnormal expression of many DEAH-box proteins (53,87,98,277-279).

Considering the high expression levels of RHAU in several lineages of the haematopoietic system (Supplementary Figure 6), a second model of mouse was conjointly developed in which RHAU could be specifically knocked-out in haematopoietic cells (90). Unlike conventional RHAU knockout mice, conditional RHAU knockout animals were viable, although manifesting a severe anaemia (Figure 16B). Further characterisation revealed that conditional knockout mice developed splenomegaly and hepatomegaly (Figure 16C), a phenotype frequently related to anaemia (280). Actually, depletion of RHAU in the haematopoietic system did not exclusively impinge on the development of erythrocytes, but impacted on the entire haematopoietic system (Figure 16D) pointing out the functional relevance of RHAU during haematopoiesis. Taken together, these data demonstrate that RHAU is an essential gene in the mouse.

RHAU senses microbial DNA in human plasmacytoid dendritic cells

A recent study in plasmacytoid dendritic cells (pDCs) has shown that RHAU as well as RNA helicase A take part in the cytoplasmic recognition of CpG motifs (unmethylated CG dinucleotides present in viral DNA, ref. 281). CpG motifs are specifically recognised by Toll-like receptor 9 (TLR9) that leads to the production of anti-viral cytokines type I interferons (IFN-α and IFN-β) through the signal transducer MyD88 protein. However, recent reports suggest the presence of a TLR9-independent recognition of viral DNA that leads to the production of IFN-α by pDCs (282-284). Interestingly, Kim and co-workers

showed that RHAU specifically recognises class A CpGs (CpG-A) and induces the expression of IFN-α through its interaction with MyD88. Furthermore, downregulation of RHAU entailed a 50 % reduction of IFN-α secretion by human pDC cell line in response to Herpes simplex virus infection. Thus, these observations led to the proposal that RHAU is implicated in the detection of viral replication through the recognition of CpG motifs. On the basis of this study, RHAU emerges as a novel component of the innate immune system in humans. Similar critical role of RNA helicases dedicated to sensing microbial nucleic acid have already been reported for several members (IFIH1/MDA5, DHX58/LGP2, DDX58/RIG-I and DICER1) of the RIG-I-like family of helicases (ref. 285,286; Figure 1).

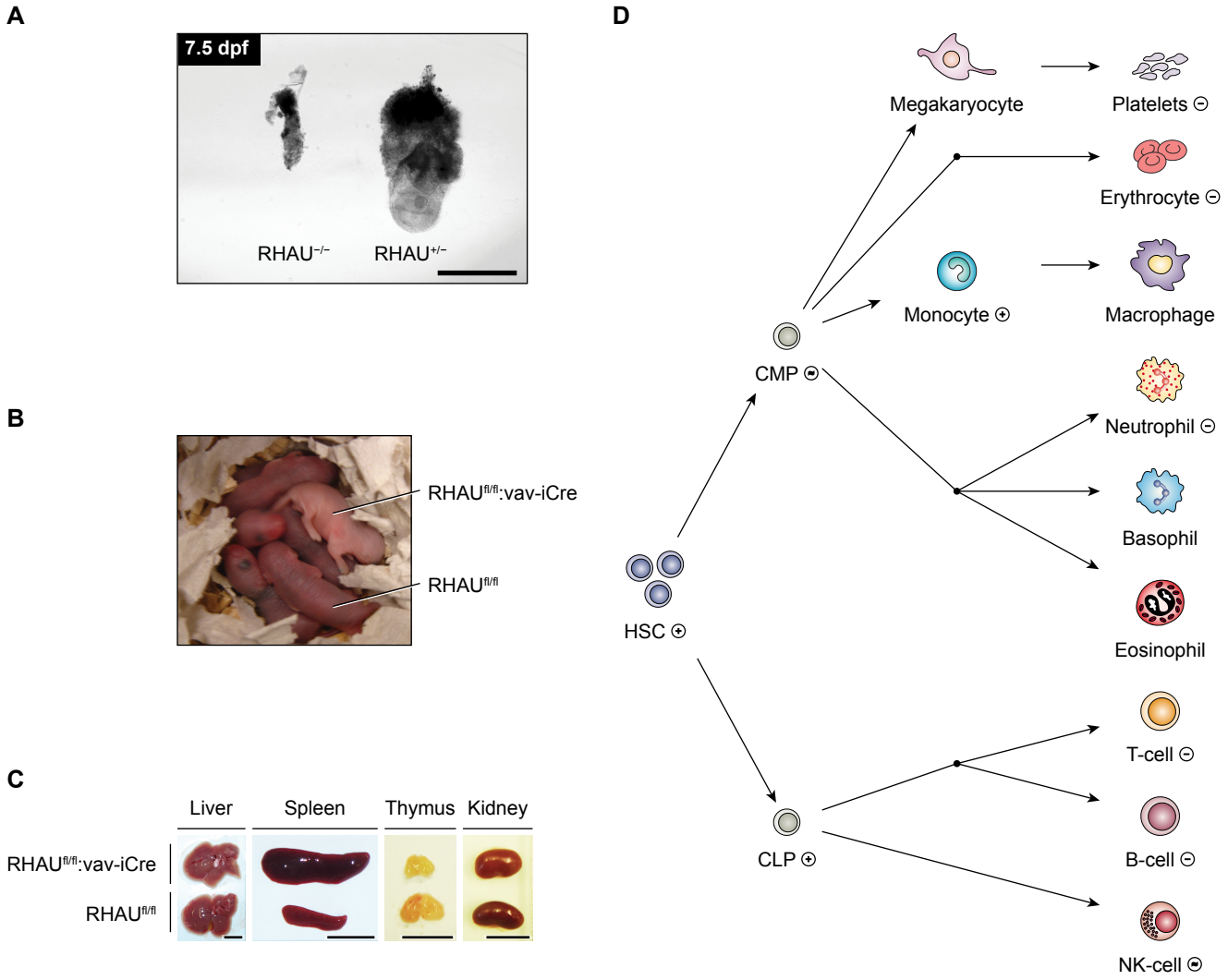


Figure 16 | **Requirement of RHAU for early embryogenesis and haematopoiesis in mouse**. (**A**) Homozygous RHAU knockout (RHAU$^{-/-}$) embryos degenerate before 7.5 days post-fertilisation (dpf). Heterozygous RHAU knockout (RHAU$^{+/-}$) embryos served as a control. Scale bar, 1 mM. (**B**) Anemic phenotype of newborn mice with conditional RHAU knockout (RHAU$^{fl/fl}$:vav-iCre) in the haematopoietic system. RHAU$^{fl/fl}$ mice served as a control. (**C**) Livers, spleens, thymus and kidneys of conditional RHAU knockout (RHAU$^{fl/fl}$:vav-iCre) and control (RHAU$^{fl/fl}$) mice. Scale bar, 1 cm. (**D**) Conceptual schematic diagram of blood cell development. The symbols on the right of cell type names denote the relative abundance of each cell type in conditional RHAU knockout (RHAU$^{fl/fl}$:vav-iCre) animals with respect to control (RHAU$^{fl/fl}$) mice. −, reduced cell number; ≈, normal cell number; +, increased cell number. HSC, haematopoietic stem cell; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; NK-cell, natural killer cell. Adapted from Lai et *al.* (90).

CHAPTER 2

Materials and Methods

Role of the amino terminal RSM domain in the recognition and resolution of guanine quadruplex-RNAs by RHAU

Plasmid constructs, cloning and mutagenesis

The plasmid used for the expression of GST-RHAU(1–200) protein in bacteria and the baculoviral expression vector used for the expression of GST-RHAU have been described previously (32,268). Plasmids for the expression of C-terminal FLAG-tagged recombinant RHAU proteins: RHAU-FLAG, RHAU(ΔGly)-FLAG, RHAU(ΔGly–RSM)-FLAG, RHAU(DAIH)-FLAG and RHAU(1–105)-FLAG have been described previously (268). CG9323 cDNA was obtained from the Drosophila Genomics Resource Center (Bloomington, IN) and was subcloned by PCR amplification into pIRES.EGFP-FLAG-N1 (268). For the RHAU(ΔRSM)-FLAG plasmid, the RSM (aa. 54–66) excision was performed using a standard overlapping PCR method (287,288). The resulting PCR product was subcloned into pIRES.EGFP-FLAG-N1. For RHAU(RSM-mut*x*)-FLAG and CG9323(RSM-mut), the RSM coding sequence was mutagenised using a variation of the classical QuikChange® (Stratagene) site-directed mutagenesis PCR method (289). Construction of all these plasmids was confirmed by sequencing. Sequences of oligonucleotides used in this work and detailed descriptions of the plasmid constructions are available upon request.

Cell culture

Human embryonic kidney HEK293T cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10 % fetal calf serum (FCS) and 2 mM L-glutamine at 37 °C in a humidified 5 % CO_2 incubator. *Spodoptera frugiperda Sf9* cells were maintained in Grace's insect

cells medium (Invitrogen) supplemented with 10 % FCS, 4.11 mM L-glutamine, 3.33 g·l^{-1} lactalbumin hydrolysate and 3.33 g·l^{-1} yeastolate at 27 °C.

Expression and purification of recombinant RHAU proteins

GST-RHAU(1–200) protein was expressed in *E. coli* strain BL21-CodonPlus(DE3)-RIPL (Stratagene). Cultures were inoculated from a single colony of freshly transformed cells and maintained in logarithmic growth at 37 °C in 2× YT medium supplemented with ampicillin (0.1 mg·ml^{-1}) to a final volume of 1 l. When the OD_{600} reached 0.3, the temperature was adjusted to 25 °C and cells were cultured to an OD_{600} of 0.6. Isopropyl β-D-thiogalactopyranoside was added to 0.4 mM and the cultures were incubated for 12 h at 25 °C with constant shaking. Cells were harvested by centrifugation and the pellets stored at −80 °C. All subsequent operations were performed at 4 °C. The cell pellets were resuspended in 50 ml of lysis buffer [1× PBS supplemented with NaCl to a final concentration of 300 mM, 1 % Triton X-100, 5 mM EDTA, 5 mM DTT, 1× protease inhibitor cocktail (Complete EDTA-free, Roche)]. Lysozyme was added to 1 mg·ml^{-1} and the suspensions were lysed for 30 min, followed by sonication (3× 10 s) to reduce viscosity. Insoluble material was removed by centrifugation (39'000*g*, 30 min, 4 °C) in a Beckman JA-17 rotor and the resulting supernatant was filtered through a 0.22-µm Express PLUS Membrane (Millipore). The resulting filtrate was applied to a 1-ml GSTrap 4B column (GE Healthcare). The column was washed with 30 ml washing buffer (1× PBS supplemented with NaCl to a final concentration of 300 mM, 5 mM DTT) and the recombinant protein was recovered in elution buffer [50 mM Tris-HCl (pH 8.0 at 4 °C), 10 mM reduced glutathione, 5 mM DTT]. The fraction containing the recombinant protein were pooled and elution buffer was exchanged to storage buffer [20 mM HEPES-KOH (pH 7.7 at room temperature), 50 mM KCl, 0.01 % Nonidet P-40, 0.5 mM EDTA, 5 mM DTT, 10 % glycerol, 2 mM AEBSF (4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride)] by repeated concentration and dilution steps in Amicon Ultra-15 centrifugal filters (Millipore). Recombinant proteins were stored at −80 °C. Purity of protein preparations was assessed by SDS-PAGE (Supplementary Figure 8) and protein concentrations were determined photometrically at 280 nm using the calculated extinction coefficient ε = 71'905 M^{-1}·cm^{-1}.

GST-RHAU protein was expressed in *Sf9* cells according to the supplier's instructions (PharMingen). Three days post-baculoviral infection, cells were harvested by centrifugation and the pellets were stored at −80 °C. All subsequent operations were performed at 4 °C. The cell pellets were resuspended in insect cell lysis buffer [10 mM Tris-HCl (pH 7.5 at 4 °C), 10 mM sodium phosphate, 300 mM NaCl, 1 % Triton X-100, 10 mM sodium pyrophosphate, 10 % glycerol, 5 mM EDTA, 5 mM DTT, 1× protease inhibitor cocktail (Complete EDTA-free, Roche)] and lysed for 30 min. All subsequent purification steps were carried out as described above for the purification of GST-RHAU(1–200) protein. Purity of protein preparations was assessed by SDS-PAGE (Supplementary Figure 8) and protein concentrations were determined photometrically at 280 nm using the calculated extinction coefficient ε = 166'615 M^{-1}·cm^{-1}.

All C-terminal FLAG-tagged recombinant RHAU and CG9323 proteins [RHAU-FLAG, RHAU(ΔGly)-FLAG, RHAU(ΔGly–RSM)-FLAG, RHAU(ΔRSM)-FLAG, RHAU(RSM-mut*x*)-FLAG, RHAU(DAIH)-FLAG, RHAU(1–105)-FLAG, CG9323-FLAG and CG9323(RSM-mut2)-FLAG] were transiently expressed in HEK293T. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 24–36 h post-transfection, washed with ice-cold PBS and resuspended in lysis buffer (1× PBS supplemented with NaCl to a final

concentration of 600 mM, 1 % Nonidet P-40, 2 mM EDTA, 2 mM AEBSF, 1× protease inhibitor cocktail) for 30 min. Cell lysates were sonicated (1× 10 s) to reduce the viscosity and insoluble material removed by centrifugation (39'000*g*, 20 min, 4 °C) in a Beckman JA-17 rotor. The soluble lysates were mixed for 5 h with 200 µl of a 50 % slurry of anti-FLAG M2-agarose affinity gel (Sigma) that had been equilibrated in lysis buffer. The resin was recovered by centrifugation, washed 3× with 1 ml of lysis buffer follow by 3 washes with 1 ml of IP-washing buffer [50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.1 % Nonidet P-40, 5 mM EDTA]. The bound proteins were eluted with 600 µl elution solution [0.1 mg·ml^{-1} FLAG peptide, 10 mM Tris-HCl (pH 7.5), 150 mM NaCl] for 10 min at 37 °C and stored at −20 °C. Purity of protein preparations was assessed by SDS-PAGE and protein concentrations were determined by Bradford assay with BSA as the standard.

Tetraplex G4-RNA preparation

Unlabelled and 5'-TAMRA-labelled 35-mer oligoribonucleotides 5'-A_{15}–G_5–A_{15}-3' were used to form tetramolecular G4-RNA and are referred to hereafter as 'rAGA'. rAGA were purchased from Dharmacon Research and were dissolved in RNase-free solution [100 mM KCl, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA] to a final concentration of 500 µM. To form tetramolecular quadruplex by annealing of rAGA, the solution was aliquoted into PCR tubes and incubated in a PCR thermocycler at 98 °C for 10 min and then held at 80 °C. EDTA was added immediately to a final concentration of 25 mM and the solution was allowed to cool slowly to room temperature. rAGA aliquots were pooled together and stored at 4 °C for 2–3 days. By this procedure, the conversion of monomeric rAGA to a stable tetramolecular quadruplex form was almost complete as judged by native PAGE (Figure 17A and C). Circular dichroism analysis of the purified, annealed rAGA oligomers revealed a typical spectrum of a parallel G4 structure with positive and negative peaks at 263 nm and 245 nm, respectively (Figure 17B). The above prepared tetramolecular G4-rAGA were stored at −20 °C.

Circular dichroism spectropolarimetry

Circular dichroism (CD) experiments were performed with an AVIV Model 202 spectrophotometer equipped with a thermoelectrically controlled cell holder. G4-rAGA at a concentration of 1 µM were prepared in RNase-free solution [10 mM Tris-HCl (pH 7.5), 1 mM EDTA] supplemented with 50 mM KCl, NaCl or LiCl. Quartz cells with 1 cm path length were used for all experiments. CD spectra were recorded at 25 °C in the UV region (200–350 nm) with 1 nm increments and an averaging time of 2 s.

Thermodynamic analysis of the stability of tetramolecular G4-rAGA structures

Preformed 5'-TAMRA-labelled tetramolecular G4-rAGA at a concentration of 100 nM were prepared in RNase-free solution [10 mM Tris-HCl (pH 7.5), 1 mM EDTA] supplemented with 50 mM KCl, NaCl or LiCl. Tetramolecular G4-rAGA structures were incubated for 5 min at various temperatures ranging from 20 to 99 °C and were immediately subjected to separation by electrophoresis for 3 h on a pre-electrophoresed 10 % polyacrylamide native gel (19:1 acrylamide:*bis* ratio) in 0.5× TBE at 25 °C. After electrophoresis, gels were scanned on a Typhoon 9210 Imager (GE Healthcare) and analysed with Multi Gauge software (Fuji). The fraction of undenatured G4-rAGA was quantitated as the ratio of the signal from the tetramolecular form to the sum of the tetramolecular and

the denatured ssRNA. The apparent temperature of mid-transition (T_m) was determined by representing the fraction of undenatured G4-rAGA as a function of the temperature. Reported T_m values are representative of three independent experiments.

Electromobility shift assay and apparent K_d determination

Recombinant RHAU proteins at concentrations from 1 to 1000 nM were incubated with 100 pM 5'-^{32}P-labelled G4-RNA in K-Res buffer [50 mM Tris-acetate (pH 7.8), 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10 % glycerol], supplemented with 10 mM EDTA and 0.2 U·µl^{-1} SUPERase-In (Ambion) in a 15-µl reaction. The reactions were incubated at 37 °C for 30 min. RNA-protein complexes were resolved on a pre-electrophoresed 6 % polyacrylamide native gel (37.5:1 acrylamide:*bis* ratio) in 0.5× TBE at 4 °C for 90 min. After electrophoresis, gels were fixed for 1 h in 10 % isopropanol/7 % acetic acid. RNA-protein complexes were detected by Phosphor-Imaging, scanned on a Typhoon 9400 Imager (GE Healthcare) and analysed with ImageQuant TL software (Nonlinear Dynamics).

G4-RNA resolvase assay

Recombinant RHAU proteins at concentrations from 1 to 100 nM were incubated with 4 nM 5'-^{32}P-labelled G4-RNA in K-Res buffer supplemented with 1 mM ATP and 0.2 U·µl^{-1} SUPERase-In in a 15-µl reaction. Reactions were allowed to proceed at 30 °C for 30 min, stopped by transfer to ice and addition of 1:10 vol. of 10× loading buffer [33 mM Tris-HCl (pH 8.0), 25 % (w·v^{-1}) Ficoll-400, 110 mM EDTA, 0.17 % SDS]. Reaction products were resolved on a pre-electrophoresed 10 % polyacrylamide native gel (19:1 acrylamide:*bis* ratio) in 0.5× TBE at 4 °C for 90 min. After electrophoresis, gels were fixed for 1 h in 10 % isopropanol/7 % acetic acid and exposed to a Phosphor-Imaging screen.

ATPase assay

Recombinant RHAU proteins at concentrations from 25 to 200 nM were incubated with 1 µl [γ-^{32}P]ATP (3'000 Ci·mmol^{-1}, 0.4 mCi·ml^{-1}) in ATPase assay buffer [50 mM Tris-HCl (pH 8.0), 100 mM KCl, 3 mM $MgCl_2$, 1 mM ATP, 1 mM DTT] supplemented with 0.4 U·µl^{-1} RNasin (Promega) and 1 µg·µl^{-1} homopolymeric poly(U) RNA (Sigma) in a 20-µl reaction. Reactions were allowed to proceed at 37 °C for 15 min. The reactions were stopped by addition of 1 ml of a 5 % (w·v^{-1}) suspension of activated charcoal (Sigma) in 20 mM phosphoric acid. The samples were incubated on ice for 10 min and the charcoal containing the adsorbed unhydrolysed ATP was pelleted by centrifugation (21'000*g*, 15 min, 4 °C). The supernatants (containing free γ-$^{32}P_i$) were transferred to new tubes and the radioactivity was quantified by liquid scintillation counting (Cerenkov counts).

RHAU binds an intramolecular G4 structure in TERC and associates with telomerase holoenzyme

Plasmid constructs, cloning and mutagenesis

The baculoviral expression vector employed for the expression of GST-RHAU and the plasmids used for the expression of the C-terminal FLAG-tagged recombinant RHAU proteins RHAU-FLAG and RHAU(ΔRSM)-FLAG were previously described (32,88,268). The pIRES.EGFP-myc-N1 vector expressing the C-terminal myc-tagged recombinant RHAU protein was derived from the previously described pIRES.EGFP-FLAG-N1/RHAU expression vector (268) by substituting the FLAG sequence with the myc epitope sequence. Human TERT cDNA (clone: IOH36343, mapped sequence: NM_198253) was obtained from ImaGenes GmbH (Berlin, Germany) and subcloned by PCR amplification into the pSL1-FLAG-N1 mammalian expression vector. The pSL1-FLAG-N1 vector was derived from pEGFP-N1 (Clontech) by replacing the EGFP open reading frame with the FLAG epitope sequence. The human TERC genomic region was PCR-amplified from the genomic DNA of HEK293T cells to yield a 2.1-kb fragment encompassing 1080 bp of the TERC promoter, its coding sequence, and 553 bp of the 3' flanking genomic region. This fragment was blunt-cloned into pGEM-T Easy vector (Promega) at the HincII/EcoRV sites. The G4 motif sequence of TERC was mutagenised using a variation of the classical QuikChange® (Stratagene) site-directed mutagenesis PCR method (289). To prepare templates for *in vitro* run-off transcription, the T7 or SP6 phage promoters were inserted upstream of the TERC coding sequence by PCR. The resulting PCR products were cloned into the pSL1-FLAG-N1 vector at the NheI/AgeI sites. Following linearization with NarI or AgeI, *in vitro* transcription of these templates yielded the TERC(1–71 nt) and full-length TERC(1–451 nt) RNA fragments, respectively. Constructions of all these plasmids were confirmed by sequencing. Sequences of oligonucleotides used in this work and detailed descriptions of the plasmid constructs are available upon request.

Cell culture and transfection

Human cervical carcinoma HeLa and embryonic kidney HEK293T cell lines were maintained in Dulbecco's modified Eagle's medium supplemented with 10 % fetal calf serum (FCS) and 2 mM L-glutamine at 37 °C in a humidified 5 % CO_2 incubator. Transient transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transfected cells were cultured for 24-36 h prior to testing for transgene expression. T-REx™-HeLa cell lines stably transfected with pTER-shRHAU or pTER-shGFP were maintained as described previously (269). To induce and maintain short hairpin RNA (shRNA) expression, cells were cultured in the presence of 1 $\mu g \cdot ml^{-1}$ doxycycline for at least 7 days prior to testing for RHAU down-regulation efficiency.

RIP-chip assay

Cells were harvested 24-36 h post-transfection, washed with ice-cold PBS and resuspended in lysis buffer [1× PBS, 1 % $v \cdot v^{-1}$ Nonidet P-40, 2 mM EDTA, 2 mM AEBSF, 1× protease inhibitor cocktail (Complete EDTA-free, Roche), 0.2 $U \cdot \mu l^{-1}$ RNasin® Plus (Promega)] for 30 min. All subsequent operations were performed at 4 °C. The lysates were cleared by centrifugation (21'000*g*, 15 min) and mixed with 10 µl of a 50 % slurry of anti-FLAG M2-agarose affinity gel (Sigma)

that had been equilibrated in lysis buffer. After gentle agitation for 5 h, the resin was recovered by centrifugation, washed 3× with 500 µl lysis buffer, followed by three washes with 500µl IP-washing buffer [50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 0.1 % v·v^{-1} Nonidet P-40, 5 mM EDTA, 0.4 U·µl^{-1} RNasin® Plus]. The resin was resuspended in 1 ml TRIzol (Invitrogen) for protein analysis and RNA extraction. For microarray analysis, 100 ng RNA was converted to cRNA according to the manufacturer's guidelines and the reaction products hybridised to GeneChip® Human Gene 1.0 ST arrays (Affymetrix). The resulting raw expression values were RMA-normalised using R/BioConductor (290) and the Oligo Package (version 1.14.0, ref. 291). Probesets were linked to Entrez Gene entries using Affymetrix annotation (NetAffx release 28, March 11, 2009), retaining a single probeset per gene. Genes not clearly detected in input samples [average $\log_2$(expression value) < 6.0] were discarded. Differential expression was determined using the Limma package (292), selecting genes with a minimal fold-change of 2.0 and an FDR-adjusted *P*-value of less than 0.01. RHAU targets were identified as transcripts enriched in RHAU-FLAG IP versus RHAU-FLAG input samples but not enriched in control IP versus control RHAU-myc input samples.

G4-RNA structure prediction and bioinformatics analysis

Human RNA sequences were retrieved from Entrez Nucleotide database. Non-overlapping putative intramolecular G4-forming sequences (PQS) and the corresponding G4-score values were computed with the QGRS Mapper algorithm (293) using the default parameters (window size = 30 nt, Min. G-group = 2, loop size = 0-36). For each transcript, the ∑(G4-score) value was calculated as the sum of all non-overlapping G4-scores computed by QGRS Mapper. To obtain a normalised ∑(G4-score) value, the ∑(G4-score) value was divided by the RNA length (kb). Randomisation of RNA sequences by single- or dinucleotide shuffling was performed using the Altschul-Erickson algorithm (294).

Protein immunoprecipitation assay

Protein immunoprecipitation experiments were performed under the same conditions employed for the RIP-chip assay. Cleared cell lysates were mixed with rProtein A or Protein G Sepharose™ Fast Flow beads (GE Healthcare) and appropriate antibodies [mouse mAb anti-RHAU (12F33, ref. 268), rabbit pAb anti-dyskerin (H-300, Santa Cruz Biotechnology)]. After gentle agitation for 5 h, the beads were washed. For telomere repeat amplification protocol (TRAP) assays, beads were resuspended in 40 µl TRAP lysis buffer [10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM $MgCl_2$, 1 mM EDTA (pH 8.0), 1 % v·v^{-1} Nonidet P-40, 0.25 mM Na-deoxycholate, 10 % v·v^{-1} glycerol, 5 mM 2-mercaptoethanol, 0.1 mM AEBSF]. For protein and RNA analysis, beads were resuspended in 1 ml TRIzol and extraction performed according to the manufacturer's instructions. Input and co-purified RNA samples were analysed by RT-qPCR. For protein analysis only, beads were directly resuspended in sodium dodecyl sulphate (SDS)-gel loading buffer. Input and immunoprecipitated protein samples were separated by SDS-PAGE and analysed by Western blotting.

RNA analysis by quantitative (RT-qPCR) and semi-quantitative RT-PCR

Reverse transcription was performed using the ImProm-II™ Reverse Transcription System (Promega) with oligo$(dT)_{15}$ or random hexamer primers, according to the manufacturer's instructions. For monitoring first-strand cDNA

synthesis, the reverse transcription reaction was performed in the presence of 20 µCi [α-^{32}P]dATP (3'000 Ci·mmol^{-1}). The reaction products were separated by agarose gel electrophoresis and visualised by Phosphor-Imaging. Quantitative and semi-quantitative PCR reactions were performed in technical duplicates using the ABsolute™ QPCR SYBR® Green ROX Mix (Thermo Fisher Scientific), according to the manufacturer's instructions, on an ABI Prism 7000 Sequence Detection System (SDS) and analysed with the ABI Prism 7000 SDS 1.0 Software (Applied Biosystems). Relative transcript levels were determined using the $2^{-\Delta Ct}$ method (295). For each primer pair (Supplementary Table I), the efficiency of amplification was determined to be equal or superior to 1.8. Control reactions lacking the reverse transcriptase or template RNA confirmed the specificity of the amplification reactions.

TRAP assays

Immunopurified ribonucleoprotein (RNP) complexes were assayed for telomerase activity by the TRAP assay (296). Four microliters of bead-immobilised RNP complexes in TRAP lysis buffer were incubated (30 min, 25 °C) in 50 µl TRAP reaction buffer [20 mM Tris–HCl (pH 8.3 at room temperature), 63 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix (50 µM each of dATP, dTTP, dGTP and dCTP), 0.05 % v·v^{-1} Tween®-20, 1 mM EGTA (pH 8.0), 4 ng·µl^{-1} Cy5-TS primer, 2 ng·µl^{-1} ACX primer (Supplementary Table I), 1 U·µl^{-1} RNasin® Plus, 0.4 mg·ml^{-1} BSA, 0.04 U·µl^{-1} Thermo-Start™ *Taq* DNA Polymerase (Thermo Fisher Scientific)]. The reaction was followed by a 15-min incubation step at 95 °C, followed by 15-17 cycles of amplification (30 s at 95 °C, 30 s at 52 °C, 45 s at 72 °C). The reaction products were resolved on a pre-electrophoresed 10 % non-denaturing polyacrylamide gel (19:1 acrylamide:*bis* ratio) in 0.5× TBE at 4 °C for 90 min. After electrophoresis, gels were fixed [500 mM NaCl, 50 % v·v^{-1} ethanol, 40 mM Na-acetate (pH 4.2)] for 30 min, scanned on a Typhoon 9400 Imager (GE Healthcare) and analysed with ImageQuant TL software (Nonlinear Dynamics). For immunodepletion experiments of RHAU, the residual telomerase activity in cell extracts was quantified by the quantitative TRAP assay (qTRAP, ref. 296). An aliqot of 250 ng protein from immunodepleted HEK293T cell extracts in TRAP lysis buffer was incubated (30 min, 25 °C) in 25 µl qTRAP reaction buffer [1× ABsolute™ QPCR SYBR® Green ROX Mix, 1 mM EGTA (pH 8.0), 4 ng·µl^{-1} TS primer, 4 ng·µl^{-1} ACX primer (Supplementary Table I), 0.2 U·µl^{-1} RNasin® Plus]. The reaction was followed by a 15-min incubation step at 95 °C, followed by 40 cycles of amplification (15 s at 95 °C, 60 s at 60 °C) on an ABI Prism 7000 Sequence Detector. The relative telomerase activity was determined using a standard curve and linear equation model (296).

Expression and purification of recombinant RHAU protein

Recombinant wild-type and ATPase deficient [RHAU(DAIH)] N-terminal GST-tagged RHAU proteins were expressed in *Sf9* cells according to the supplier's instructions (PharMingen) and purified to homogeneity as described previously (88). Purified recombinant GST-RHAU proteins were stored at −80 °C. Purity of protein preparations was assessed by SDS-PAGE (Supplementary Figure 11) and protein concentrations determined photometrically at 280 nm using the calculated extinction coefficient ε = 166'615 M^{-1}·cm^{-1}.

In vitro synthesis of ^{32}P-labelled TERC transcripts and intramolecular G4-RNA preparation

Synthetic radio-labelled wild-type and mutant (G4-MT) telomerase RNAs were prepared by *in vitro* transcription using 50 μCi [α-^{32}P]UTP (3'000 Ci·mmol^{-1}) and T7 or SP6 RNA polymerases (Promega), respectively, according to the manufacturer's instructions. The transcripts were purified by denaturing PAGE, ethanol precipitated and recovered by centrifugation. The purified RNAs were resuspended in potassium- or lithium-based storage buffer [10 mM Li-cacodylate (pH 7.4), 100 mM KCl or LiCl] and annealed by heating at 95 °C for 2 min, followed by slow cooling to room temperature. G4-annealed radio-labelled RNAs were stored at −80 °C.

RNA electromobility shift assay (REMSA)

Purified recombinant GST-RHAU protein at concentrations from 1-320 nM were incubated with 100 pM ^{32}P-labelled G4-RNA in K-Res buffer [50 mM Tris-acetate (pH 7.8), 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10 % glycerol], supplemented with 10 mM EDTA and 0.2 U·μl^{-1} SUPERase-In (Ambion) in a 10-μl reaction. The reactions were equilibrated at 22 °C for 30 min. RNA-protein complexes were resolved on a pre-electrophoresed 6 % non-denaturing polyacrylamide gel (37.5:1 acrylamide:*bis* ratio) in 0.5× TBE at 4 °C for 90 min. After electrophoresis, gels were fixed for 1 h in 10 % isopropanol/7 % acetic acid. RNA-protein complexes were detected by Phosphor-Imaging, scanned on a Typhoon 9400 Imager (GE Healthcare) and analysed with ImageQuant TL software (Nonlinear Dynamics).

Structural model of the helicase core and HA regions of RHAU

Homology modelling and model quality estimation

The three-dimensional (3-D) model of RHAU was built by homology modelling based on the high-resolution (1.9 Å) and high-quality (R = 0.184, R_{free} = 0.224) X-ray structure of the *S. cerivisiae* Prp43 protein (RCSB PDB ID: 2XAU chain A, ref. 113). The pairwise sequence alignment between RHAU and Prp43 template was obtained from a multiple sequence alignment of various *E. coli*, *S. cerevisiae*, *D. melanogaster* and *H. sapiens* DEAH-box proteins (Supplementary Figure 1). Homology modelling was carried out using the MODELLER (v. 9.9) program (297,298). Quality assessment of the modelled RHAU structure was carried out with QMEAN6 (299) and ModFOLD (300) tools. Three-dimensional protein structure superposition and calculations of the RMSD values were performed using the CEalign function in PyMOL (v. 1.5, ref. 301) by fitting C_α-backbones of the analysed protein molecules. The ADP bound model of RHAU was generated by superposing the helicase core region of Prp43·ADP complex to that of RHAU and masking Prp43 structure. The nucleic acid bound model of RHAU was obtained by fitting Hel308·DNA complex (RCSB PDB ID: 2P6R) to that of RHAU and hiding Hel308 structure.

CHAPTER 3

Results

Characterisation of the tetramolecular G4 structure formed by rAGA oligoribonucleotides

AIMS AND RATIONALE

In vivo tetramolecular G4 structures are probably less frequent than intramolecular G4 structures because of their very slow folding kinetic, however they have been extensively employed for *in vitro* investigations of protein with G4 binding and resolving activities (33,89,227-229,243,302,303). Despite their somewhat unusual aspect, tetramolecular G4 structures may be seen as simpler models of biologically relevant G-tetrads (204). The conformation of the stem of the tetramolecular quadruplexes is indeed very close to the central core of intramolecular parallel quadruplexes found in human telomeric repeats or other motifs (304,305). Importantly, the use of a tetramolecular G4 structure facilitates unwinding analysis, insofar as the tetraplex form can be readily separated from the unwound single stranded product by electrophoresis. The investigation of the unwinding of intramolecular G4 structure is indeed tricky and requires trapping the unwound nucleic acid with a complementary strand to prevent its readoption of G4-conformation. Nevertheless, the usage of a tetramolecular G4 substrate to characterise the G4 resolvase activity of RHAU does not deprive its biological significance insofar as we also provide evidence that RHAU binds intramolecular G4 structures.

RESULTS AND DISCUSSION

In a part of this work, we employed the tetramolecular G4-rAGA substrate to characterise the G4-RNA binding and resolving activity of RHAU *in vitro*. AGA oligonucleotide was already employed as a G4-DNA substrate to characterise the enzymatic activity of RHAU (33). G4-rAGA

substrate was derived from the Z33 oligomer by substituting the 3' and 5' overhangs flanking the guanine core with two tails of 15 adenosine residues. Considering that adenosine cannot base-pair with itself or with the five central guanosine residues, we could in that way prevent the formation of undesired secondary structures that would bias the study of G4 binding and resolving activity of RHAU.

As shown in Figure 17A, rAGA oligoribonucleotides form a high ordered structure when they are incubated at a high concentration (micromolar range) in a potassium-containing solution. As analysed by native PAGE, the characteristic monomer band was also diminished, indicating the formation of a larger RNA structure in a manner dependent on RNA concentration. By comparison with a 25-bp dsRNA resulting from the hybridisation of rAGA oligoribonucleotide with its complementary strand 5'-U_{15}–C_5–U_{15}-3' (rUCU), the size of the slow mobility band turned out to be of higher complexity, since its migration was much slower than that of the dsRNA species. When diluted at a low concentration (nanomolar range) in a solution deprived of monovalent cations (K^+, Na^+ or Li^+), the rAGA oligoribonucleotide remained in a monomeric state. The conversion of the monomeric form to the high ordered structure was also suppressed when two guanines of the central core were replaced with adenines. This demonstrates the significance of the five central guanines for the stability of the structure. In contrast, the flanking adenosine residues could all be substituted with uridines without affecting the formation of the high ordered structure indicating that they were not essential for the self-assembly of rAGA oligomer.

Circular dichroism analysis of the purified high ordered structure formed by rAGA oligomers revealed a positive and a negative peak at 263 nm and 245 nm, respectively (Figure 17B). This characteristic CD spectrum signature with a strong maximum ellipticity at 263 nm is typical of a parallel G4 structure whose all stacked quartets have the same polarity (306-308). Antiparallel G4 structures have indeed been shown to have a different profile with positive and negative peaks at 295 nm and a 265 nm, respectively. In fact, in G4-RNA structures, the parallel strand conformation is strongly favoured over the antiparallel one. This distinctive feature arises because of the C3'-*endo* sugar pucker conformation of the ribose (as opposed to the C2'-*endo* conformation of deoxyribose) that entails the base to adopt the *anti* conformation (309). Circular dichroism spectrum analysis also revealed that the amplitude of the signal at 263 nm and 245 nm was reduced in the absence of cations. This suggests that the G4-rAGA structure is more stable in the presence of monovalent cations, which is consistent with the known alkali metal cation dependency of G4 structures (174,175,179). Although K^+ and Na^+ ions are known to be better stabilisers of G4 structures than Li^+ ions, this effect is not apparent at 25 °C but occurs at higher temperatures (Figure 17C and D).

Compared to conventional double-stranded nucleic acids, tetramolecular G4 structures are extremely stable (170,173,185,204,310). Moreover the thermal stability of most of them shows a strict alkali cation dependency in the order of: $K^+ > Na^+ >> Li^+$. From a kinetic point of view, formation of tetramolecular G4 structures is a very slow process that follows a fourth-order reaction with respect to strand concentration. In contrast, the dissociation of tetramolecular G4 structures follows a first order kinetic. Therefore at low concentrations, thermal denaturation of short parallel quadruplexes is often irreversible. We actually took advantage of this irreversibility to determine the thermodynamic stability of preformed G4-rAGA structure as a function of the nature of the monovalent cation by electrophoresis for direct visualisation. As shown in Figure 17C and D, the apparent temperature of mid-transition (T_m) of G4-rAGA structure ranged from 60 °C in LiCl to more than 100 °C in KCl with

an intermediate value of 88 °C in NaCl. This result shows that K^+ stabilises G4-rAGA structure better than Na^+ or Li^+, which is consistent with the known hierarchy of tetramolecular G4 structure stabilisation by different monovalent cations (179). Taken together, these data provide strong evidence that rAGA oligoribonucleotides form a stable and parallel interstrand G4 structure and its high ordered form constitutes an adequate substrate to investigate the G4-RNA binding and resolving activity of RHAU protein.

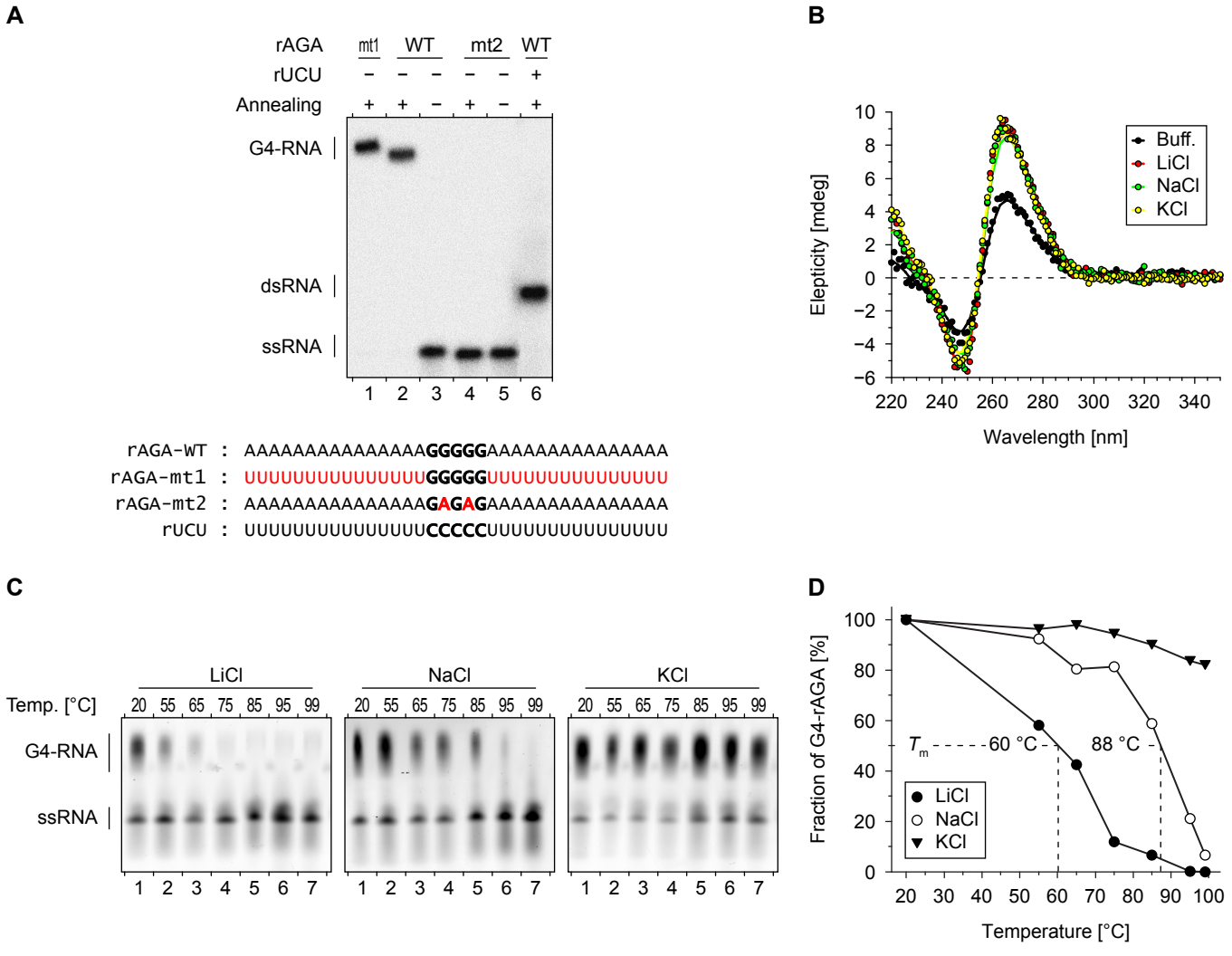


Figure 17 | **Characterisation of the tetramolecular G4-rAGA structure**. (**A**) Native PAGE analysis of monomeric and annealed forms of rAGA oligomer. The positions of the single-stranded oligoribonucleotides (ssRNA) and of the annealed double-stranded (dsRNA) or tetraplex (G4-RNA) RNA are denoted on the left. Sequences of the employed oligomers are indicated underneath with mutated residues marked in red. (**B**) Circular dicroism spectra of G4-rAGA structure at 25 °C in the presence of 50 mM of either LiCl, NaCl or KCl. (**C**) Thermodynamic stability analysis of preformed G4-rAGA structure as a function of the nature of the monovalent cations. 5'-TAMRA-labelled tetramolecular rAGA at a concentration of 100 nM were incubated at various temperatures for 5 min in the presence of 50 mM of either LiCl, NaCl or KCl and immediately subjected to separation by electrophoresis under native conditions. (**D**) Determination of the apparent temperature of mid-transition (T_m) of G4-rAGA structure as a function of the nature of the monovalent cations.

Role of the amino terminal RSM domain in the recognition and resolution of guanine quadruplex-RNAs by RHAU

AIMS AND RATIONALE

RHAU displays high affinity and specificity for G4 nucleic acid structures *in vitro* (33,89), however little is know regarding the mechanisms by which RHAU recognises and resoves its substrates. Previous work showed that RHAU associates with mRNAs and relocalises to stress granules upon translational arrest induced by various environmental stresses (268). Deletion analysis of the N-terminal region of RHAU revealed that a region of the first 105 amino acids was critical for RNA binding and relocalisation of RHAU to stress granules. Importantly, this 105 amino acid-long region alone had the ability *in vivo* to bind RNA and relocalise to stress granules. The apparent significance of the first 105 amino acids for RHAU interaction with RNA prompted us to determine whether the N-terminal domain of RHAU also contributes to the recognition of G4 structures *in vitro*. To address this question, a series of N-terminal deletion mutants of RHAU was generated as shown in Figure 18A. Through biochemical analysis of these mutants, we have uncovered the functional importance of the first 105 amino acids of RHAU for interaction with G4 structures and further revealed among this region that the previously identified RSM (RHAU-specific motif, aa. 54–66) domain (268) is essential to the high affinity for G4 structures shown by RHAU. Here, we show that the N-terminal region of RHAU alone can bind to G4-RNA structures, albeit with lower affinity than the full-length protein. Besides, we also show that the G4 resolving activity of RHAU is conserved in higher eukaryotes, insofar as CG9323, the *Drosophila* orthologue of RHAU, readily unwinds G4 structures in the presence of ATP. Finally, we show that a variant form of CG9323 harbouring mutations in the RSM domain manifests reduced G4 binding activity, suggesting that RSM functions similarly in *Drosophila* and human RHAU proteins.

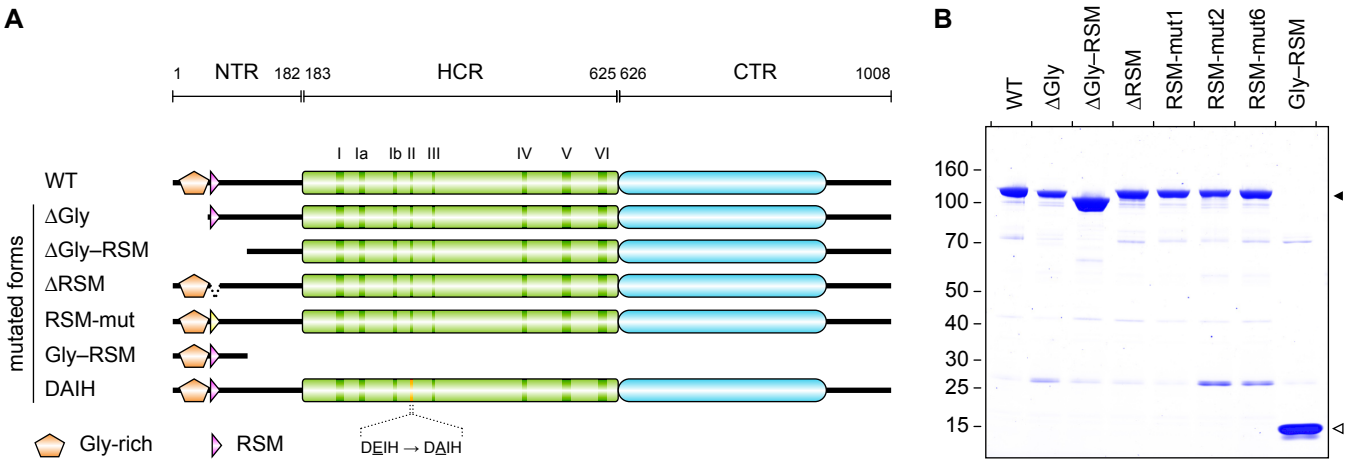


Figure 18 | **N-terminal deletion mutants of RHAU**. (**A**) Schematic representation of the 1008-aa. RHAU protein and its N-terminal truncated mutants. The conserved ATPase/helicase motifs I–VI of the DEAH-box family are indicated within the helicase core region (HCR) by vertical bars. The HCR is flanked by N-terminal (NTR) and C-terminal (CTR) regions of 182 aa and 383 aa, respectively. The N-terminal Gly-rich (aa 10–51) and RSM (RHAU-specific motif, aa. 54–66) domains are indicated. WT, 1–1008 aa.; ΔGly, 50–1008 aa.; ΔGly–RSM, 105–1008 aa.; ΔRSM, RHAU harbouring a deletion of aa. 54–66; RSM-mut, RHAU with mutagenised RSM (see Figure 23A for details about aa. substitutions); Gly–RSM, 1–105 aa.; DAIH, D335A ATPase deficient mutant. (**B**) SDS-PAGE separation and Coomassie staining of purified FLAG-tagged recombinant wild-type and mutant RHAU proteins (2 µg protein per lane). The positions and sizes (kDa) of marker proteins are indicated at the left. The marks (◀ and ◁) on the right indicate the positions of the purified proteins.

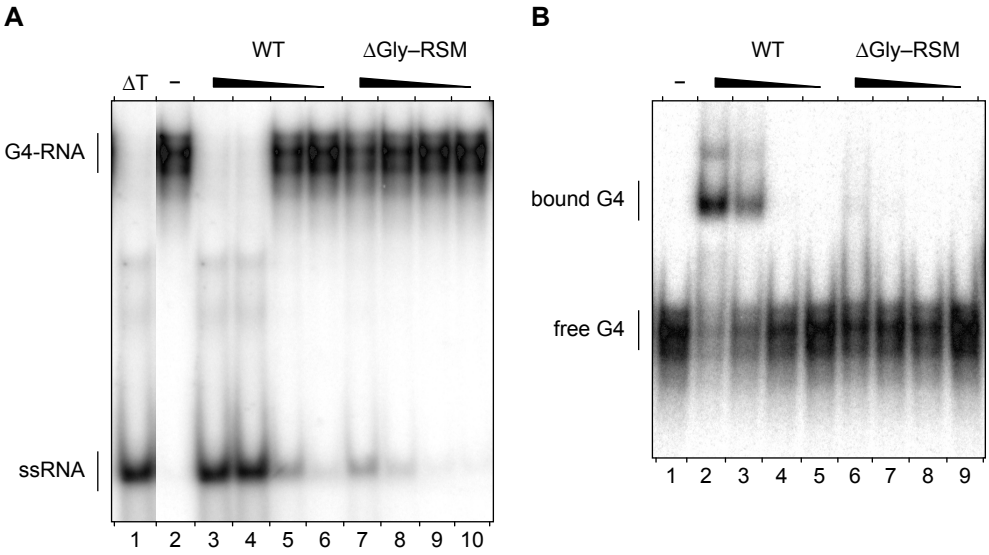


Figure 19 | **G4-RNA unwinding and binding by N-terminal truncated RHAU proteins**. (**A**) G4-RNA unwinding assay: radiolabelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (1, 3, 10 and 30 nM) of wild-type (WT) RHAU or ΔGly–RSM mutant. The reaction products were resolved by native PAGE after disrupting RNA-protein interactions with SDS. An autoradiogram of the gel is shown. The positions of the tetraplex substrate RNA (G4-RNA) and the unwound single-stranded product (ssRNA) are denoted on the left. An aliquot of the tetraplex substrate that was heat-denatured (95 °C, 5 min) and then quenched (ΔT) serves as a marker for the position of single-stranded RNA. (**B**) G4-RNA binding assay: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (1, 3, 10 and 30 nM) of WT RHAU or ΔGly–RSM mutant in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. The positions of the free tetramolecular RNA substrate and the protein-RNA complex are indicated on the left.

RESULTS

The first 105 amino acids of RHAU are required for binding and resolving G4 structures

We demonstrated previously *in vivo* and *in vitro* that amino acids 1–105 of RHAU delineate a functional domain of prime importance for the interaction of RHAU with RNA (268). To assess the significance of this region for the G4 resolvase activity of RHAU, we constructed a series of FLAG-tagged N-terminal truncated mutants of RHAU (Figure 18A). These truncated forms were expressed in HEK293T cells and purified from soluble lysates by anti-FLAG immunoaffinity chromatography. The purity of these preparations as judged by Coomassie staining after SDS-PAGE was very similar (Figure 18B). To test whether the truncated form of RHAU lacking the first 105 amino acids [RHAU(ΔGly–RSM)] could resolve G4 structures, RHAU protein was incubated with ^{32}P-labelled tetramolecular rAGA and ATP. The products were analysed by native PAGE after disruption of RNA-protein interactions by addition of SDS. The free single-stranded ^{32}P-labelled oligos migrated faster than the quadruplex substrate. As shown previously, wild-type (WT) RHAU efficiently resolved G4 structures into single-stranded oligos (Figure 19A). The labelled product of the resolvase reaction co-migrated during electrophoresis with the single-stranded species released by thermal denaturation of the substrate and was proportional to the level of input protein. In contrast to RHAU(WT), RHAU(ΔGly–RSM) failed to resolve the G4 substrate, suggesting that the N-terminal region of RHAU is essential for its G4 resolvase activity.

Being critical for the enzymatic activity and containing an atypical RNA-binding domain (268), the N-terminal region of RHAU is most likely to play a role in substrate recognition. Therefore, we examined whether this N-terminal region was also essential for the recognition of G4 structures. To address this question, we performed RNA electromobility shift assays (REMSA) using G4 substrate as the ligand. RHAU protein was incubated with ^{32}P-labelled G4 in the absence of ATP (to prevent G4 resolution) and the mixtures were analysed by native PAGE. In the absence of protein, the ^{32}P-labelled G4 structures migrated as a single species in the gel (Figure 19B). Addition of increasing amounts of RHAU(WT) protein resulted in the appearance of a protein–G4 complex with reduced mobility. The bound complexes appeared as two main band regions in the gel, which might reflect multiple forms of the complexes. By contrast, the RHAU(ΔGly–RSM) mutant protein failed to form a stable complex with

G4 structures. Taken together, these results demonstrate that the N-terminal region encompassing residues 1–105 of RHAU protein is indispensable for both binding and the resolving of G4 structures by RHAU. These results are in agreement with our previous report that the N-terminal region of RHAU is critical for its RNA binding *in vivo* and its relocalisation to stress granules (268).

The N-terminal region of RHAU binds but cannot alone resolve G4 structures

In view of their uniqueness, the N-terminal regions of DEAH-box proteins may have regulatory functions such as subcellular localisation and/or interaction with RNAs or other proteins (6,37). With *prima facie* evidence that the N-terminal region of RHAU is important for both binding to and resolving G4 structures, we examined whether this region has G4 binding activity. As shown by REMSA, increasing amounts of RHAU(1–105) protein formed proportionately slow-migrating complexes with G4-RNA (Figure 20A). That the observed mobility shift was due to binding of RHAU(1–105) protein was confirmed by addition of anti-FLAG antibody to binding reactions. As shown in Figure 20B, anti-FLAG antibody caused a supershift of the protein-probe complex but not of the free probe.

Given that some G4 binding proteins destabilise G4 structures in a non-catalytic fashion without requiring ATP hydrolysis (262), we examined whether the N-terminal region of RHAU may also destabilise G4 structures in a similar manner. However, as shown in Figure 20C, binding of the N-terminal region to G4

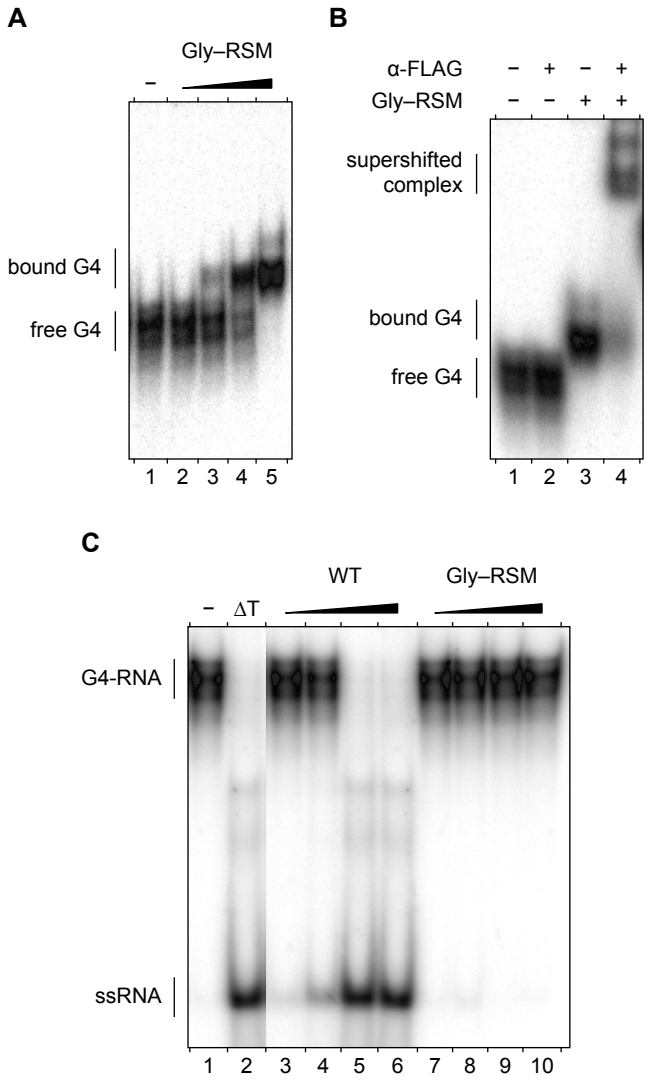


Figure 20 | **G4-RNA binding and resolving activities of the N-terminal domain of RHAU**. (**A**) Gel mobility shift assay for G4-RNA binding: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (30, 100, 300 and 1000 nM) of RHAU(Gly–RSM) in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**B**) Supershift of G4-RNA binding complex by a specific antibody: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated with RHAU(Gly–RSM) at 300 nM in the presence or absence of anti-FLAG antibodies (2.5 µg per lane), as indicated. The reaction mixtures were electrophoresed on a native gel. An autoradiogram of the gel is shown. The positions of the free tetramolecular RNA substrate, the protein-RNA complex and the supershifted complex are indicated on the left. (**C**) G4-RNA unwinding assay: radiolabelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (1, 3, 10 and 30 nM) of WT RHAU or Gly–RSM RHAU mutant (30, 100, 300 and 1000 nM). The reaction products were resolved by native PAGE after disrupting of RNA-protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the tetraplex substrate that was heat-denatured (95 °C, 5 min) and then quenched (ΔT) serves as a marker for the position of single-stranded RNA.

structures is not sufficient to destabilise it. In agreement with the observation that RHAU cannot resolve G4 structures in the absence of ATP (33), this result demonstrates that the N-terminal region alone lacks G4 destabilising activity and that the G4 resolving activity is likely a property of the central catalytically active core domain.

The helicase core domain, together with the N-terminal region, contributes to tight G4 binding of RHAU

Considering that RHAU has a high affinity for G4 structures (89), we next examined whether the N-terminal region itself accounts for the high affinity for G4 structures of the whole protein. N-terminal region [RHAU(1–200)] and full-length RHAU (WT) were expressed as GST fusion proteins in bacteria and *Sf9* insect cells, respectively, and purified to homogeneity as shown in Supplementary Figure 8. Their apparent dissociation constants (K_d) for G4 structures were determined by REMSA. Titration of RHAU(WT) and RHAU(1–200) proteins gave half-saturation points for G4 binding of 14 nM and 440 nM, respectively (Figure 21A and B). This 30-fold difference between the two proteins indicates that the N-terminal region does not constitute by itself an independent and high-affinity G4-RNA binding domain. We propose that the helicase core domain provides RHAU with substantial additional binding activity through interactions with the RNA phosphate backbone, as already shown for many DExD/H-box proteins (27,118,138,139,141,311,312). It should be mentioned that the observed K_d value obtained here is higher than that previously reported (89), which may reflect the difference of the type and location of tags attached to RHAU (N-GST versus C-His_6) and of the way how proteins were purified.

The RSM domain, but not the Gly-rich sequence, in the N-terminal region is crucial for the recognition and resolution of G4 structures by RHAU

The functional N-terminal region encompassing residues 1–105 of RHAU protein consists of two abutting domains (Figure 18A): a 41 amino acid-long low-complexity Gly-rich domain followed by the evolutionary conserved 13 amino acid-long RSM (aa. 54–66). We demonstrated previously *in vivo* that deletion of

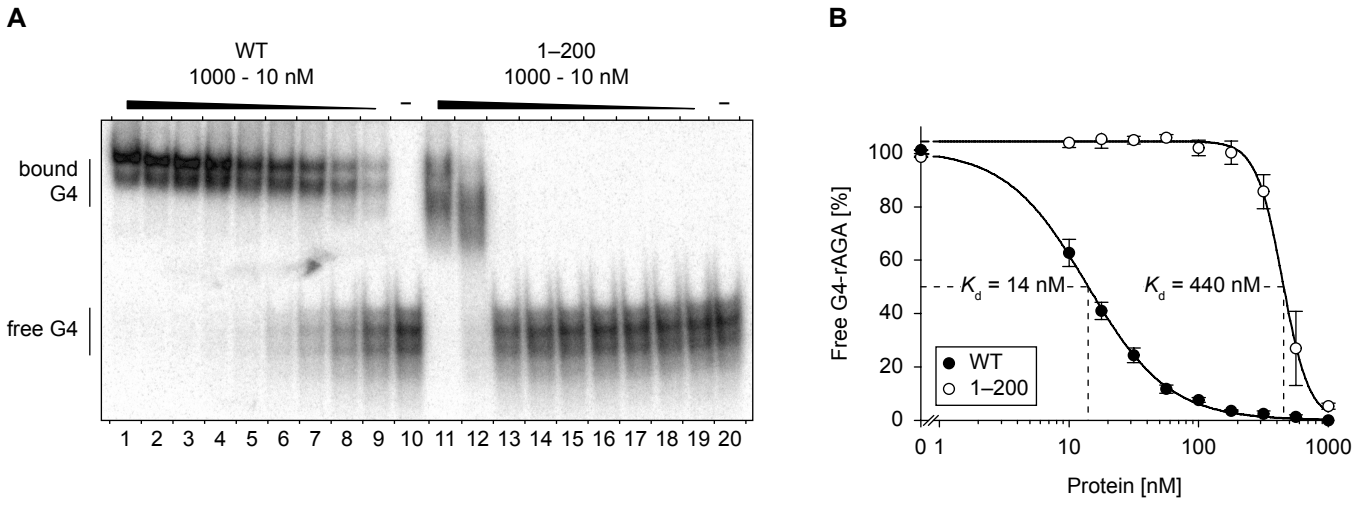


Figure 21 | **G4-RNA binding properties of RHAU and the N-terminal region**. (**A**) Gel mobility shift assay for G4-RNA binding: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts of either GST-tagged WT RHAU or RHAU(1–200) in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. At concentrations from 10 to 1000 nM, GST protein alone had no effect on G4-RNA mobility (not shown). (**B**) Quantification of gel mobility shift assays for wild-type RHAU (WT, ●) and RHAU N-terminal region (1–200, ○) binding to tetramolecular rAGA. The data represent means ± SEM from three independent experiments.

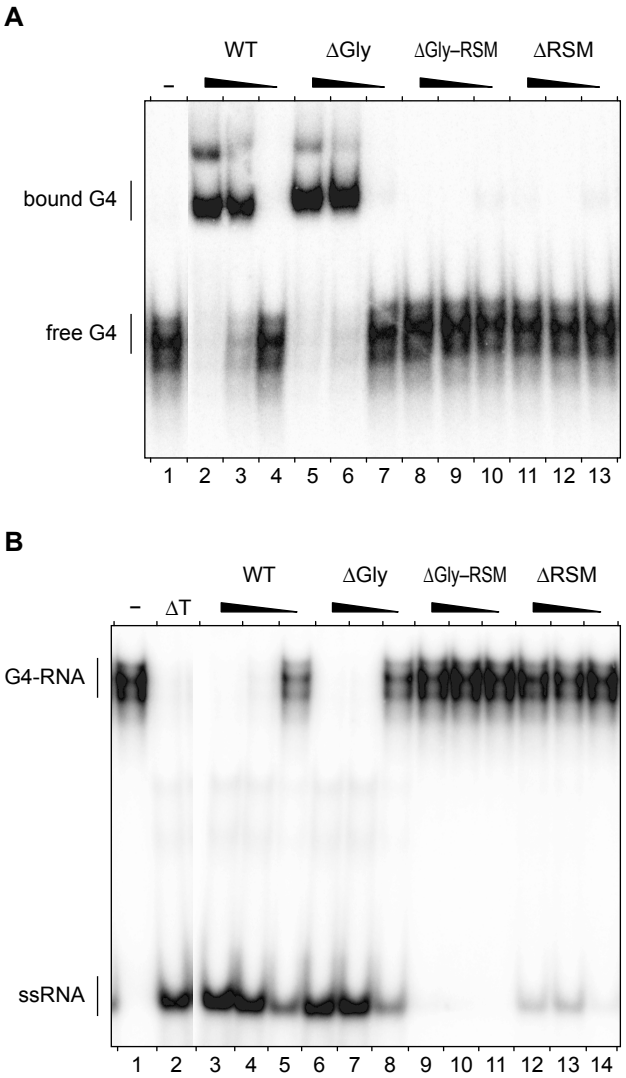


Figure 22 | **G4-RNA binding and unwinding by Gly-rich and RSM truncated RHAU proteins**. (**A**) Gel mobility shift assay for G4-RNA binding: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU, and ΔGly, ΔGly–RSM or ΔRSM RHAU mutants in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**B**) G4-RNA unwinding assay: radiolabelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU, or ΔGly, ΔGly–RSM or ΔRSM RHAU mutants. The reaction products were resolved by native PAGE after disrupting of RNA-protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the tetraplex substrate that was heat-denatured (95 °C, 5 min) and then quenched (ΔT) serves as a marker for the position of single-stranded RNA.

the Gly-rich region [RHAU(ΔGly)] significantly impinged on the relocalisation of the protein to stress granules and on its association with RNA (268). To characterise the individual contributions of each of the two domains to the interaction of RHAU with G4 structures, we performed REMSA experiments using corresponding deletion mutants (Figure 22A). As already shown in Figure 19B, RHAU(ΔGly–RSM) lacking the first 105 amino acids failed to form a stable complex with G4 structures. In contrast, RHAU(ΔGly) still harbouring the RSM was able to bind to G4 structures with the same efficiency as RHAU(WT). Contrary to the deletion of the Gly-rich sequence, a RHAU mutant with an RSM deletion [RHAU(ΔRSM)] abrogated the interaction with G4 structures. We further assessed the consequence of loss of G4-recognition on the *in vitro* resolvase activity of RHAU (Figure 22B). As expected, RHAU(ΔRSM), like RHAU(ΔGly–RSM), failed to resolve G4 structures, while RHAU(ΔGly), like RHAU(WT), did. Similar results with strict RSM-dependency could also be reproduced with other G4-RNA structures (data not shown). Together, these results indicate that the presence of the RSM but no further sequence of the N-terminal region is a prerequisite for the recognition and the resolution of G4 structures by RHAU *in vitro*.

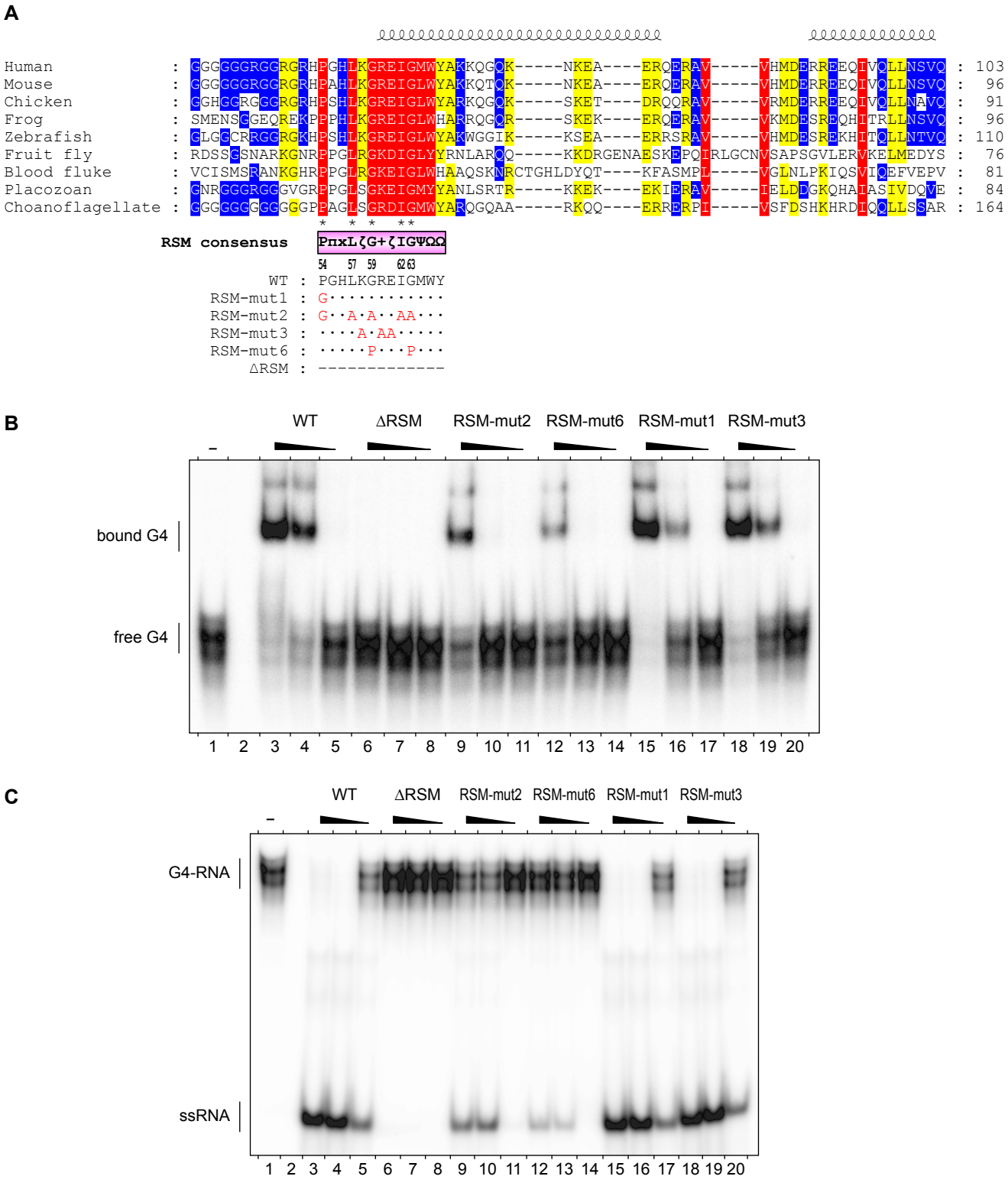


Figure 23 | **G4-RNA binding and unwinding by RSM-mutated forms of RHAU**. (**A**) Conservation of the RSM among RHAU orthologues throughout evolution. Multiple sequence alignment was carried out with MAFFT (version 6, ref. 313). Similarity analysis was made by GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Similarity is shown in red for 100 %, yellow for 99–80 % and blue for 79–60 %. Amino acids that are identical between the nine sequences are indicated by asterisks below the alignment. Secondary structure prediction was performed by JPRED (314) and is indicated on the top. The RSM and its consensus sequence derived from 40 different RHAU orthologues (Supplementary Figure 9) are shown below the alignment (π = small side chain, ζ = hydrophilic, [+] = basic, Ψ = aliphatic, Ω = aromatic). The site-directed substitutions of the RSM mutants employed in this study are listed with the amino acid changes indicated underneath. Species and accession numbers of RHAU orthologues listed are: human (*Homo sapiens*, NP_065916), mouse (*Mus musculus*, NP_082412), chicken (*Gallus gallus*, XP_422834), frog (*Xenopus tropicalis*, ENSXETP00000016958), zebrafish (*Danio rerio*, NP_001122016), fruit fly (*Drosophila melanogaster*, NP_610056), blood fluke (*Schistosoma mansoni*, XP_002577014), placozoan (*Trichoplax adhaerens*, XP_002110272), choanoflagellate (*Monosiga brevicollis*, XP_001747335). (**B**) Gel mobility shift assay for G4-RNA binding: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU or the indicated RSM

Conserved residues within the RSM domain are essential for the recognition of G4 structures by RHAU

Multiple sequence alignments of RHAU orthologues of various species from choanoflagelates to humans unveiled a highly conserved cluster of 13 amino acids presenting the RSM domain embedded in a moderately conserved region of about 50 amino acids (aa. 54–100; Figure 23A). In contrast, the rest of the N-terminus surrounding this region is poorly conserved. The RSM consensus sequence as determined from 40 different RHAU orthologue sequences is: P–π–x–L–ζ–G–[+]–ζ–I–G–Ψ–Ω–Ω† (Supplementary Figure 9). The motif consists of five invariant amino acids (Pro-54, Leu-57, Gly-59, Ile-62 and Gly-63) and seven highly conserved residues of similar biochemical properties (π-55, ζ-58, [+]-60, ζ-61, Ψ-64, Ω-65, Ω-66). In addition, computation-based secondary structure prediction of the sequences used for alignment suggested that RSM is partially structured, sitting astride an unstructured loop (aa. 54–59) and an α-helix (aa. 60–66; Figure 23A). We further examined by site-specific mutagenesis *in vitro* the contribution of conserved amino acids within the RSM domain to the recognition and resolution of G4 structures (Figure 23A). To this end, conserved residues of the RSM were substituted with alanines, prolines or glycines. We focussed on Pro-54, Gly-59 and Gly-63 because they may provide the RSM with substantial structural rigidity (Pro-54) or flexibility (Gly-59 and Gly-63). As shown in Figure 23B, mutation of the five invariant residues of the RSM (RSM-mut2) considerably reduced the binding affinity of RHAU for G4-RNA, albeit to a lesser extent than the RSM-deleted form of RHAU (ΔRSM). Interestingly, substitution of the invariant Gly-59 and Gly-63 residues with prolines (RSM-mut6) caused a stronger reduction of G4 binding activity than the RHAU(RSM-mut2) in which these two amino acids are mutated to alanines. Since prolines may induce a structural constraint on the RSM, this result suggests that recognition of G4 structures may depend on the conformational organisation of the RSM. Finally, as observed with RHAU(RSM-mut1) and RHAU(RSM-mut3) mutants, Pro-54 as well as the polar residues Lys-58, Arg-60 and Glu-61 appear to be dispensable.

As expected, RHAU proteins with mutated RSM and displaying strong G4-RNA binding deficiency (RSM-mut2 and 6) also had reduced G4 resolving activity (Figure 23C). In contrast, despite the slight reduction of G4 binding activity observed for RHAU(RSM-mut1) and RHAU(RSM-mut3), these two mutants unwound G4 structures with an efficiency comparable to wild-type RHAU. We concluded from these *in vitro* mutagenesis experiments that the highly conserved residues in the RSM domain are essential for RHAU G4 structure binding and resolving activities.

ATPase activity of RHAU N-terminal truncated mutants

Hitherto, it could not be excluded that the introduced mutations caused significant conformational changes in the helicase core domain that resulted in enzymatically defective proteins. To define whether loss of G4 resolvase activity was due to loss of G4 binding by the N-terminal domain or to impairment of the basic activity of the helicase core domain, we compared the ATPase activities

† Consensus RSM is listed according to the Seefeld convention (315), symbol nomenclature stands for: π (pi) = small side chain, ζ (zeta) = hydrophilic, [+] = basic, Ψ (Psi) = aliphatic, Ω (Omega) = aromatic.

mutants in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**C**) G4-RNA unwinding assay: radiolabelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU or the indicated RSM mutants. The reaction products were resolved by native PAGE after disrupting RNA-protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the duplex substrate that was heat-denatured (95 °C, 5 min) and then quenched (ΔT) serves as a marker for the position of single-stranded RNA.

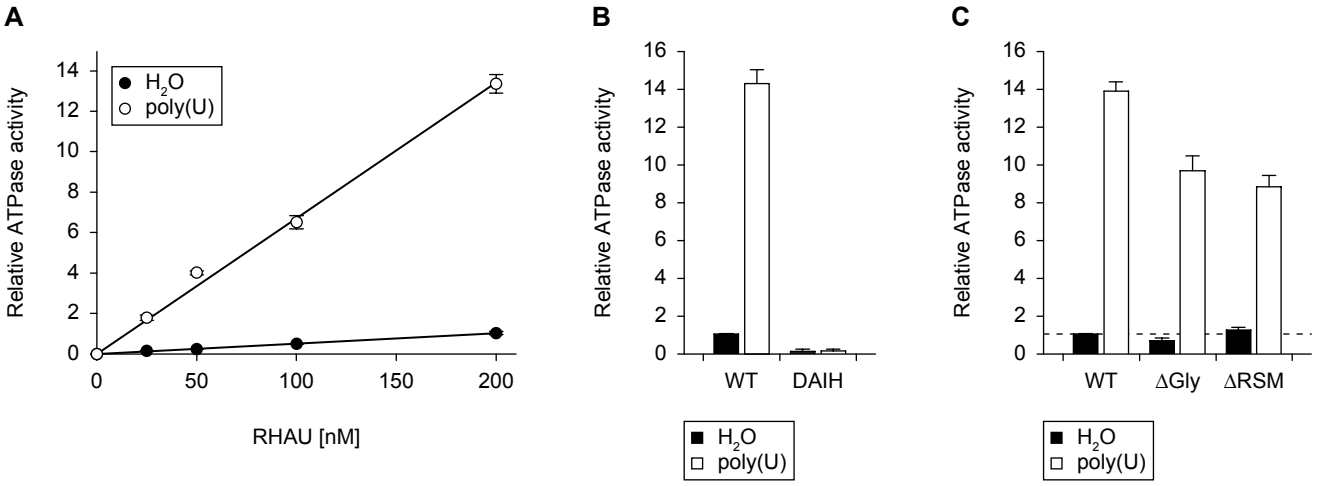


Figure 24 | **ATPase activity of wild-type RHAU and N-terminal mutants**. (**A**) The extent of ATP hydrolysis by wild-type RHAU in the presence (○) or absence (●) of poly(U) plotted as a function of input RHAU. The ATPase activity is expressed as a function of the activity obtained with WT RHAU at a concentration of 200 nM when poly(U) was omitted. (**B**) Influence of poly(U) on the ATPase activity of WT RHAU and DAIH ATPase-deficient mutant (200 nM). The ATPase activity was represented relatively to that of WT RHAU in the absence of poly(U), that was set to 1. (**C**) Influence of poly(U) on the ATPase activity of WT RHAU and ΔGly or ΔRSM RHAU mutants (200 nM). The basal ATPase activity (without nucleic acid cofactor) of WT RHAU is indicated (dashed line). The data represent the means ± SEM from three independent experiments

of RHAU mutants and wild-type RHAU. As shown previously (32), RHAU exhibited significant ATPase activity in the absence of nucleic acids, which was stimulated substantially by the presence of homopolymeric poly(U) (Figure 24A). Both the RNA-dependent and -independent ATPase activities were proportional to the amount of input RHAU protein. To exclude the possibility that the RNA-independent ATPase activity observed was due to contaminants derived from HEK293T cells, we substituted the Glu-335 residue with Ala within the Walker B site (DEIH → DAIH), which abolishes RHAU ATPase activity (32). RHAU(DAIH) protein was prepared according to the protocol employed for wild-type RHAU protein, with comparable yield and purity (Supplementary Figure 11). In contrast to wild-type RHAU, RHAU(DAIH) showed no ATPase activity, even in the presence of poly(U) (Figure 24B). Thus, we concluded that wild-type RHAU harbours intrinsic RNA-independent ATPase activity, which can be further stimulated by homopolymeric RNA.

To clarify the biochemical basis of the absence of G4 resolvase activity in the RHAU(ΔRSM) mutant, we measured the rate of ATP hydrolysis by this mutant and compared it to that of the wild-type RHAU and the RHAU(ΔGly) mutant that was still proficient in G4-unwinding. In the absence of RNA cofactor, the ATPase activity of RHAU(ΔRSM) protein was the same as RHAU(WT) and RHAU(ΔGly) proteins (Figure 24C), suggesting that the deletion of the Gly-rich (ΔGly) or the RSM (ΔRSM) domains does not affect the basal ATPase activity of RHAU. However, the extent of poly(U)-dependent stimulation of ATPase activity of RHAU(ΔRSM) was about 60 % of that of RHAU(WT) and similar to that of RHAU(ΔGly). Taken together, these results indicate that deletion of the N-terminal region of RHAU does not cause any significant conformational changes to the helicase core but renders the RHAU protein less responsive to RNA in the stimulation of its ATPase activity. Given that RHAU(ΔRSM) retained significant RNA-dependent ATPase activity, equivalent to the G4 resolvase proficient RHAU(ΔGly) mutant, we concluded that the lack of G4 resolving activity in the RHAU(ΔRSM) mutant resulted from the loss of G4 binding by the N-terminal domain.

CG9323, the *Drosophila* orthologue of RHAU efficiently unwinds G4-RNA

Based on sequence analysis, RHAU has clear orthologues in almost all species of the animal kingdom ranging from choanoflagellates to humans. A multiple sequence alignment between eight of these orthologues showed the helicase core (aa. 183–625) together with the C-terminal region (aa. 626–1008) of RHAU to be evolutionary conserved (Supplementary Figure 5). In contrast, sequences

of the N-terminal region (aa. 1–182) show little similarity with the exception of the highly conserved RSM (aa. 54–66) and its surrounding region (aa. 54–100). Considering the apparent importance of the RSM in the recognition of G4 structures and the high conservation of its sequence together with the rest of RHAU protein, we surmised that the G4 binding and resolving activity of the RHAU protein are conserved among higher eukaryotes. To validate this hypothesis, we cloned and characterised CG9323, the *Drosophila* orthologue of RHAU.

CG9323 is a 942 amino acid-long protein, which like RHAU comprises all the typical signature motifs of the DEAH-box family of RNA helicases (Figure 25).

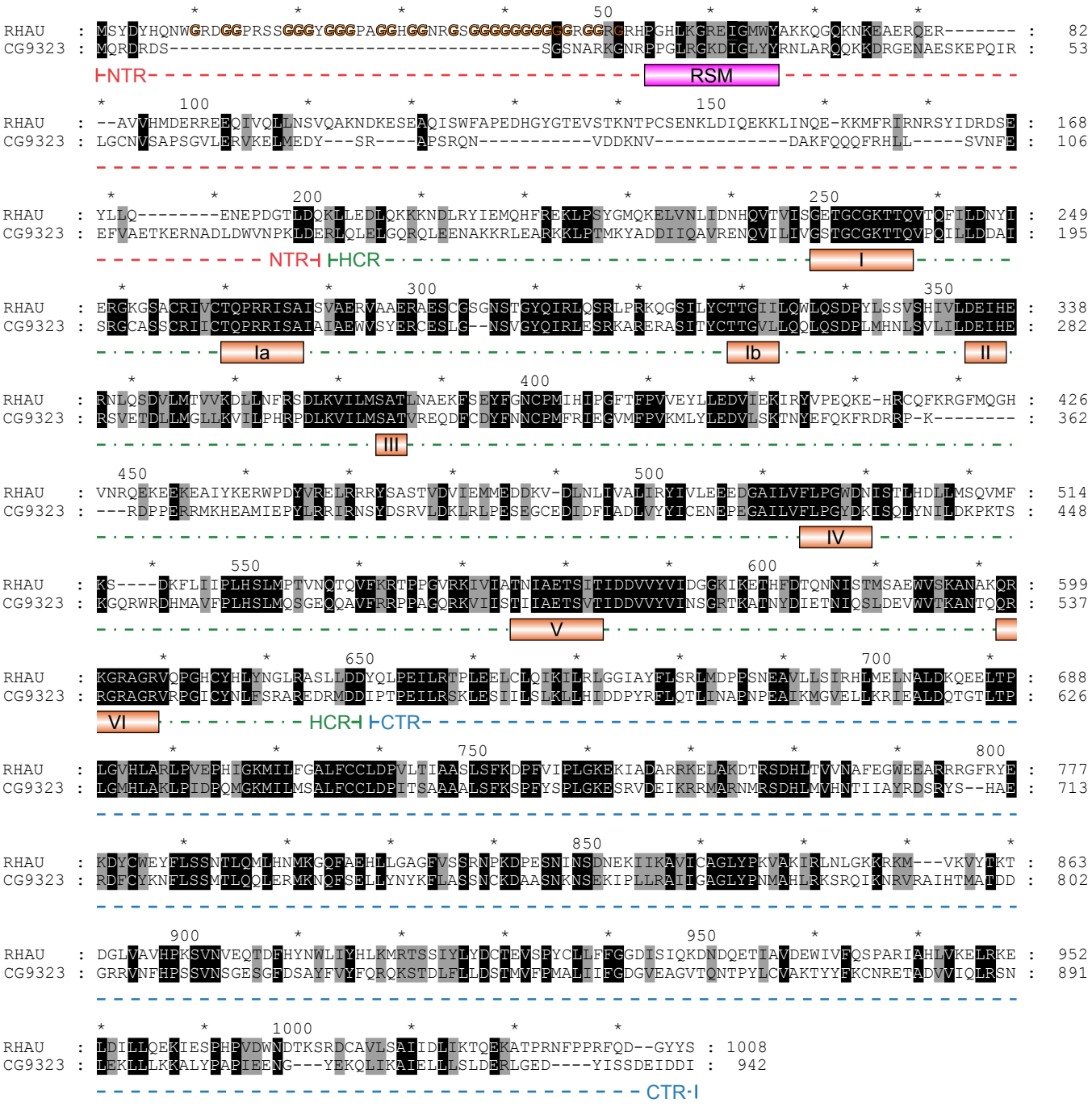
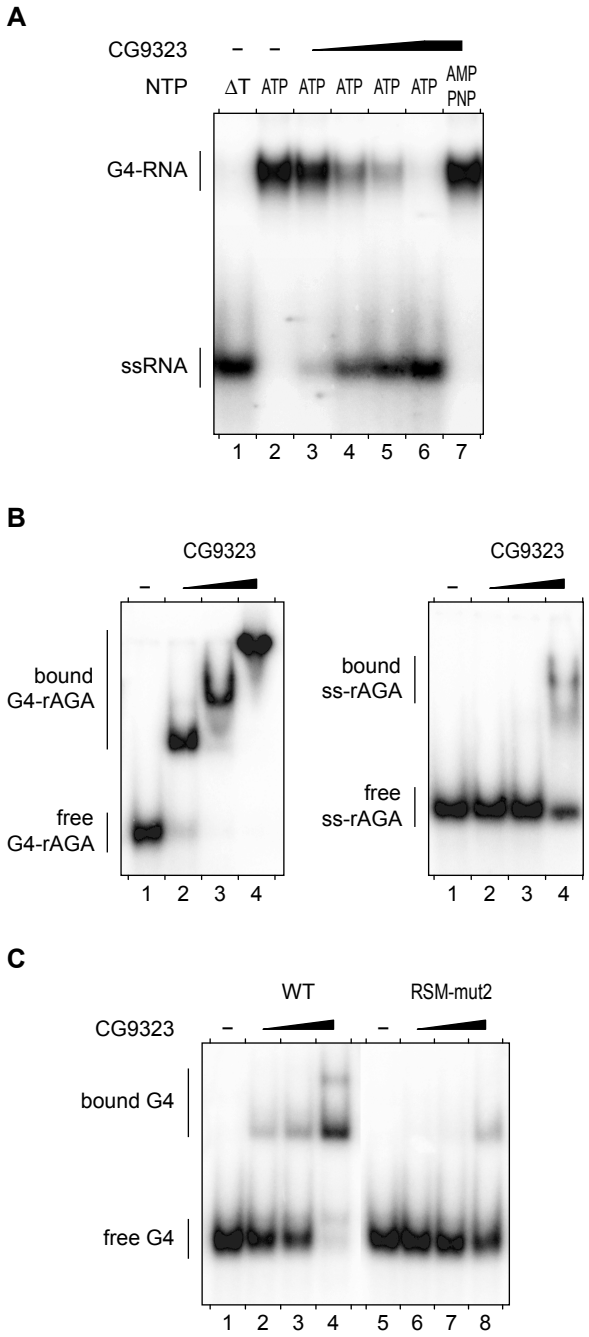


Figure 25 | **Pairwise alignment of RHAU with its *Drosophila* orthologue CG9323**. Amino acids that are identical or similar between the two sequences are shaded in black and grey, respectively. The RSM domain as well as helicase motifs I–VI are indicated below the sequences. Gly residues of the Gly-rich domain (aa. 10–51) of RHAU are depicted in orange. N-terminal (NTR, aa. 1–182), helicase core (HCR, aa. 183–625) and C-terminal (CTR, aa. 626–1008) regions are delineated with a coloured dashed line in red, green and blue, respectively.



The overall identity and similarity between CG9323 and RHAU are 34 % and 52 %, respectively, but not evenly distributed along the entire sequence. Sequence similarity with RHAU is particularly high (41 % identity and 60 % similarity) for the helicase core and the C-terminal regions of CG9323. Nevertheless, apart from the conserved RSM, the CG9323 N-terminal region lacks evident sequence resemblance to RHAU. In addition, the CG9323 N-terminal region is 25 % shorter than that of RHAU, mainly due to the absence of the Gly-rich sequence upstream of the RSM.

To check whether the *Drosophila* orthologue of RHAU retains G4 binding and resolving activities, CG9323 was expressed as a FLAG-tagged recombinant protein and immunopurified to homogeneity as shown in Supplementary Figure 10. The ability of purified recombinant CG9323 protein to unwind and to bind G4 structures was assessed under the conditions previously employed for RHAU. As shown in Figure 26A, CG9323 efficiently unwound the G4-RNA substrate. As for RHAU, the extent of products resolved by CG9323 was proportional to the input protein. Furthermore, CG9323 failed to resolve G4 substrates in the presence of the non-hydrolysable ATP analogue AMP-PNP, indicating that G4-unwinding by CG9323 requires the hydrolysis of nucleosides triphosphate.

Similar to RHAU, CG9323 also formed a stable complex with G4 substrates in the absence of hydrolysable rNTPs (Figure 26B). In binding experiments, CG9323 turned out to be more specific for G4-RNA relative to ssRNAs, since CG9323 bound tetramolecular rAGA with a 10-fold higher affinity than monomeric single-stranded rAGA oligoribonucleotides of the same sequence. This observation, by analogy, is consistent with earlier observations that RHAU has poor sequence-specific recognition of ssRNAs (32,89). Finally, a variant form of CG9323, in which highly conserved residues of the RSM were mutagenised, displayed reduced G4 binding activity, suggesting that the RSM is essential for the recognition of G4 structures by CG9323 (Figure 26C). In conclusion, these results provide compelling evidence that the G4 binding and resolving activities of RHAU have been conserved from *Drosophila* to human, with the RSM playing a pivotal role.

DISCUSSION

Owing to their bulky and thermodynamically stable features, G4 structures have been shown in many respects to impede normal nucleic acid metabolism (190,200,208,209,228,229,316,317). To cope with this problem, proteins are produced that mitigate

Figure 26 | **G4-RNA unwinding and binding by CG9323**. (**A**) G4-RNA unwinding assay: radiolabelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP or AMP-PNP (as indicated) without protein (−) or with increasing amounts (2, 6, 20 and 60 nM) of purified recombinant CG9323 protein. The reaction products were resolved by native PAGE after disrupting RNA-protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the duplex substrate that was heat-denatured (95 °C, 5 min) and then quenched (ΔT) serves as a marker for the position of single-stranded RNA. (**B**) G4- and single-stranded-RNA binding assay: tetramolecular (left) or monomeric (right) radiolabelled rAGA at a concentration of 100 pM were incubated without protein (−) or with increasing amounts (10, 30 and 100 nM) of CG9323 in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**C**) G4-RNA binding assay: radiolabelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (3, 10 and 30 nM) of WT or RSM-mut2 CG9323 proteins in the absence of ATP. The reaction mixture were analysed by native PAGE. An autoradiogram of the gel is shown.

effects of these atypical stable structures. Eight human helicases, including RHAU, have been shown so far to harbour G4 resolving activity *in vitro*. These include the SF1 helicases Pif1 (233) and Dna2 (246), as well as the RecQ family proteins BLM (227,243) and WRN (226), the Rad3-like helicases FANCJ (228) and DDX11 (230) as well as the DEAH-box RNA helicase A (DHX9, ref. 238). Apart from DHX9 and RHAU, these proteins are all DNA helicases that have been clearly implicated in the maintenance of genome integrity (22,23,248,249,318). Both RHAU and DHX9 belong to the DEAH-box family of RNA helicases and show very little sequence similarity with the above mentioned helicases. Curiously, RHAU was initially identified as a G4-DNA resolvase enzyme (33). The possibility that G4-RNA structures are targets of RHAU emerged from *in vivo* UV-crosslinking results showing RHAU binding mainly to RNA (268). Subsequent characterisation of RHAU demonstrated its aptitude to resolve G4-RNA better than G4-DNA (89). This finding was a remarkable breakthrough, since RHAU was the first helicase to possess G4-RNA resolvase activity. Furthermore, RHAU is one of the rare DExD/H-box proteins that exhibit high affinity and specificity for its substrate *in vitro* independent of accessory proteins. Since this finding, efforts have been made to understand the mechanism underlying recognition of G4 structures by RHAU. The present study shows that the N-terminal region of RHAU is essential and responsible for binding of RHAU to G4 structures. Further investigations dissecting the N-terminal region, coupled with site-directed mutagenesis, have demonstrated that the RSM makes a decisive contribution to the high affinity of RHAU for G4 structures. Sequence comparisons of RHAU orthologues from various species showed the RSM to be the unique highly conserved part of the N-terminal region. Hence, we predicted that all orthologous forms should possess G4 resolving activity based on the functional significance and sequence conservation of the RSM domain and the catalytic region. This hypothesis was supported by the robust ATP-dependent G4-RNA resolvase activity found for CG9323, the *Drosophila* form of RHAU. As expected, and also shown for RHAU, the binding of CG9323 to G4-RNAs depended on RSM integrity, which further indicated similar functions for this motif in both proteins.

Recognition of G4-RNA by RHAU depends on the N-terminal RSM

RHAU shares with most helicases a global scheme of modular architecture that combines a conserved central helicase core domain with peripheral regions of various lengths and sequences (319). Together with previous findings (268,320-322), our results indicate that the helicase core alone cannot account for the high specificity of function usually attributed to DExD/H-box proteins. In this regard, these data also agree with numerous structural observations that the helicase core region of DEAD-box proteins interacts essentially in a non-sequence-specific manner with the phosphoribose backbone of single-stranded nucleic acid (27,138,139,311,312). Such contacts suffice to discriminate RNA from DNA by means of the 2'-hydroxyl groups of the ribose moieties, but not to distinguish between sequences of varying nucleotide composition. In the present work, we have shown the importance of the unique N-terminal flanking region in adapting the conserved catalytic core to a specific function. The present investigation has also shown clearly that, although necessary, ATPase activity by itself is not sufficient for RHAU to unwind G4 structures. Our data strongly suggest that the establishment of a stable complex between RHAU and its G4 substrate is a prerequisite for the subsequent ATPase-dependent unwinding of the G4 structure. We propose that the N-terminal RSM endows the enzyme with specificity by binding the G4 substrate, thereby positioning

the helicase core in close proximity to the substrate. Likewise, critical roles for N- and C-terminal flanking regions have been reported for several DExD/H-box proteins, exerting specific functions by interacting with particular RNA species/structures or with other regulatory proteins (320,323-325). For example, the prototypical yeast DEAH-box proteins Prp16 and Prp22 and the human orthologue HRH1 (alias DHX8 or hPrp22) have been shown to associate with the spliceosome *via* their non-conserved N-terminal regions (110,116,117). The *Drosophila* maleless (MLE) protein, more closely related to RHAU, harbours two copies of dsRNA binding motifs (dsRBM) in its N-terminal region (326). Their deletion, as with the N-terminal truncation of RHAU, caused the loss of RNA-binding and unwinding activities *in vitro* and subcellular mislocalisation of the protein *in vivo* (148). Together, these examples emphasise the role of the peripheral domains of DExD/H-box proteins in adapting a common catalytic core to a broad spectrum of specific functions.

Potential role of the RSM in RHAU relocalisation to stress granules

The present work underscores the essential nature of the conserved RSM, which endows RHAU with a high affinity for G4 structures *in vitro*. Determining the mechanisms whereby RHAU resolves G4 substrates *in vivo* is an important issue to be addressed in future. It is tempting to draw a parallel between the present observation and a previous report that the stress-induced recruitment of RHAU to stress granules is mediated by interactions of RHAU with RNA (268). Similar to our observations, the region identified as essential for this activity included the RSM together with the upstream Gly-rich domain, underscoring the functional relevance of the RSM and its surrounding region for the biological activity of RHAU. The N-terminal region of RHAU alone was also found to be sufficient to drive RNA binding *in vivo* as well to relocalise to stress granules (268). This observation, however, contrasts with the data presented here in that, although required for functional specificity, the N-terminal region alone does not constitute an independent and high affinity G4-RNA-binding domain. In this regard, our data rather suggest that both the N-terminal and helicase core regions are required for the productive interaction of RHAU with G4 structures. Since the nucleic acid binding motifs Ia, Ib, IV and V of DEAH-box proteins have been proposed to contact RNA (118,141), we surmise that, for RHAU, the helicase core may also provide the protein with substantial binding activity by interacting in a non-specific manner with the phosphoribose backbone of the single-stranded tail flanking the tetramolecular G4 structure. Further investigations into the functional contribution of the RNA binding site of the helicase core are needed to further challenge this hypothesis. In addition, our previous finding that the ATPase-deficient form of RHAU [RHAU(DAIH)] stalls in stress granules (268) agrees with the observation that, once bound to G4-RNA, RHAU(DAIH) cannot dissociate itself from its substrate, even in the presence of ATP (data not shown). This indicates that the G4-unwinding reaction requires ATP hydrolysis, rather than ATP binding *per se*, and further suggests that the release of the RNA substrate occurs only after G4-unwinding. Stress granules may constitute a favourable environment for the formation of intermolecular G4-RNA structures, as they are temporary sites of accumulation of stress-induced stalled translation initiation RNP complexes (271,272). However, an important question that we have not yet addressed is whether RHAU relocalises to stress granules upon binding to G4-RNA structures. Interestingly, however, we noticed that RSM-mutated forms of RHAU that are deficient for G4 structure recognition *in vitro*, manifest reduced association with RNA *in vivo* concomitantly with reduced relocalisation to stress granules (Chalupnikova, K.,

unpublished data). Thus, this finding raises the possibility that at least a fraction of RHAU is recruited to stress granules *via* interactions with G4 structures.

The mechanism by which RHAU recognises G4 structures is an important issue that needed to be addressed to improve our understanding of RHAU. So far, however, little is known about its biological function as a potential G4 resolvase enzyme. Recently, G4 structures in RNA have attracted considerable attention as a plausible means of regulating gene expression (260,327). Formation of G4 structures in the 5'-untranslated region has been shown to affect mRNA translation (190,208,209) and bioinformatics studies have identified more than 50'000 potential G4 structures near splicing and polyadenylation sites of various human and mouse genes. This raises the possibility that G4-formation impedes RNA metabolism at many different stages (214). At the moment, we lack a corresponding understanding of how cells negotiate with G4 structures in various RNA molecules. However, from the findings presented here, RHAU emerges as a promising regulator of G4 structure-based RNA metabolism and merits future investigation of its potential roles in different biological contexts.

RHAU binds an intramolecular G4 structure in TERC and associates with telomerase holoenzyme

AIMS AND RATIONALE

Although considerable information is available regarding the enzymatic activity of RHAU *in vitro* (33,39,88,89), almost nothing is known about its biological function as a G4 binding/resolving enzyme *in vivo*. To address this question, we sought RNAs bound by RHAU in living cells, surmising that the identification of RHAU-bound RNAs and understanding the effects of RHAU should provide important clues to its function *in vivo*. To this end, we employed high-throughput gene array technologies to identify and quantify the co-purified RNAs on a genome-wide scale. With this screen, we identified about 100 RNAs significantly enriched by RHAU. Computational analysis of RNA sequences for potential intramolecular G4 structures revealed the preferential association of RHAU with transcripts bearing G4-forming motifs, suggesting direct targeting of G4-RNAs by RHAU. Amongst the RNAs with the potential to form G4 structures and selectively enriched with RHAU, we identified TERC as a *bona fide* target of RHAU. Characterisation of the RHAU-TERC interaction *in vivo* and *in vitro* showed binding of TERC by RHAU to be strictly dependent on the formation of a G4 structure on the 5'-extremity of TERC RNA and to require the N-terminal RSM domain of RHAU. Finally, we have demonstrated that RHAU not only interacts with TERC *stricto sensu*, but is also part of the fully assembled telomerase complex through direct interaction with TERC G4 structure. Taken together, these data demonstrate that intramolecular G4-RNAs are naturally occurring substrates of RHAU *in vivo*. Moreover, they provide indirect but strong support for the existence of a G4-RNA structure in the telomerase RNP.

RESULTS

Microarray identification of RHAU-associated RNAs

To identify endogenous RNAs associated with RHAU *in vivo* on a genome-wide scale, we designed a RIP-chip assay (RNA immunoprecipitation coupled to microarray analysis). Subsets of RHAU target RNAs were isolated by immunoprecipitation (IP) assays under optimised conditions that preserved RNA-protein complexes. Briefly, HeLa cells were transfected with a vector expressing FLAG-tagged or myc-tagged RHAU. Immunoprecipitations of non-cross-linked whole-cell extracts were carried out using anti-FLAG antibodies. Anti-FLAG IP from cells expressing myc-tagged RHAU was employed as a control (IP_{ctrl}) to assess nonspecific interactions that may occur during RIP. Following IP, co-fractionated RNAs were recovered and purified by standard phenol-chloroform extraction and converted to cRNA. Products were subsequently hybridised to human oligonucleotide arrays.

Western blot analysis of immunoprecipitated proteins revealed that RHAU-FLAG, but not RHAU-myc, was efficiently enriched from HeLa cell extracts following IP with anti-FLAG antibody (Figure 27A, bottom). Oligo(dT)-primed reverser-transcription of co-immunoprecipitated RNAs showed the presence of polyadenylated RNAs with sizes similar to the input (Figure 27A, top, compare lane 2 to lanes 3 and 4). With regard to the RHAU-FLAG IP fraction, very little background RNA signal was detected in the control immunoprecipitate (lane 1), demonstrating specific co-immunoprecipitation of mRNAs with RHAU. The association between RHAU and target RNAs was deemed specific as the

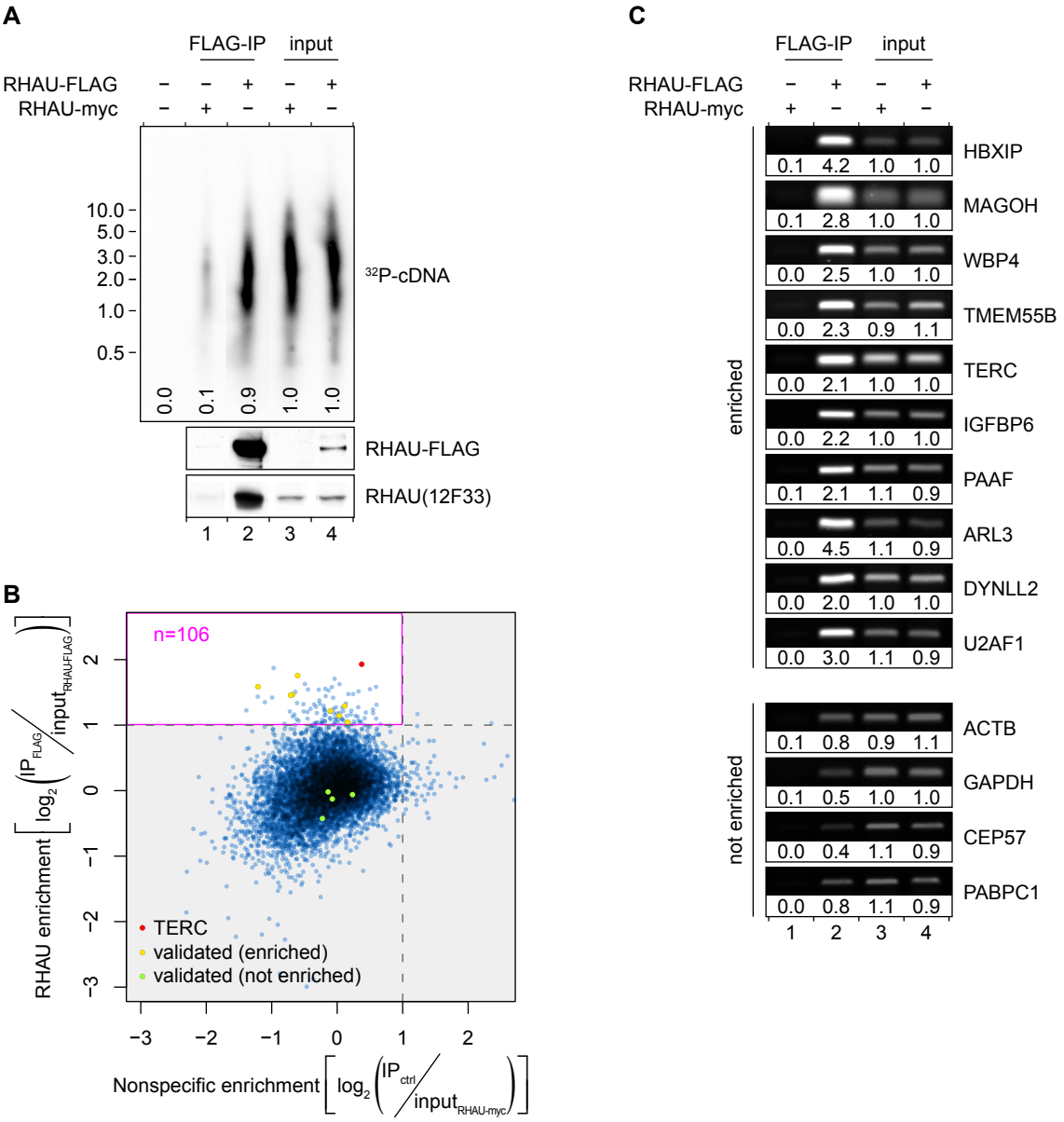


Figure 27 | **Analysis of the RNAs co-immunoprecipitated with RHAU RNPs**. (**A**) RHAU associates with poly(A)$^+$ RNAs in HeLa cells. First strand cDNA synthesis from input (lanes 3 and 4) and co-immunoprecipitated (lanes 1 and 2) RNAs was monitored by incorporation of [α-^{32}P]dATP. The RT-PCR reaction products were separated by agarose gel electrophoresis. An autoradiogram of the gel is shown. The positions and sizes (kb) of marker DNAs are indicated at the left. Signal intensities are expressed relative to the average signal intensity of the input fractions (lanes 3 and 4). Expression and specific enrichment of RHAU-FLAG in input and FLAG-IP fractions were verified by Western blot analysis. (**B**) Scatter plot representation of the differential enrichment of RHAU-bound versus nonspecific RNAs as quantified by microarray analysis. The pink rectangle delineates the area of RNA specifically enriched by RHAU. RNAs whose enrichment was further validated by semi-quantitative RT-PCR are indicated. (**C**) Validation of novel RHAU-bound target RNAs. The abundance of 10 potential RHAU targets and 4 non-targeted RNAs in total input RNA (lanes 3 and 4), control IP (lane 1) and RHAU-FLAG IP (lane 2) fractions was monitored by semi-quantitative RT-PCR. Reaction products were separated by agarose gel electrophoresis and visualised by SYBR Green staining. Band intensities are expressed relative to the average signal intensity of the input fractions (lanes 3 and 4).

anti-FLAG antibody exhibited no obvious cross-reactivity with other cellular proteins (data not shown).

Microarray analysis of the purified RNAs recovered from input and immunoprecipitated fractions revealed that 9'354 (49 %) of the 19'089 total genes available on the chip were significantly expressed in HeLa cells overexpressing RHAU (data not shown). In order to maximise the chance of identifying true RNA targets of RHAU as well as to minimise the occurrence of false positives, only those RNAs were chosen that were significantly (adjusted *P*-value < 0.01) enriched and were at least two-fold more abundant in the RHAU-FLAG IP fraction than in the control. Of these potential RHAU targets, we discarded

those that were also significantly enriched in the RHAU-myc expressing cells and might thus be RHAU-independent RNAs binding non-specifically to the anti-FLAG antibody matrix.

Thus, of the 9'354 genes expressed in HeLa cells, 108 RNAs (1.2 %) were found to be significantly enriched in RHAU-FLAG IP fraction compared with total input RNA (Figure 27B). Finally, after subtraction of two non-target RNAs based on the above-mentioned criteria that were associated with the antibody or the beads, 106 RNAs were judged to be specifically enriched by RHAU (Supplementary Table II). The most abundant transcripts of these potential RHAU targets were selected for further analysis.

Validation of potential RHAU target RNAs

To assess independently the validity and reproducibility of the identified RHAU-associated transcripts, RNA abundance in total input RNA, control IP and RHAU-FLAG IP fractions from fresh whole-cell extracts were analysed by semi-quantitative RT-PCR. As shown in Figure 27C, ten RNAs randomly selected from the group of plausible RHAU targets were confirmed to be enriched more than two-fold in the RHAU-FLAG IP fraction than in the control IP or input fractions. Besides, four non-targeted RNAs were included as negative control to monitor RHAU binding to non-specific RNAs. None of them were found to be enriched in the RHAU-FLAG IP fraction despite their relatively high abundance in the input fraction. Taken together, these independents results validate the previous RIP-chip data. We assume that the remaining RNA species are also part of RHAU RNPs, although additional experiments would be necessary to confirm this supposition.

G4-content analysis for RNAs enriched by RHAU

As RHAU shows a high affinity for G4-RNA structures *in vitro* (89), we next carried out a bioinformatics search for RNAs with potential intramolecular G4 structures. The rationale was that if RHAU binds G4-RNA *in vivo*, the proportion of potential G4-forming sequences should be higher among RNAs enriched by RHAU than among non-enriched RNAs. Of the various available methods for predicting intramolecular G4 motif sequences, we used QGRS Mapper that identifies and scores each potential G4-forming sequence according to their predicted stabilities (293). Being aware that only limited experimental data so far support the scoring method of QGRS Mapper, we employed the algorithm to identify potential G4-forming sequences, but we also included the G4-score-based analysis as Supplementary Material (Supplementary Figure 12), since the two approaches provided similar results. In fact, with almost eight potential intramolecular G4-forming sequences (PQS) per kilobase, the occurrence of G4 motif sequences was higher in the fraction enriched by RHAU than in any other fraction (Figure 28A and Supplementary Figure 12A). In addition, there was a weak ($R^2 = 0.10$) but significant positive correlation between the predicted G4 motif density per transcript and the magnitude of RNA enrichment by RHAU. Randomisation of RHAU target sequences, retaining single- or even di-nucleotide frequencies of the original transcripts, significantly reduced the proportion of potential G4-forming sequences (Figure 28B and Supplementary Figure 12B). Thus, the proportion of G4 motif sequences among RHAU-associated transcripts cannot be explained by a mere bias in nucleotide composition.

It should be mentioned that our prediction is likely to overestimate the number of G4-forming motif sequences per transcript. In fact, we also considered G4 structures consisting of only two successive G-tetrads as these metastable structures have also been demonstrated to have biologically relevant

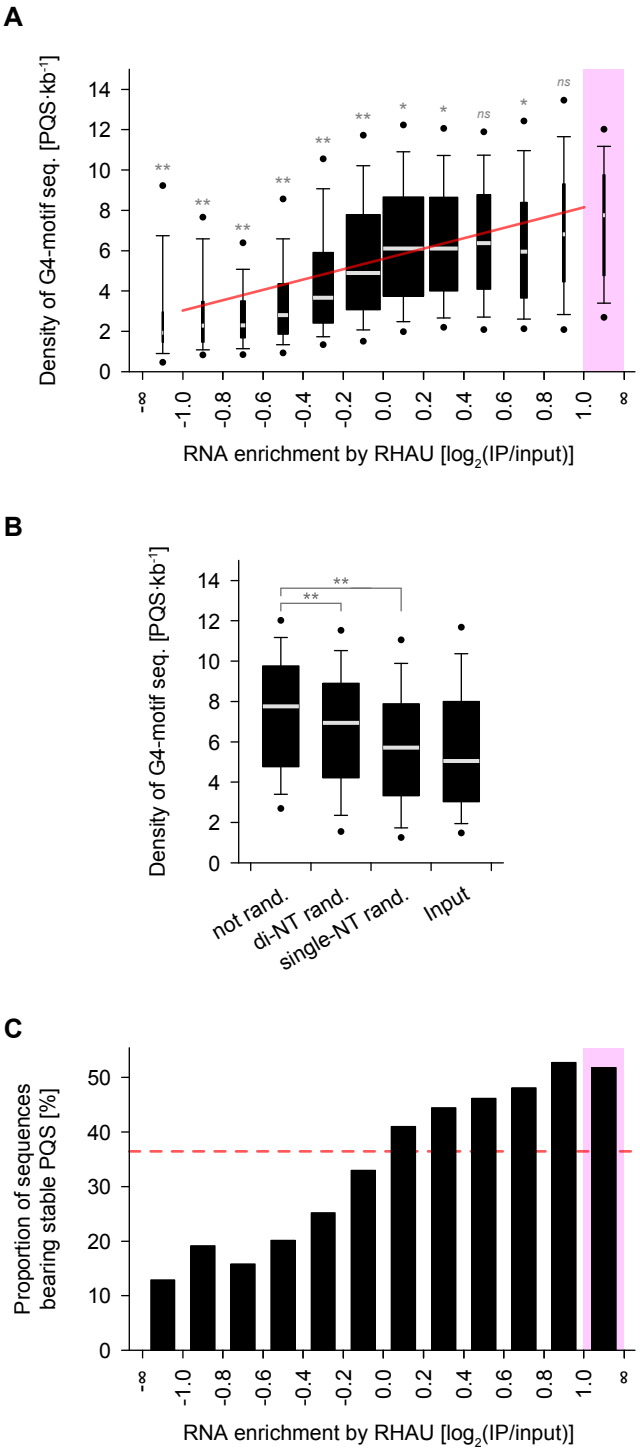


Figure 28 | **Computational analysis of potential intramolecular G4-forming sequences (PQS) among RNAs enriched by RHAU**. (**A**) Box plot representation of the density of potential intramolecular G4-forming sequences per transcript (PQS·kb^{-1}) as a function of their level of enrichment by RHAU. The lower and upper boundaries of the boxes indicate the 25th and 75th percentiles, respectively. The white lines within the boxes mark the median. The width of the boxes is proportional to the number of transcripts found within a group. The error bars indicate the 10th and 90th percentiles and the filled circles the 5th and 95th percentiles. The pink rectangle refers to the group of 106 RNAs specifically enriched by RHAU and serves as a control group for multiple comparison analysis. Statistical significance was determined by Kruskal-Wallis one-way ANOVA on ranks and the Dunn's test. $^{ns}P \geq 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$. The correlation between the two variables was estimated by linear regression analysis ($y = 2.56x + 5.60$, $R^2 = 0.10$, $P < 0.001$) and is indicated as a red line on the graph. (**B**) PQS analysis among randomised RHAU target sequences. The sequences of the 106 transcripts specifically enriched by RHAU were shuffled, retaining single- or di-nucleotide base composition of the original transcripts. The box plot represents the density of PQS (PQS·kb^{-1}) among the original (not rand.), di-nucleotide shuffled (di-NT rand.) and single-nucleotide shuffled (single-NT rand.) sequences. Statistical significance was determined by one-way repeated measures ANOVA and the Dunnett's test. (**C**) Proportions of RNAs bearing stable PQS as a function of their level of enrichment by RHAU. The pink rectangle refers to the group of 106 RNAs specifically enriched by RHAU. The dashed red line denotes the proportion of sequences showing stable PQS in the input fraction. Significant ($P < 0.001$) association between the magnitudes of the two variables was estimated by the chi-square-test-for-trend.

roles (317,328,329). However, after discarding these less-stable G4 structures from the analysis, we still observed a significant relationship between the proportion of sequences bearing potentially stable G4-forming sequences (composed of three or more stacked G-quartets) and the magnitude of RNA enrichment by RHAU (Figure 28C). Indeed, with more than half of the sequences presenting potentially stable G4-forming sequences, the group of 106 RNAs identified as true RHAU targets presents one of the highest incidences of potentially stable G4-forming motifs. In summary, the present data are consistent with preferential association of RHAU with transcripts containing potential G4-forming sequences, suggesting direct recognition of G4-RNA structures by RHAU. To further test this hypothesis, we selected one of the identified RHAU targets and addressed the molecular basis of its interaction with RHAU.

RHAU associates with TERC through its G4 motif sequence

Within the pool of transcripts enriched by RHAU and showing a high potential to form stable G4, TERC is the only RNA reported previously to form a stable and parallel G4 structure *in vitro* (Figure 29A; ref. 224). TERC was also found to be one of the most abundant RNAs enriched by RHAU in several independent RIP-chip assays using various cell lines (Supplementary Table II and unpublished data). To address the role of the TERC G4 structure in its efficient co-immunoprecipitation by RHAU, we cloned the TERC gene, including its promoter and 3' flanking genomic sequence. The 5' extremity of the TERC sequence was subsequently mutated (G4-MT) or truncated (ΔG4) to prevent G4 formation (Figure 29B). To avoid effects on the structurally conserved P1 helix by the introduced substitutions, guanine residues that are part of both the predicted G4 structure and the P1a helix region were left intact. The recombinant forms of TERC were transiently transfected into HEK293T cells and the accumulation of stable TERC transcripts was monitored by RT-qPCR (Figure 29C). After 24 hours, HEK293T cells transiently transfected with the TERC(WT) construct but not with the vector alone showed a substantial (~200-fold) increase in TERC abundance relative to endogenous TERC levels. Importantly, mutations of the G4 motif (G4-MT or ΔG4) did not influence the steady-state level of exogenous TERC expression in these cells since recombinant wild-type, G4-MT and ΔG4 forms of TERC accumulated to comparable levels.

To examine the significance of the G4 motif sequence for co-precipitation of TERC with RHAU, we repeated the RIP assay using HEK293T cells transiently overexpressing the wild-type or G4-MT and ΔG4 mutated forms of TERC. Immunoprecipitations were carried out on endogenous RHAU using a monoclonal antibody against RHAU and TERC RNA abundance was analysed by RT-qPCR (Figure 29D). As already shown for endogenous TERC, a substantial amount of overexpressed TERC(WT) was also recovered with endogenous RHAU, as evidenced by a 25-fold enrichment of TERC transcript over the mock transfection (Figure 29D, compare lanes 4 and 2). In marked contrast, both G4 motif mutated forms of TERC were barely co-immunoprecipitated with RHAU, judged by the 95 % reduction in TERC abundance in the corresponding fractions compared with the WT control (Figure 29D, compare lanes 5 and 6 to lane 4). The drastic reduction in immunoprecipitation of the TERC mutants was not due to nonspecific RNA degradation in these fractions. Indeed, comparable levels of nonspecifically co-immunoprecipitated GAPDH RNA were found to contaminate all of these fractions. Taken together, these results indicate that the intact G4 motif sequence is a prerequisite for TERC association with RHAU *in vivo*.

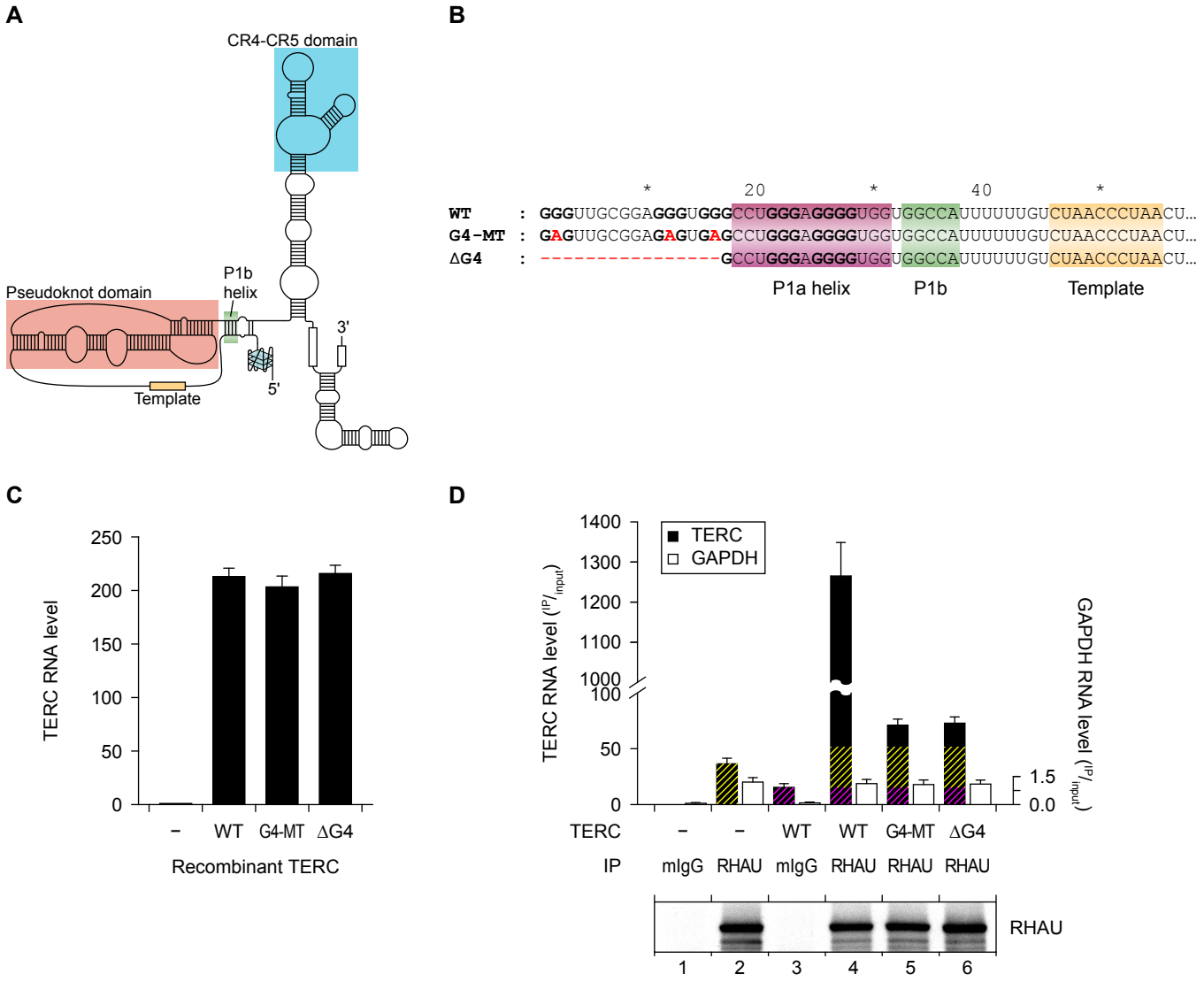


Figure 29 | **RHAU associates with TERC through its G4-motif sequence**. (**A**) Schematic representation of the secondary structure of human TERC bearing a parallel G4 structure in the 5'-extremity as described by Mergny et *al.* (224). (**B**) Nucleotide sequence of the WT and G4 motif mutant (G4-MT and ΔG4) forms of TERC. Guanine tracts that are predicted to form a stable G4 structure are shown bold. Nucleotide substitutions or deletions in mutant forms of TERC are marked in red. The P1 helix subdomains as well as the template region are indicated. (**C**) RNA expression levels of endogenous and recombinant WT and G4 motif mutant forms of TERC in HEK293T cells. Expression was quantified by RT-qPCR, normalised to GAPDH expression and endogenous TERC levels set to 1. Data represent the mean ± SEM of three independent experiments. (**D**) RT-qPCR analysis of the abundance of WT and G4 motif mutant forms of TERC that co-immunoprecipitated with endogenous RHAU protein. RNA levels in IP fractions are represented in function of their respective abundance in the input fraction. Immunoprecipitation with mouse IgGs (mIgG) served as a control to assess nonspecific interactions. In lanes 4, 5 and 6, yellow hatches represent the fraction related to endogenous TERC signal and violet hatches represent the fraction of TERC RNA that non-specifically interacts with the antibody matrix or the beads. Data represent the mean ± SEM of five independent experiments. Comparable efficiency of RHAU immunoprecipitation in the various fractions was verified by Western blot analysis with anti-RHAU antibodies.

RHAU binds TERC through a G4 structure in the TERC 5'-region

In the absence of Mg·NTPs, RHAU specifically binds to tetramolecular G4-RNA structures with high affinity (88,89). The strict requirement for the G4 motif sequence for effective recovery of TERC by RHAU *in vivo* suggested that RHAU may also form a stable complex with TERC through direct binding of the G4-RNA forming structure. To address this question *in vitro*, we performed RNA electromobility shift assays (REMSA) using purified recombinant GST-tagged RHAU protein and *in vitro* transcribed radiolabelled full-length (1–451 nt) and 5'-end (1–71 nt) TERC fragments. In the absence of protein, the ^{32}P-labelled probes migrated as a single species in the gel (Figure 30A and B). Addition of increasing amounts of RHAU to both full-length and 5'-end TERC

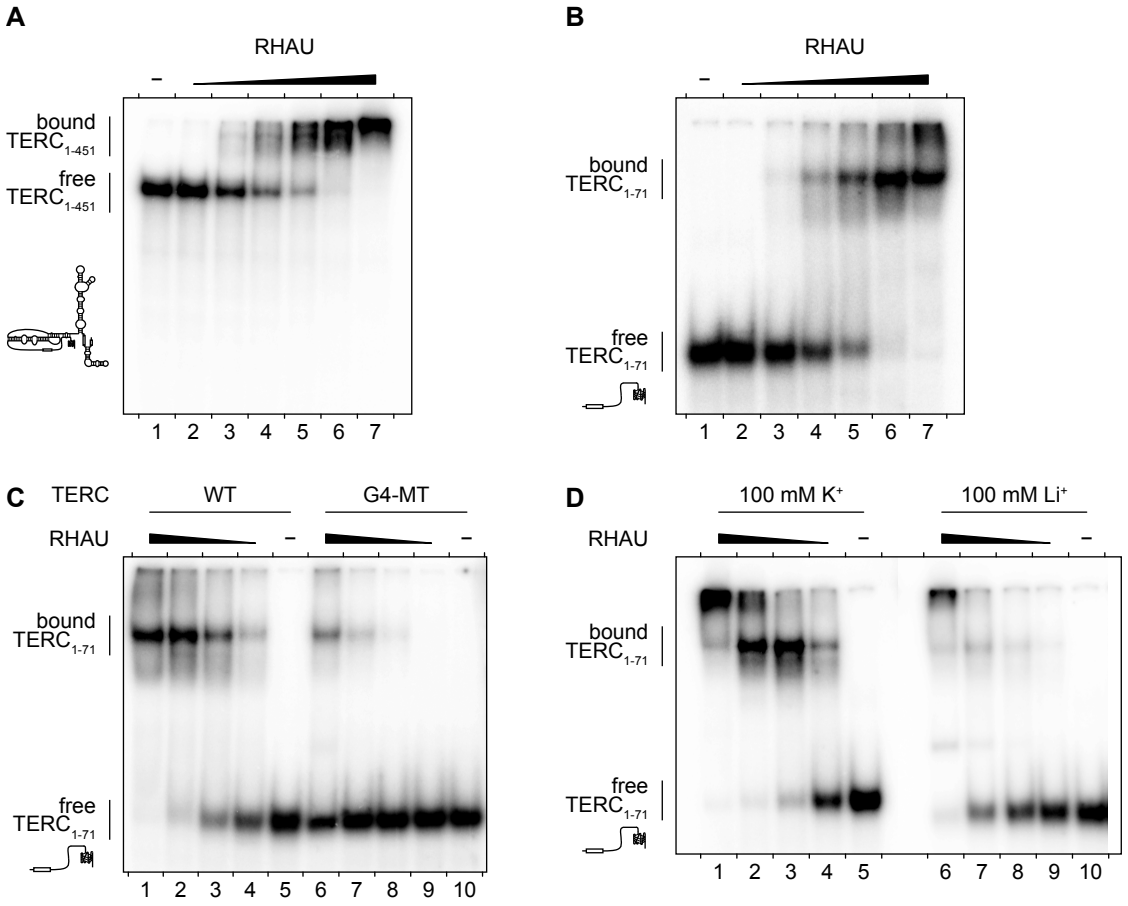
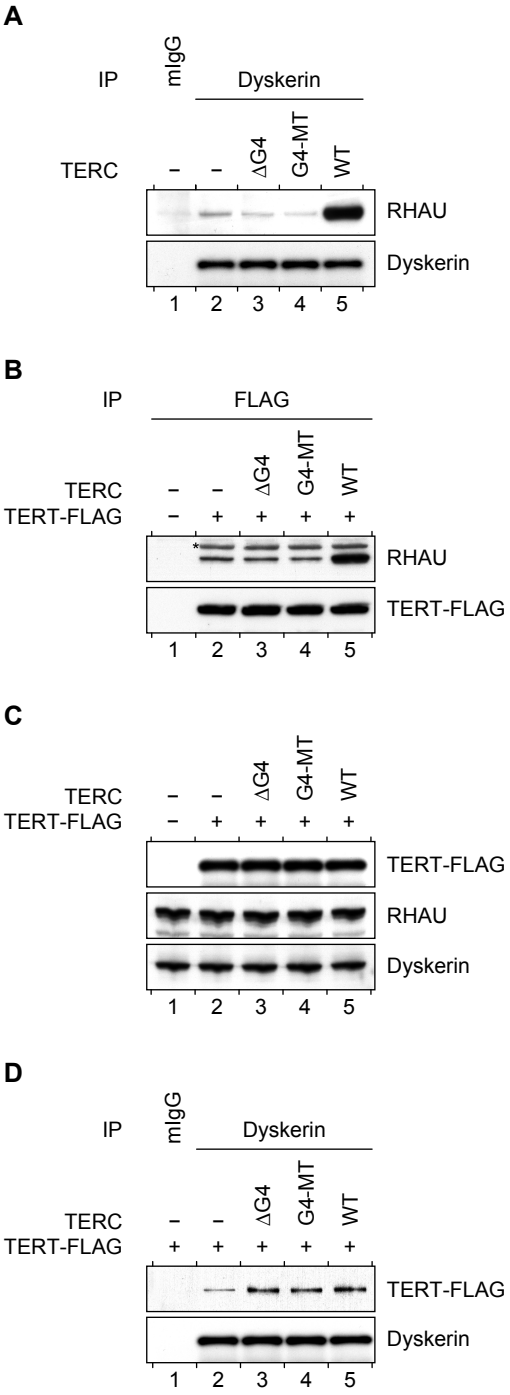


Figure 30 | **Gel mobility shift assay for TERC G4 binding by RHAU**. (**A**) Radiolabelled full-length (1–451 nt) and (**B**) 5'-end (1–71 nt) TERC fragments at a concentration of 100 pM were incubated without protein (−) or with increasing amounts (1, 3.2, 10, 32, 100 and 320 nM) of GST-tagged RHAU in the absence of ATP. The reaction mixtures were analysed by non-denaturing PAGE. An autoradiogram of the gel is shown. The positions of the free RNA substrate and the protein-RNA complex are indicated on the left. At concentrations up to 320 nM, GST protein alone had no effect on TERC mobility (data not shown). (**C**) RNA-binding assay with WT and G4 motif mutant (G4-MT) forms of TERC 5'-end (1–71 nt) fragments. Radio-labelled TERC fragments were incubated without protein (−) or with increasing amounts (3.2, 10, 32 and 100 nM) of GST-tagged RHAU. (**D**) Cation dependency of RHAU interaction with TERC. The 5'-end WT TERC fragment was incubated under standard (100 mM K^+) or lithium-based (100 mM Li^+) REMSA conditions without protein (−) or with increasing amounts (10, 32, 100 and 320 nM) of GST-tagged RHAU.

fragments resulted in the appearance of a high-affinity (estimated K_d of 10 nM) ribonucleoprotein complex of reduced mobility. In contrast, the stability of the RHAU-TERC interaction was strongly impaired (~20-fold reduction) when formation of the G4 structure was prevented by mutation of the G4 motif sequence (G4-MT, Figure 30C). Similarly, conditions that are unfavourable to G4 stability (substitution of K^+ for Li^+) impaired the RHAU-TERC interaction to comparable extent (Figure 30D). Replacing K^+ by Li^+ was indeed shown to strongly reduce the thermodynamic stability of the TERC G4 structure (224). Thus, these results demonstrate a direct and specific interaction between RHAU and TERC dependent on RNA folding into a stable G4 structure. This is consistent with the above finding that RHAU can co-immunoprecipitate the wild-type but not the G4 motif mutated forms of TERC. Together these observations argue that a fraction of TERC RNA forms a G4 structure that can be further bound by RHAU, *in vivo*.

RHAU associates with telomerase RNPs by direct interaction with TERC

In cells, the biogenesis of the active telomerase enzyme follows a stepwise RNP assembly. Upon transcription, nascent TERC is first bound by proteins dyskerin, NHP2 and NOP10 and subsequently joined by GAR1 (for review, see



ref. 330). The co-transcriptional binding of this RNP complex is essential for the processing and accumulation of TERC transcripts. Finally, the catalytic protein component telomerase reverse transcriptase (TERT) assembles with the processed telomerase RNPs and forms the active telomerase holoenzyme. To determine whether RHAU was exclusively interacting with TERC during the transcriptional process or whether it was also part of the telomerase RNP, we performed co-immunoprecipitation assays and examined whether endogenous RHAU co-fractionated with endogenous dyskerin or FLAG-tagged TERT proteins. Immunoblot analysis showed that RHAU was present in both dyskerin and TERT immunoprecipitated fractions but absent in the control IP (Figure 31A and B, compare lanes 1 and 2). The association of RHAU with components of the telomerase holoenzyme proved to be strictly dependent on the ability of RHAU to bind the 5'-end of TERC, since overexpression of wild-type but not the G4 motif mutated forms of TERC resulted in extensive enrichment of coprecipitating RHAU protein (Figure 31A and B, compare lane 2 to lanes 3, 4 and 5). Although mutations of the TERC G4 motif sequence severely impinged on the recruitment of RHAU to the telomerase, they had no apparent repercussions on the steady-state expression level (Figure 31C) nor on the assembly of the core components of telomerase. Indeed, as shown in Figure 31D (compare lane 2 to lanes 3, 4 and 5), similar levels of TERT protein co-fractionated with dyskerin following transient overexpression of either wild-type or G4 motif mutated forms of TERC. These results suggest that RHAU does not strictly bind TERC in the course of its biogenesis but that a fraction of RHAU also associates with the fully assembled telomerase holoenzyme through direct interaction with the G4 motif sequence of TERC.

RHAU associates with telomerase activity

As RHAU binding to TERC requires a G4 structure, the finding that RHAU co-immunoprecipitated with components of the telomerase complex suggested the existence of an alternatively folded form of TERC bearing a G4 structure in a fraction of telomerase RNP. To address this further, immunopurified RHAU RNP complexes were analysed for telomerase activity by TRAP assay. In all the subsequent experiments, parallel immunopurification of dyskerin served as a positive control for telomerase activity (331). In agreement with previous observations, TRAP assays of antibody-bound complexes confirmed that substantial telomerase activity was recovered with RHAU but not with control IgGs (Figure 32A, top). Coomassie staining demonstrated similar yields of immunoprecipitated RHAU and dyskerin proteins (Figure 32A, bottom) and protein identity was subsequently verified by Western blotting (Figure 32B).

Figure 31 | **Association of RHAU with telomerase holoenzyme subunits**. (**A**) TERC G4 motif dependent association of RHAU with dyskerin. Proteins from whole HEK293T cell lysates of either mock-transfected (−) cells or transiently expressing WT or G4 motif mutant (G4-MT, ΔG4) forms of TERC were immunoprecipitated with either control mouse IgGs (mIgG) or anti-dyskerin antibodies. Immunopurified RNP complexes were separated by SDS-PAGE and probed with anti-RHAU or anti-dyskerin antibodies. (**B**) TERC G4 motif-dependent association of RHAU with TERT. Protein immunoprecipitation experiments with anti-FLAG antibodies were performed with whole cells lysates of HEK293T cells transiently expressing TERT-FLAG protein together with WT or G4 motif mutant (G4-MT, ΔG4) forms of TERC. Immunopurified RNP complexes were separated by SDS-PAGE and probed with anti-RHAU or anti-FLAG antibodies. The asterisk denotes immunoprecipitated TERT-FLAG protein that cross-reacted unspecifically with the horseradish peroxidase-conjugated secondary antibody. (**C**) Western blot analysis of comparable protein expression in input HEK293T cell lysates. (**D**) Protein immunoprecipitation experiments with anti-dyskerin antibodies were performed with whole cells lysates of HEK293T cells transiently expressing TERT-FLAG protein together with WT or G4 motif mutant (G4-MT, ΔG4) forms of TERC. Immunopurified RNP complexes were separated by SDS-PAGE and probed with anti-FLAG or anti-dyskerin antibodies.

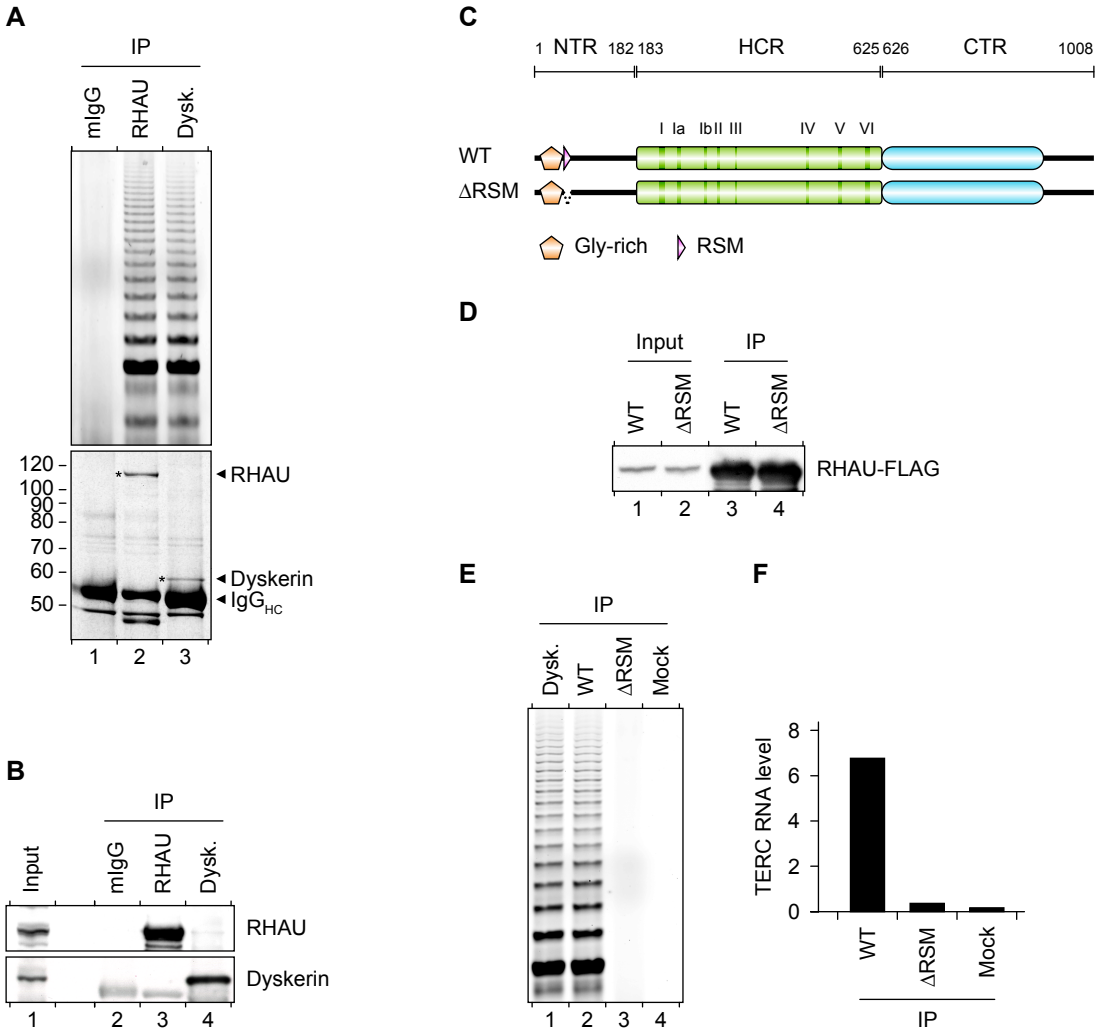


Figure 32 | **Association of RHAU with telomerase activity**. (**A**) TRAP assay of immunopurified endogenous RHAU and dyskerin RNP complexes. RHAU or dyskerin RNP complexes were enriched by immunoprecipitation and a fraction of the immunopurified RNP was essayed for telomerase activity by the TRAP assay. Mouse IgGs (mIgG) served as a control to assess nonspecific interactions to the antibody matrix. A fraction of the immunopurified RNP was separated by SDS-PAGE. The Coomassie stained gel was scanned on a LI-COR Odyssey infrared imaging system. Positions of the immunoprecipitated proteins as well as the immunoglobulin heavy chains (IgG_{HC}) are indicated at the right. Positions and sizes (kDa) of marker proteins are shown at the left. (**B**) Western blot analysis of the immunopurified RNP complexes assayed for telomerase activity. (**C**) Schematic representation of WT and G4 binding-deficient (ΔRSM) forms of RHAU. The conserved ATPase/helicase motifs I–VI of the DEAH-box family are indicated within the helicase core region (HCR) by vertical bars. The N-terminal Gly-rich (aa. 10–51) and RSM (RHAU-specific motif, aa. 54–66) domains are indicated. (**D**) Western blot analysis of the expression and enrichment of FLAG-tagged WT and ΔRSM forms of RHAU. (**E**) TRAP assay of immunopurified FLAG-tagged WT and ΔRSM forms of RHAU RNP complexes. (**F**) RT-qPCR analysis of the levels of endogenous TERC co-immunoprecipitating with FLAG-tagged WT and ΔRSM forms of RHAU RNP complexes. TERC RNA levels in IP fractions were normalised to endogenous levels of the input.

To further examine the apparent significance of G4 binding by RHAU for its recruitment to the telomerase holoenzyme, we transiently expressed a FLAG-tagged mutant version of RHAU (ΔRSM, Figure 32C) that was unable to bind to G4 structures (88). Western blot analysis confirmed that the wild-type and G4 binding-deficient (ΔRSM) forms of RHAU were expressed and recovered to similar levels (Figure 32D). Immunopurified RHAU fractions were further assayed for telomerase activity and enrichment of TERC (Figure 32E and F). As previously observed for endogenous RHAU, immunopurified FLAG-tagged RHAU(WT) efficiently recovered telomerase activity and co-precipitated TERC. In contrast, negligible telomerase activity and TERC signal were retrieved from the FLAG-tagged RHAU(ΔRSM) mutant, showing that G4 binding by RHAU is a prerequisite for RHAU binding to TERC. Altogether, these findings provide persuasive evidence that a fraction of RHAU associates with a subpopulation of

the active telomerase holoenzyme *in vivo* through direct interaction with the G4-motif sequence of TERC.

Finally, to determine the proportion of telomerase activity associated with RHAU, we carried out immunodepletion studies of RHAU in HEK293T cell lysates and subsequently quantified the residual TERC level and telomerase activity within these fractions (Figure 33A and B). Depletion with RHAU antibody significantly ablated telomerase activity in an aliquot of the supernatant by 25 %, while nearly 70 % of the telomerase activity of the input fraction was lost following depletion of endogenous dyskerin. As a control, immunodepletion with control mouse IgGs produced no significant reduction in telomerase activity. Western blot analysis of the supernatants confirmed that both RHAU and dyskerin proteins were effectively depleted from the corresponding fractions (Figure 33C). Furthermore, the quantification by RT-PCR of RNA levels in the immunodepleted fractions was consistent with the above findings. Indeed, approximately one quarter of input TERC, but not unrelated RNAs such as β-actin mRNA, vanished following depletion of endogenous RHAU. These results not only corroborate the association of RHAU with telomerase, but also show that the fraction of telomerase interacting with RHAU accounts for merely 25 % of the total telomerase activity. Nevertheless, these data provide indirect but explicit evidence of the existence of an alternatively folded TERC structure bearing a G4 scaffold in the fully assembled telomerase holoenzyme.

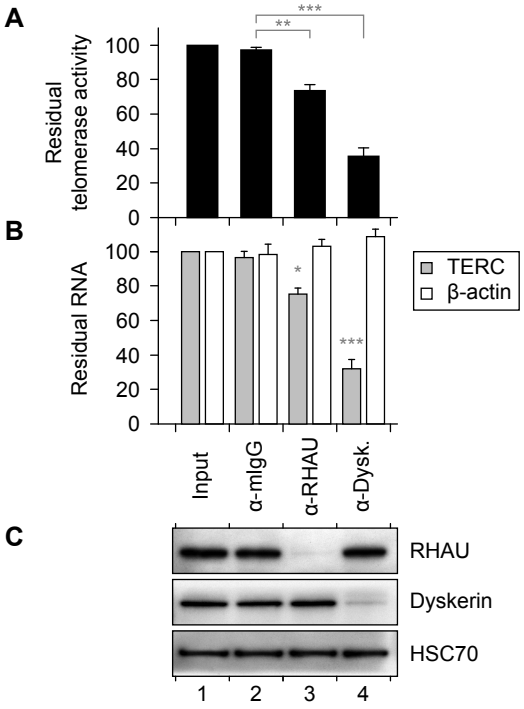
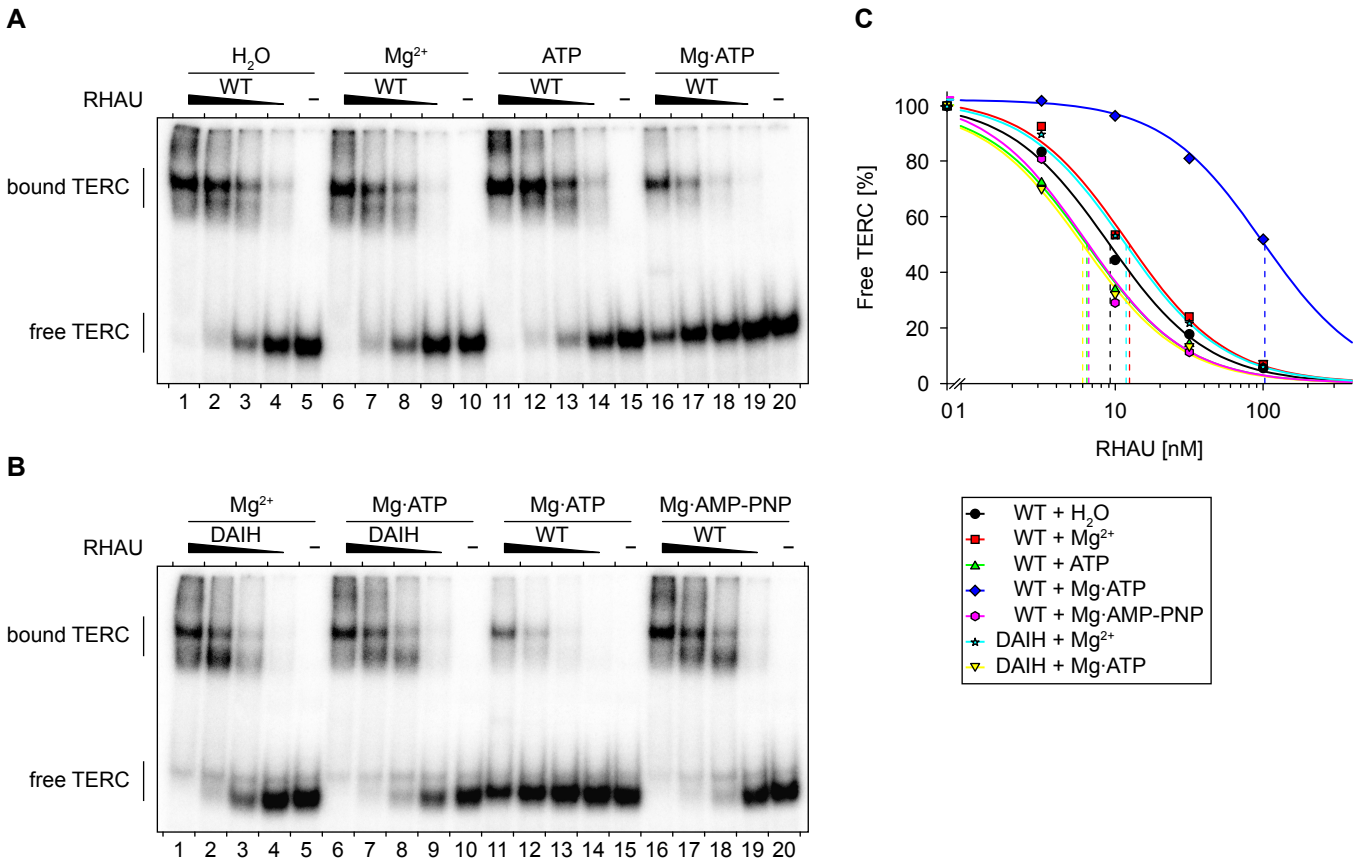


Figure 33 | **Quantification of the fraction of TERC and telomerase activity associated with RHAU**. (**A**) qTRAP analysis of the residual telomerase activity present in immunodepleted HEK293T cell lysates. Data represent the mean ± SEM of three independent experiments. Statistical significance was determined by one-way ANOVA and the Bonferroni *t*-test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001. (**B**) RT-qPCR analysis of the residual levels of TERC and β-actin RNAs in immunodepleted HEK293T cell lysates. The RNA signal was normalised to GAPDH. Data represent the mean ± SEM of three independent experiments. Statistical significance was determined by one-way ANOVA and the Bonferroni *t*-test. (**C**) Western blot analysis of input and immunodepleted HEK293T cell lysates. Heat shock protein 70 cognate (HSC70) was used as a loading control.

ATPase-dependent interaction of RHAU with TERC

The fact that only one quarter of telomerase activity is linked to RHAU may reflect a dynamic interaction between RHAU and TERC. Indeed, most of the DEAH-box RNA helicases only transiently interact with nucleic acid, because they do not remain bound after the ATPase-dependent remodelling of their substrate (47,80,94,332). To gain insight into the role of ATP-hydrolysis by RHAU in its interaction with TERC, we repeated gel mobility shift assays with 5' TERC fragments in the presence of Mg^{2+} or ATP or both. As shown in Figure 34A, ATP or Mg^{2+} alone had little effect on the binding affinity of RHAU



for TERC. In contrast, under conditions supporting ATP hydrolysis (Mg·ATP), RHAU displayed reduced binding affinity for TERC, as evidenced by a 10-fold increase in the apparent equilibrium K_d value (Figure 34C). Such a reduction in the affinity of RHAU for TERC was not observed when Mg·ATP was substituted with the non-hydrolysable ATP analogue Mg·AMP-PNP (Figure 34B and C). Thus, the present data suggest that intrinsic RHAU ATP hydrolysis rather than ATP binding is essential for disrupting the interaction of RHAU with TERC. To further validate this, we turned to the previously described RHAU(DAIH) mutant in which the E335A amino acid substitution within the Walker B site abolishes RHAU ATPase activity (88). Unlike wild-type RHAU, the binding affinity of RHAU(DAIH) mutant for TERC was similar in the absence and presence of Mg·ATP (Figure 34B and C). The absence of RHAU helicase activity as a consequence of the loss of ATPase activity probably prevents RHAU dissociation from its substrate. In fact, a similar reduction in *in vivo* RNA binding dynamics after suppression of RHAU intrinsic ATPase activity was reported previously in studies of the shuttling of RHAU in cytoplasmic stress granules (268). Taken together, these data corroborate the idea that RHAU is not permanently associated with TERC but dissociates upon ATP hydrolysis.

Apart from a high affinity for G4 structures, RHAU has also been shown to couple ATP hydrolysis with tetramolecular G4-RNA resolving activity (88,89). These biochemical properties prompted us to test whether RHAU could catalyse the conversion of TERC G4 structure to ssRNA form. However, the faster rate of refolding of intramolecular G4 structures with regard to tetramolecular G4-RNAs made it technically difficult to monitor the G4 resolving activity of RHAU using standard non-denaturing gel electrophoresis.

Figure 34 | **Role of Mg·ATP on the interaction of RHAU with TERC**. (**A**) Gel mobility shift assay for TERC G4 binding by RHAU. Radiolabelled 5'-end (1–71 nt) TERC fragment at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (3.2, 10, 32 and 100 nM) of purified recombinant GST-tagged RHAU in the presence of 1 mM Mg^{2+}, ATP or Mg·ATP (as indicated). The reaction mixtures were analysed by non-denaturing PAGE. An autoradiogram of the gel is shown. (**B**) Gel mobility shift assay of TERC G4 binding by the ATPase deficient RHAU(DAIH) mutant. Radiolabelled 5'-end TERC fragment was incubated without protein (−) or with increasing amounts (3.2, 10, 32 and 100 nM) of purified recombinant WT or DAIH mutant GST-tagged RHAU in the presence of 1 mM Mg^{2+}, Mg·ATP or Mg·AMP-PNP (as indicated). (**C**) Quantification of gel mobility shift assays of WT and RHAU(DAIH) mutant binding to TERC 5'-end (1–71 nt) fragment. The data represent the mean of three independent experiments. Error bars for SEM are omitted for clarity.

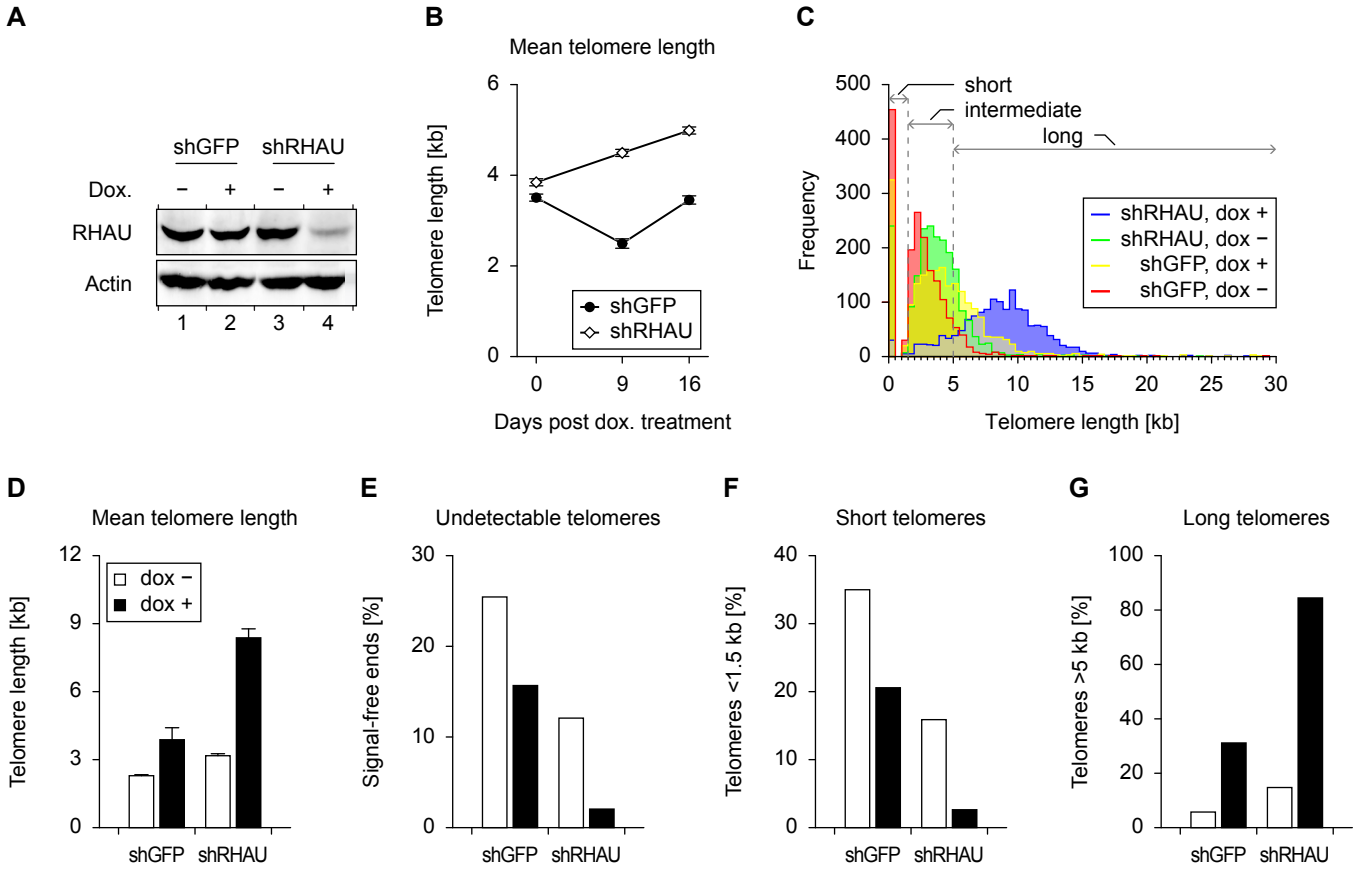


Figure 35 | **Telomeres get gradually longer in RHAU knock-down cells**. (**A**) Western blot analysis of the efficiency of RHAU downregulation in HeLa T-REx cell lines stably transfected with doxycycline-inducible vector of shRNA against RHAU (shRHAU). (**B**) Mean telomere length of HeLa T-REx cell lines stably downregulated for RHAU after 0, 9 and 16 days of culture in the presence of 1 μg·ml^{-1} doxycycline. (**C–G**) Telomere length analysis of HeLa T-REx cell lines stably transfected with doxycycline-inducible shRNA expression plasmids after 42 days of culture in the presence of 1 μg·ml^{-1} doxycycline. (**D**) Mean telomere length. (**E**) Undetectable telomeres. (**F**) Short (< 1.5 kb) telomeres. (**G**) Long (> 5 kb) telomeres.

Telomere lengthening phenotype in RHAU knocked-down cells

As evident in some human telomeropathies, deficiencies in proteins essential to telomerase biogenesis have often been related to decreased telomerase activity and accelerated telomere erosion phenotypes (333). Thus, to clarify the functional impact of RHAU on telomerase activity, we undertook to assess telomere length in T-REx-HeLa cell lines stably transfected with doxycycline-inducible vector of shRNA against RHAU (shRHAU). As a control, we included a shRNA construct directed against the heterologous GFP mRNA (shGFP). After cells were treated for seven days with doxycycline, RHAU protein levels were efficiently reduced in shRHAU cells, whereas doxycycline had no detectable effect on shGFP control cells (Figure 35A). The cells were cultured further for 16 more days with doxycycline and telomere length was assessed by quantitative FISH (Q-FISH, ref. 334). As shown in Figure 35B, we found that in average the cells in which RHAU was down-regulated had slightly longer telomeres (~5.0 kb) than did the control cells (~3.5 kb). As the effect was not very pronounced after 16 days of culture, the experiment was repeated but the cells were let to divide for a 42-day period. After six weeks of culture, the mean telomere length from RHAU knocked-down cells was about 3-fold higher compared to telomeres from control cells (Figure 35D). Although RHAU knocked-down cells had marked telomere length heterogeneity (Figure 35C), short and undetectable telomeres in these cells were nearly inexistent (Figure 35E and F). Reciprocally, more than 80 % of RHAU knocked-down cells presented telomeres longer than 5 kb while only 30 % of control cell telomeres exceeded this value (Figure 35G). Taken together, these two time course experiments show that telomeres get gradually longer (~100 bp per day) in cells where RHAU expression was dampened, suggesting a possible direct or indirect negative role of RHAU on human telomere homeostasis.

DISCUSSION

The propensity of nucleic acid guanine-rich sequences to self-assemble into G4 structures *in vitro* has been recognised for several decades. Although RNA is also prone to form such structures, G4-RNA has not attracted as much attention as G4-DNA. Nevertheless, a growing body of evidence indicates a significant role for G4 structure formation during RNA metabolism (207,335). As a consequence, G4 structures emerged as a plausible post-transcriptional means of regulating the function of coding and non-coding RNAs. However, little is known about the mechanisms by which G4-RNA formation is regulated in the cells. Indeed, only a few proteins have been reported so far to interact with G4-RNAs *in vitro* (254,259,262,265,267) and, apart from FMRP and FMR2P, experimental data on their biological activity subsequent to their interaction with G4-RNAs is scarce. Of these G4-RNA binding proteins, RHAU is one of the rare proteins that exhibits robust *in vitro* ATPase-dependent G4-RNA resolving activity, in addition to a high affinity and specificity for its target RNAs (89). These particular biochemical features promote RHAU as an ideal candidate for regulating G4-dependent RNA metabolism, although its *in vivo* RNA targets have not yet been determined. The present study aimed to identify naturally occurring RHAU targets with the emphasis on RNAs containing potential G4-forming sequences. By RIP-chip assays, 106 RNAs were found to be significantly enriched with RHAU. Importantly, more than half of these RNAs contained G-rich sequences with potential to form stable G4 structures. In addition, there was a weak but significant correlation between the predicted G4 motif density per transcript and the magnitude of RNA enrichment by RHAU. Nevertheless, because RHAU appears to show significant affinity for RNAs not bearing G4-forming sequences, we do not exclude the possibility that RHAU has affinity to other RNA structural features.

In-depth studies revealed that TERC, one of the identified RNAs, was a *bona fide* target of RHAU. Several independent RIP-chip assays using various cell lines further showed TERC to be one of the most abundant RNAs enriched by RHAU. TERC bears a 5' G-rich sequence that was previously shown to adopt a stable intramolecular G4 structure *in vitro* (224). Further investigations dissecting the basis of the interaction between RHAU and TERC showed RHAU binding to be strictly dependent on G4 structure formation in the 5' region of TERC and to require the N-terminal RSM domain of RHAU, an ancillary domain necessary for specific recognition of G4 by RHAU (88). Moreover, RHAU was found to interact transiently with TERC in a manner dependent on its own ATPase activity. Together, these data provide the first evidence of a specific and direct interaction between a G4 resolvase enzyme and a potentially relevant intramolecular G4-RNA substrate. Furthermore, in agreement with our previous reports (88,89), these data attest that G4-RNAs are naturally occurring substrates of RHAU *in vivo*.

The 5' extremity of TERC folds into an intramolecular G4 structure *in vivo*

In agreement with the observations presented here, Mergny and co-workers reported previously the formation *in vitro* of a G4 structure in the 5' region of human TERC (224). However, as for the great majority of nucleic acid structures, direct experimental demonstration of G4 structures *in vivo* is inherently difficult to prove. As such, our results do not directly prove the existence of G4 structures *in vivo*, but strongly support the notion that TERC can form a G4 structure in the cells. Notably, RHAU can only bind TERC provided that all conditions necessary for the formation of an intramolecular G4 structure (RNA

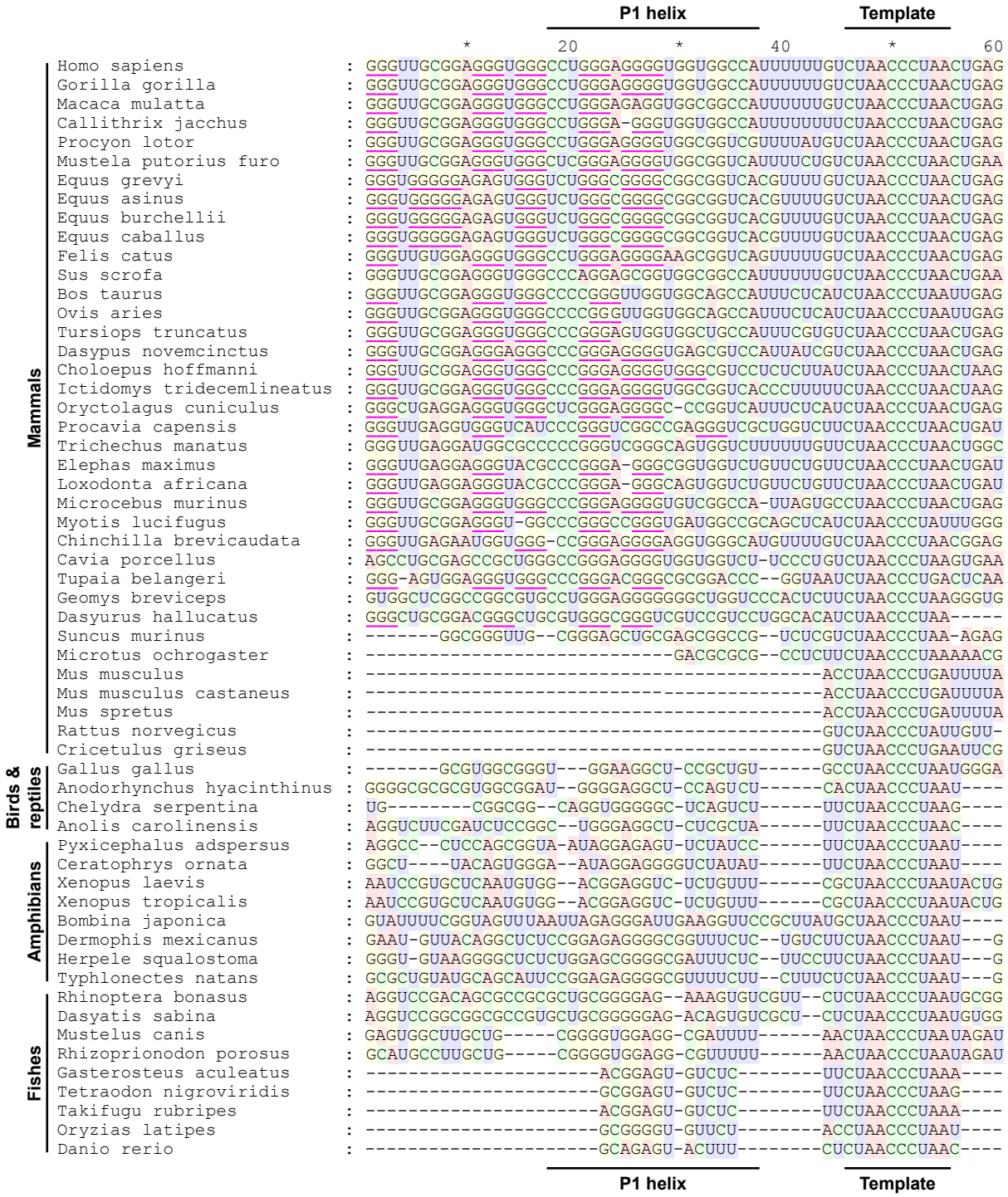


Figure 36 | **Conservation of the 5' G4 motif sequence among mammalian orthologues of TERC**. Multiple sequence alignment of the 5'-regions of various TERC orthologues was carried out with MAFFT (version 6, ref. 313). Guanine tracts that are predicted to form a stable G4 structure are underlined. The P1 helix as well as the template region are indicated.

sequence and ionic conditions) are met. Modifying any one of the conditions is sufficient to reduce substantially the affinity of RHAU for TERC. Moreover, we predict that formation of a G4 structure in TERC is not exclusive to human cells, since the majority of the mammalian orthologues of TERC – with the notable exception of the rodent forms – also harbours a potential 5' G4-forming sequence (Figure 36). Insofar as the G4-RNA binding and resolving activities of RHAU are conserved in higher eukaryotes (88), it would not be surprising to find that RHAU binds the telomerase RNA moiety of other mammals.

As reported previously (224), folding of the 5' region of TERC into a G4 structure is likely to interfere with the widely adopted secondary structure of human TERC (Figure 37). Indeed, according to the standard model of TERC secondary structure (336), two of the guanine tracts that are part of the

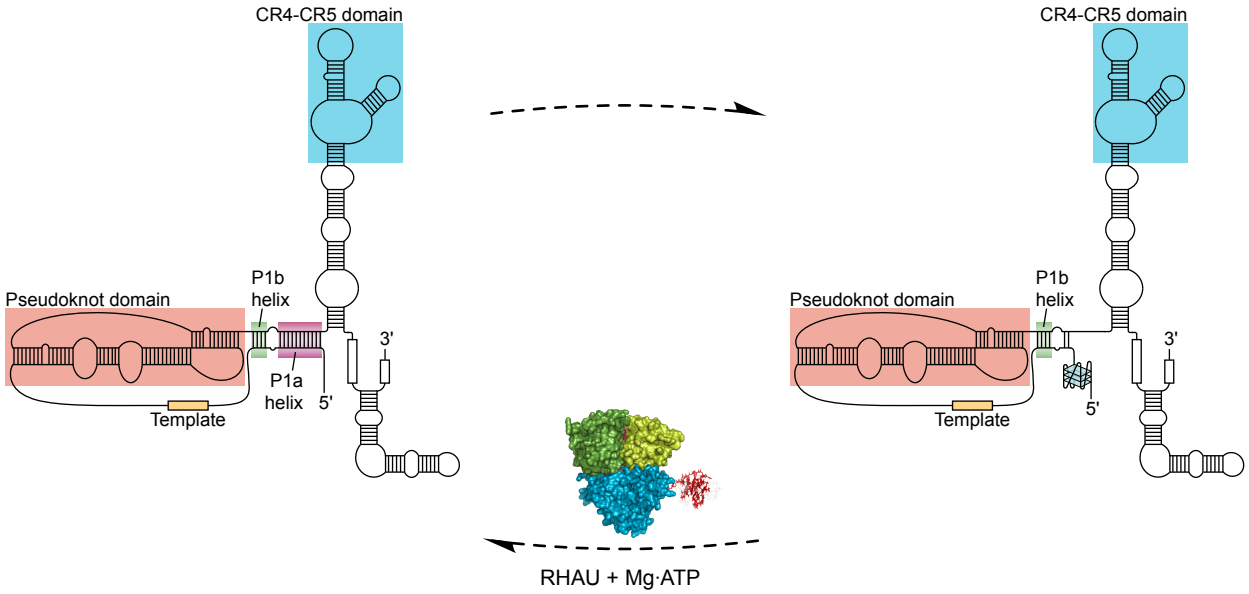


Figure 37 | **Dynamic equilibrium between the P1a duplex- and G4-folded forms of TERC**. RHAU may couple ATP-hydrolysis with unwinding of the 5' G4 structure to restore helix P1 base-pairing.

intramolecular G4 also form the P1a helix. As a consequence, the P1a helix and G4 represent two mutually exclusive structures and may correspond to different functional states of TERC. Insofar as helix P1 is required for template boundary definition in mammalian telomerase (337), formation of a G4 structure in TERC may be detrimental to telomerase activity. On the other hand, G4 structure formation may protect TERC from degradation during telomerase biogenesis. Such a scenario was suggested quite recently by Collins and co-workers, who found that nucleotide substitutions within the G4-forming sequence markedly reduced accumulation of the mature form of TERC, thereby causing telomere shortening (34). Similar to our data, they also found that RHAU associates with the G4-forming sequence of TERC *in vivo*, but in contrast to our finding, detected negligible telomerase activity with immunopurified RHAU fraction. Thus, further experiments are necessary to clarify the functional impact of G4 formation in TERC on telomerase biogenesis and activity.

In cells, the P1 duplex- and G4-folded forms of TERC are likely to coexist in dynamic equilibrium. However, under physiological conditions, restoration of helix P1 base pairing from the G4-folded conformation of TERC is likely to require a catalyst due to the high thermodynamic stability of the quadruplex (224). The biochemical properties and substrate specificity of RHAU make it a likely candidate for catalysing such a reaction. Quite apart from its high affinity for TERC, RHAU manifests robust tetramolecular G4-RNA resolving activity (88,89) and was also shown to unwind various intramolecular G4-DNA structures *in vitro* (39). These data suggest that RHAU can resolve G4 structures irrespective of the strand stoichiometry of the G4 stem. Although unwinding of TERC G4 by RHAU has not yet been demonstrated experimentally, our observation that RHAU dissociates from TERC upon ATP-hydrolysis, but not upon ATP-binding, supports the idea that RHAU couples ATP-hydrolysis to conformational changes of TERC, resulting in the resolution of the G4 and further release of RHAU from its target RNA.

Finally, our data show clearly that RHAU does not exclusively interact with TERC during the early stages of telomerase RNP assembly, but that it is also part of a fraction of the fully assembled telomerase holoenzyme. Notably, RHAU could be co-immunoprecipitated with TERT, the catalytic subunit of telomerase, and substantial telomerase activity could also be recovered with purified endogenous RHAU. Insofar as the association of RHAU with telomerase components relies exclusively on the presence of a G4 structure

in TERC, these data provide circumstantial arguments for the existence of an alternatively conformational state of TERC bearing a G4 structure in a fraction of telomerase holoenzyme. Based on the fraction of telomerase activity co-precipitating with RHAU, we can estimate that at least 25 % of telomerase in HEK293T cells bears a G4 structure in its RNA moiety. Whether this minor fraction retains partial or fully telomerase activity is still an open question that needs to be addressed in future. Besides, further studies are necessary to provide explanations for the apparent negative effect of RHAU on human telomere homeostasis. Currently, these results are somewhat inconsistent with the proposal of a non-functional G4 conformational state of TERC or with the observed telomere shortening phenotype in cells expressing G4 motif mutated forms of TERC (34). However, since RHAU interacts with many RNAs in cells and manifests disparate biological functions, it is above all necessary to clarify whether the telomere lengthening phenotype in RHAU knockdown cells results from a direct or indirect effect of RHAU on telomerase activity.

RHAU may target other RNAs containing G4 structures

Apart from TERC, 54 further RNAs found to be associated with RHAU are predicted to form stable G4 structures (Supplementary Table II). The determination of the modalities and the significance of these interactions are essential issues to be addressed in the future. Nonetheless, the finding that RHAU associates preferentially with transcripts bearing potential G4-forming sequences strongly suggests that RHAU targets G4-RNA structures in cells. Although nearly three-quarters of these G4-forming sequences are located in 5' or 3' UTRs of mRNAs (Supplementary Table II), it is important to stress that, to date, we have not observed any suppressive action of RHAU on translational repression by G4 formation in 5'-UTRs. Therefore, RHAU is unlikely to function as a general translational activator but may instead intervene in other aspects of mRNA metabolism, such as pre-mRNA processing or mRNA turnover. Indeed, RHAU localises predominantly in the nucleus, where it concentrates in nuclear speckles (269). These are sites of high transcriptional activity and mRNA splicing. Furthermore, RHAU relocalises to cytoplasmic stress granules in response to various cellular stresses (268). Although recruitment of RHAU to stress granules is mediated by interaction with RNA, we have not yet examined whether this phenomenon depends upon binding of RHAU to G4-RNA structures. However, considering that RSM-domain mutated forms of RHAU deficient in G4-RNA binding also show reduced relocalisation to stress granules, it is likely that a fraction of RHAU binds G4 structures in stress granules.

Collectively, the data presented here provides compelling evidence that RHAU is a *bona fide* G4 binding protein and of the existence of a G4 structure in TERC. Our results corroborate another study conducted by Collins and co-workers, who found that the presence of the G4 motif sequence in TERC was an essential determinant for its interaction with RHAU and for optimal progression of telomerase assembly (34). Therefore, the G4-dependent interaction between RHAU and TERC seems fairly well established. Nonetheless, it remains to be further clarified whether RHAU also contributes to regulating telomerase activity apart from its role during the biogenesis process. Besides, considering that G4 structures may also form at telomeres or on telomeric RNA repeats (TERRA), one should keep in mind that RHAU may also intervene in telomere homeostasis independently of its activity on telomerase. Finally, various lines of evidence argue that telomerase has extra non-canonical functions which are unrelated to telomere maintenance (338). Insofar as TERC is still required for some of these functions, one should consider whether RHAU also plays a role in these telomere lengthening-independent processes.

Structural model of the helicase core and HA regions of RHAU

AIMS AND RATIONALE

As previously shown for Prp43, acquiring structural data of helicases is a crucial step towards a better understanding of the mechanisms by which these molecular motors bind and process their substrates. However, hitherto, our attempts to purify high yields of soluble recombinant RHAU protein have not been successful which has precluded further crystallographic studies. Nonetheless, we took advantage of the reasonable degree of sequence similarity between paralogous DEAH-box proteins to build a homology model of RHAU. Homology modelling is currently the most accurate computational method to generate reliable three-dimensional models of proteins presenting high sequence similarity to another known protein structure (339,340). In order to interpret our biochemical data from a structural perspective, we generated a model of RHAU based on the high-quality and high-resolution crystal structure of the prototypical DEAH-box protein Prp43 (113). With this approach, it emerges that the helicase core and the C-terminal HA region of RHAU are likely adopting a comparable fold to that of Prp43. These two regions in RHAU likely define the minimal functional catalytic unit for translocation along single-stranded nucleic acid. Overall, these computational structural data are consistent with our biochemical data and offer some valuable clues as to way RHAU may bind and unwind G4 nucleic acid structures.

RESULTS AND DISCUSSION

Modelling and model quality assessment

Template selection and sequence alignments identified Prp43 as the best structural candidate with the highest similarity and sequence coverage for building a model of RHAU. Indeed, a large part of Prp43 amino-acid sequence (aa. 81–721) displays reasonable similarity (30 % identity and 50 % similarity) over 70 % (aa. 195–913) of RHAU sequence (Supplementary Figure 14).

The three-dimensional model was generated by MODELLER (297) and had first to be refined as the program failed to provide structural predictions for three regions (1–182, 400–470 and 918–1008) of RHAU due to insufficient sequence similarity with Prp43. Nonetheless, MODELLER successfully produced a 664-amino-acid model of RHAU including the helicase core and the HA region. ModFOLD (300) and QMEAN (299) programs were further employed to estimate the overall quality of the refined model (Supplementary Figure 13). QMEAN program credited the modelled structure a QMEAN Z-score value of −1.50 that is indicative of an intermediate overall quality model. Likewise, the computed RHAU structure obtained a ModFOLD global model quality score value of 0.216 with a 2.3 % probability of incorrectness. As shown in Supplementary Figure 13D and F the less reliable sections of the computational model correspond to the C-terminal part (aa. 700–917) and some loop regions. Nevertheless, on the basis of the results provided by the two verification algorithms (Supplementary Figure 13B and E), we considered the modelled structure of RHAU of sufficient quality to provide meaningful insights regarding the mechanisms of nucleic acids binding and rNTP selectivity.

Overall comparison of RHAU and Prp43 structures

Despite a moderate sequence similarity, the helicase core and the HA region of RHAU show a remarkable structural similarity to that of Prp43, with

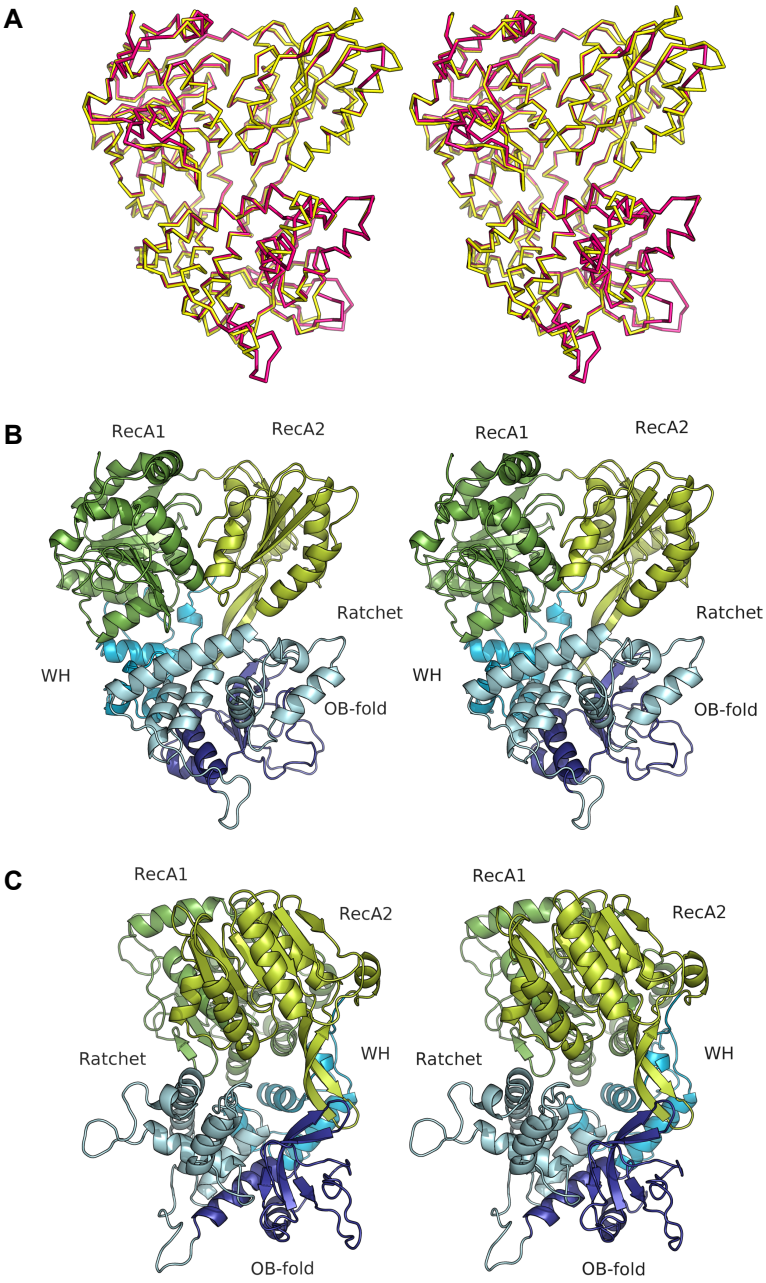


Figure 38 | **Structural model of the helicase core and HA regions of RHAU**. (**A**) Three-dimensional protein structure superposition between RHAU model (pink) and Prp43 (yellow, RCSB PDB ID: 2XAU chain A). (**B**) Front and (**C**) side views of RHAU structural model (residues 183–399 + 471–917). The helicase core (RecA1 and RecA2 domains) as well as the HA regions [winged-helix (WH), Ratchet and OB-fold domains] are coloured in green and blue, respectively.

a C_{α} RMSD value of 1.71 Å over 632 residues (Figure 38A). The main structural discrepancies include several mismatched loops corresponding to residues: 255–258, 512–516, 542–544, 655–656, 769–779, 815–826, 850–858 and 875–882. Besides, as mentioned earlier, a 71 amino acid-long region (400–470) in the N-terminal region of the RecA2 domain of RHAU could not be modelled because of the lack of sequence similarity with Prp43. Such an extension of the RecA2 domain is only found in the DEAH-box proteins of the RHA group and is totally absent in the other (spliceosomal) DEAH-box proteins like Prp43 (Supplementary Figure 1). Besides, no known function has yet been attributed to this short sequence. Even so, the overall structural organisation of the modelled region of RHAU is nearly identical to that of Prp43. Indeed, the helicase core region folds into two universal and tandemly linked RecA-like domains while the HA region consists of three distinct domains: (a.) a winged-helix (WH) domain, (b.) a seven-helical bundle, referred as the 'ratchet' domain, and (c.) an OB-fold motif arranged in a five-stranded β-barrel (Figure 38B and C). Therefore, in view of the structural similarities of the helicase core and HA regions of RHAU with those of Prp43, it is probable that both proteins function and unwind RNA by

a comparable mechanism. This result at least agrees with the concept that most DEAH-box proteins share a similar overall structural organisation that combines a minimal functional unit, the helicase core and the HA domains, with flanking regions of various lengths and sequences (30). Unfortunately, both homology and *ab initio* modelling methods did not permit to get meaningful predictions for the unique N- (aa. 1–182) and C-terminal (aa. 918–1008) peripheral regions of RHAU. In consequence, our understanding of how the N-terminal region binds and endows affinity for G4 structures remains very limited.

Recognition and binding of NTPs by RHAU

The interface between the two RecA-like domains in RHAU forms a cavity that defines the NTP-binding and hydrolysis site (Figures 39A and 40B). RHAU likely binds NTPs in analogous fashion to Prp43. Indeed, RHAU lacks the Q-motif that provides ATP-binding specificity for most of SF2 helicases (Supplementary Figure 14). Instead, RHAU harbours the R- and F-motifs which non-specifically maintain the nucleobase moiety ring by hydrophobic effect between the side chains of Arg-275 and Phe-537. Accordingly, just like other DEAH-box proteins (57,83,96,107,118,131,132), RHAU was expected not to manifest a particular preference for ATP when catalysing the conversion of G4 structure to single-stranded nucleic acid. To further validate this hypothesis, RHAU G4 unwinding

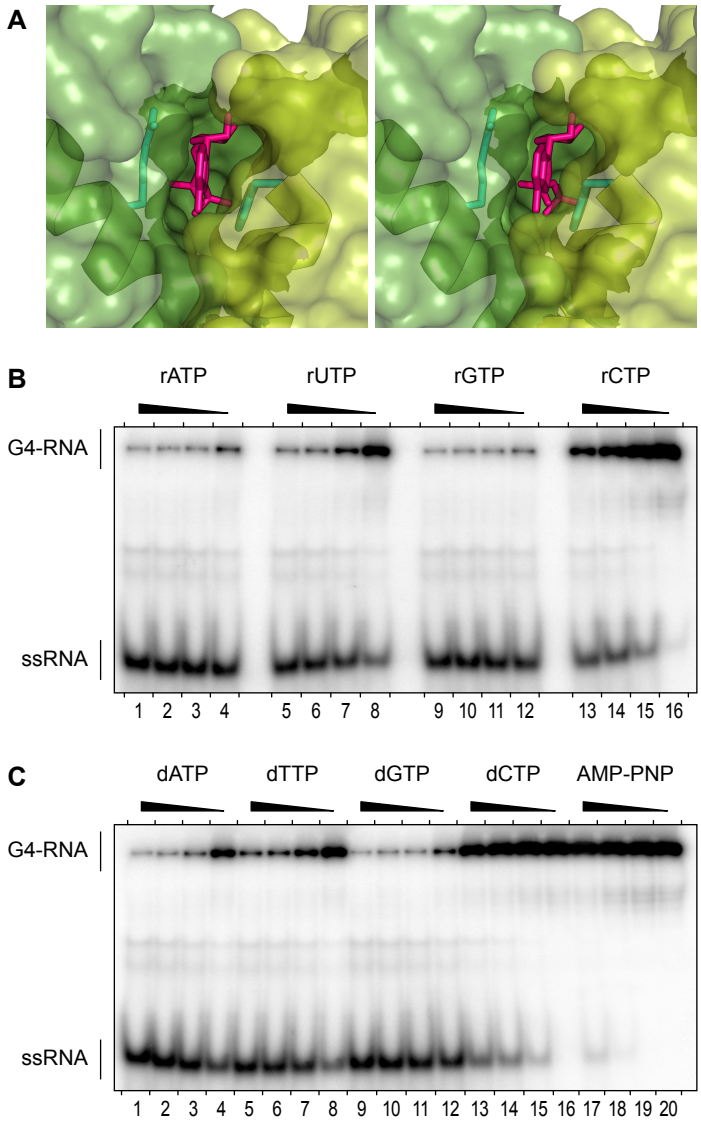


Figure 39 | **Binding and recognition of NTPs by RHAU**. (**A**) Close-up view of the NTP binding and hydrolysis site in RHAU. NTP (pink) is bound in a cavity formed at the interface of the RecA1 (dark green) and the RecA2 (light green) domains. The nucleotide base is stacked between the side chains (cyan) of Arg-275 (R-motif, RecA1) and Phe-537 (F-motif, RecA2). (**B**) All rNTPs and (**C**) dNTPs cofactors support unwinding of G4-RNA structures by RHAU. Radiolabelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of purified recombinant GST-tagged RHAU and increasing amounts (0.03, 0.10, 0.32 and 1.00 mM) of NTPs or AMP-PNP (as indicated). The reaction products were resolved by native PAGE. An autoradiogram of the gel is shown.

activity was investigated in the presence of various rNTP and dNTP cofactors. As shown in Figure 39B and C, each of the rNTPs and dNTPs, but not AMP-PNP, supported unwinding of G4-rAGA substrate by RHAU. However, the unwinding activity was substantially higher with purine cofactors than with pyrimidine rNTPs/dNTPs, with CTP/dCTP being almost inert. A comparable slight preference for purine over pyrimidine nucleotides was previously observed for the DEAH-box proteins Prp2 (107) and Prp16 (96) but not for Prp43 that showed maximal catalytic activity in the presence of UTP/dTTP (118). Thus, although our structural model of RHAU provided some valuable insights into the mechanism of nucleotide binding by RHAU, it did not permit to further clarify the molecular bases behind the observed purine nucleotide bias. Finally, considering the relatively high degree of amino acid conservation for the R- and F-motifs in RHAU as well as in other DEAH-box proteins, we predict that mutation of either the Arg or Phe residues in DEAH-box helicases should considerably reduce their respective binding affinities for NTP cofactors.

Nucleic acid binding and G4 structure recognition by RHAU

In Prp43, the junction between the helicase core and the HA region forms a groove (Figure 6C) that was proposed to accommodate the translocated single-stranded nucleic acid (113). According to this model, the RNA is held in a sequence-independent manner by amino acid residues of the helicase motifs Ia, Ib, Ic, IV, IVa and V as well as some residues of the HA region. Based on the high structural similarities between Prp43 and the 3'-to-5' helicases NS3 and Hel308, the 5' double-stranded substrate region of the translocated RNA is likely positioned proximal to the RecA2 side of Prp43 (113). This would be at least consistent with the observation that most of DEAH-box helicases translocate with a 3'-to-5' directionality (Table III and references therein). Similar to Prp43, RHAU harbours a cavity between the helicase core and the HA region (Figures 38C and 40A). Its diameter is wide enough to hold a single nucleic acid strand but there is clearly not sufficient space to accommodate a ~26-Å-wide G4-scaffold. This suggests that the four stranded structure lies outside of the core protein and that only the 3' single-stranded extremity of the nucleic acid spans the helicase domain (Figure 40B). That the helicase core region of RHAU does not directly bind the G4 structure is consistent with our previous observations showing that much of the G4 binding affinity is endowed by the N-terminal region (88). Although our computational approach did not permit to get structural information about this peripheral region (aa. 1–182), the N-terminal RSM likely binds the G4 structure near the RecA2 domain (Figure 40C). Furthermore, our current model suggests that the bound 3' tail of the translocated RNA consists of about ten nucleotides. This is clearly in agreement with the observation that unwinding of G4 structures by RHAU is dependent on the presence of a minimal 9-nt 3'-overhang (Figure 15C).

Even if little is known regarding the G4 unwinding mechanism of RHAU, its core structure has all the aspects of a processive helicase such as Hel308. RHAU notably shares with Hel308 the WH and Ratchet domains that have previously shown to play an essential role for DNA unwinding (29). A particular common feature is the ~20 amino acid-long central α-helix (α22 in Supplementary Figure 14) of the Ratchet domain that may function as a hook for directional transport of the single-stranded nucleic acid across the helicase core groove. For Hel308, Büttner et *al.* suggested that antagonistic interactions between the base moieties of the translocated nucleic acid and side-chain residues of the Ratchet domain confer extra stability to the unwound DNA and are probably important for translocation processivity (29). Thus, in view of the overall structural similarities between RHAU and Hel308, both helicases may share

an analogous NTP-driven single-strand translocation mechanism for resolving nucleic acid structures. Nevertheless, a doubt subsists regarding the degree of similarity between the two translocation mechanisms. Indeed, the helicase core region of RHAU and Prp43 more resembles to that of the viral NS3-like helicases than to that of Hel308. In particular, RHAU and Prp43 harbour a long and twisted antiparallel β-hairpin in their RecA2 domain that protrudes into a cleft of the HA region (Figures 38C and 40A). Walbott et *al.* proposed that in Prp43 this β-hairpin functions similarly to that of Hel308 and may be responsible for double-strand separation (113). However, studies on NS3 helicase indicates that the corresponding β-hairpin rather functions as a pivot point that permits to lock the RecA2 domain orientation in the ADP-bound state relative to the rest of the protein (125). Therefore, there is still some degree of uncertainty regarding the role of this β-hairpin and additional investigations are necessary to clarify the mechanisms underlying unwinding of nucleic acid structure by DEAH-box helicases. Solving a particular DEAH-box protein structure in different conformational states would be one way to put the above ideas under experimental testing. In such a case, RHAU bound to TERC G4 structure would certainly provide an appropriate prototype.

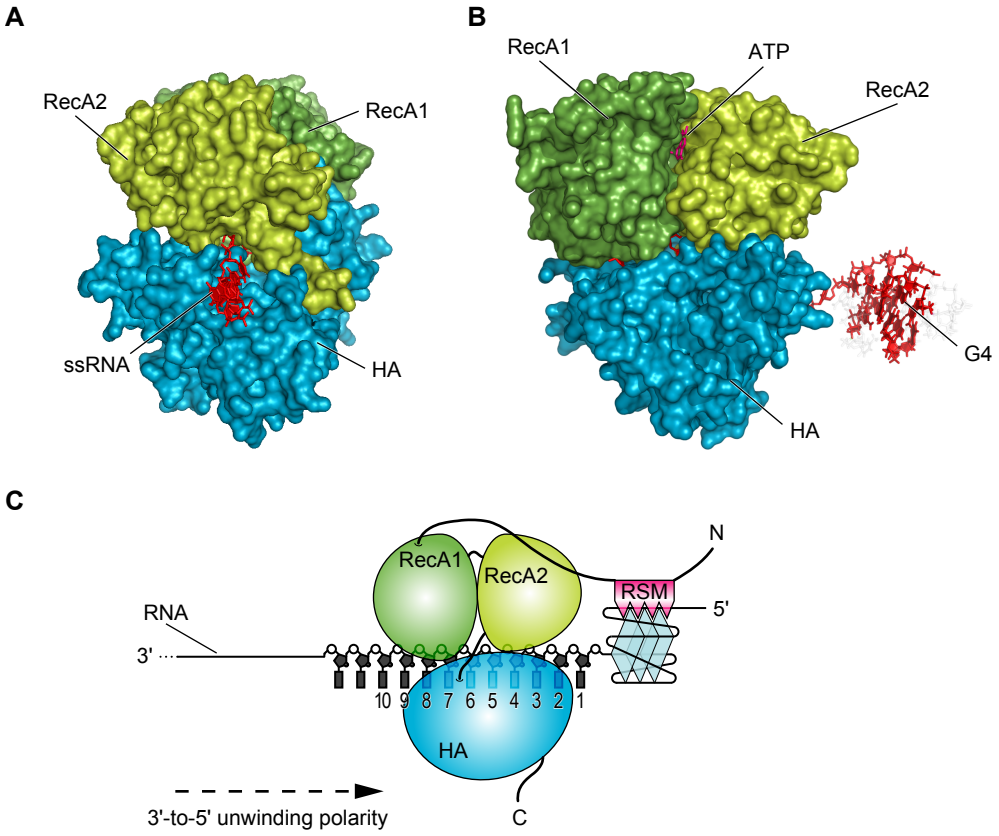


Figure 40 | **G4 structure binding by RHAU**. (**A**) Side view of the modelled RHAU bound to ssRNA. (**B**) Front view of the modelled RHAU bound to G4-RNA. (**C**) Schematic and hypothetical representation of the binding and RSM-depend recognition of G4 structure by RHAU.

CHAPTER 4

Discussion

Substantial effort has been expendend in characterising the biochemical properties and biological role of RHAU since its early description as a facilitator of uPA mRNA deadenylation and decay in 2004 (32). Since then, RHAU was found to be essential for embryogenesis and haematopoiesis in mouse (90), to sense microbial nucleic acids in human dendritic cells (281,341) and to mediate transcription of tissue-nonspecific alkaline phosphatase in osteoblasts (TNAP, ref. 342). Subcellular localisation studies identified RHAU as a nucleocytoplasmic shuttling protein with a predominant nuclear localisation (32,269). However, RHAU was also found to accumulate in cytoplasmic stress granules upon stress-induced translational inhibition (268). Thus, it becomes clear that RHAU harbours multiple facets taking part in various aspects of nucleic acid metabolism and may show an apparent eclectic diversity in the recognition of nucleic acid structures.

In the present study, a great deal of attention has been expended in examining the G4 binding and G4 resolving activities of RHAU. Actually, when this research work was initiated, not much was known about the biochemical properties of RHAU. Unlike most DEAH-box proteins, RHAU did not manifest unwinding activity on ordinary dsDNA or dsRNA substrates (Tran, H., unpublished data). Besides, although RHAU had been identified as an ARE-associated factor of uPA mRNA, it could not bind the ARE^{uPA} sequence directly, but required the presence of the auxiliary ARE-binding protein NFAR1/NF90 as a cofactor (Akimitsu, N., unpublished data). Thus, G4 nucleic acid structures constituted an ideal substrate to characterise the enzymatic activity of RHAU *in vitro*. Of importance, the finding that RHAU binds an intramolecular G4 structure in TERC has later provided proof-of-concept that G4-RNA structures serves as physiologically relevant target for RHAU in cells. Even so, we still lack a good understanding of the actual contribution of RHAU in resolving G4 structures in cells and about the fraction of RHAU that is dedicated to that task. A valuable clue to the first question came from immunodepletion experiments identifying RHAU as the main source of

tetramolecular G4 resolving activity in HeLa cell lysates (33,89). As for the second query, our RIP-chip analysis revealed a preferential association of RHAU with RNAs bearing PQS, suggesting that RHAU may target other RNAs containing G4 structures in cells. Nonetheless, these data need to be interpreted with caution since the presently available G4 prediction algorithms cannot reliably predict the likelihood of G4 formation in cellular RNAs. Notably, these predictions tools do not allow for the presence of nearby competing secondary structures and/or the cellular context of RNAs bound to proteins. Therefore, only systematic and empirical investigations will permit to confirm the presence of extra G4 structures in RHAU RNA targets.

Linking up RHAU-dependent phenotypes with the processing of G4 nucleic acid structures

Possible reduction of G4-induced inhibition of gene expression by RHAU

At present, there is admittly quite limited functional data underpinning a model in which RHAU would play a decisive role in regulating G4 in cells. Formation of G4 structures in promoters and 5'-UTRs is well documented to hinder gene expression to various extents depending on the position and stability of the G4 (207,316). Thus, one would assume that a reduction of RHAU levels in cells would alter expression of genes bearing PQS. However, thus far, only contrasting results were obtained from computational search for PQS of deregulated genes in RHAU-deficient cells. Transcriptome analysis of RHAU knockout proerythroblast unveiled a higher prevalence of PQS in promoters of genes showing differential expression patterns (90), while no significant difference was previously observed in HeLa cells in which RHAU expression was dampened by RNA interference (269). It should also be stressed that we have not yet observed any suppressive action of RHAU on translational repression by G4 formed in 5'-UTRs (343). Nonetheless, RHAU may instead be implicated in some other aspects of nucleic acid metabolism since G4 structures have also been shown to affect subcellular sorting of mRNAs (344), transcription termination and polyadenylation (218,219), pre-mRNA splicing and mRNA turnover (213,215,216). Besides, overall, RHAU appears to be facultative for most fundamental cellular processes since knockdown of RHAU does not affect the viability or the proliferation of many common human and mouse cell lines when cultured under standard conditions. On the other hand, the penetrance of the RHAU knockout phenotype in haematopoitic cells is variable in severity, time of onset, and lineages concerned. Altogether, these data rather indicate that RHAU is mostly required at some specific stages that control cell lineage commitment. This would be at least consistent with the observations that PQS are more prevalent in genes requiring higher levels of regulatory control (187). Therefore, in order to uncover biologically relevant RNA targets of RHAU, one should consider repeating RIP-chip assays using cells presenting more pronounced phenotypes upon altered RHAU expression.

Implication of RHAU in mouse telomeres homeostasis

We have not yet examined a possible and direct contribution of RHAU in telomere homeostasis in other models than human cell lines. However, it turns out that the murine transcript of *Terc* is substantially shorter than its human paralogue and cannot adopt a G4 conformation at its 5'-end (Figure 36). As a consequence, we can almost rule out the possibility that the abnormal progression of gastrulation and haematopoiesis in RHAU knockout mice arises

because of reduced accumulation of mature *Terc* transcripts. Beyond that, mice are generally not considered as a model of choice to address telomere erosion and its implications on cellular senescence. Indeed, mouse telomeres are about 5 to 10 times longer than their equivalents in human cells. Thus, telomere erosion phenotypes arise only after several generations and usually animals do not manifest the habitual symptoms of human telomeropathies (e.g. bone marrow failure, hepatic cirrhosis, pulmonary fibrosis). In consequence, RHAU knockout mice and murine cell lines in general cannot serve as appropriate models to characterise the interaction of human RHAU with telomerase holoenzyme.

Does RHAU function as an intracellular sensor of microbial G4 structures?

In a recent study, Kim and co-workers suggested that RHAU function as an intracellular sensor of microbial DNA in plasmacytoid dendritic cells (281). The authors proposed that RHAU senses viral CpG motifs (unmethylated CG dinucleotides in bacterial and viral genomic DNA) and promotes the expression of the anti-viral cytokine IFN-α via the MyD88/IRF-7 signalling pathway (Figure 41). However, one interesting but uninvestigated aspect of this study is that RHAU specifically recognises class A CpG oligodeoxynucleotides (CpG-A ODN) but not class B. CpG-A ODNs differ from class B in that they bear poly(dG) sequences at their extremities (345,346). It was previously shown that CpG-A ODNs self-assemble to higher order nanoparticles via oligomerisation of their poly(dG) tails into tetramolecular G4 structures. Of importance, this G4-dependent oligomerisation process is prerequisite to trigger the IFN-α response in plasmacytoid dendritic cells (347). Therefore, it might be possible that contrary to Kim and colleagues' model RHAU recognises the G4 scaffolds formed by CpG-A ODNs rather than the oligodeoxynucleotide sequence alone. Indeed, previous *in vitro* binding competition studies showed that RHAU does not bind specifically to ssDNA sequences and present much higher affinity and specificity for G4-DNA structures (33). In order to consolidate their hypothesis, Kim and colleagues challenged plasmacytoid dendritic cells with Herpes simplex (CpG-rich dsDNA virus) or Influenza A (ssRNA virus employed as a control) viruses and showed that downregulation of RHAU reduced the production of IFN-α upon infection with Herpes simplex but not upon infection with Influenza A. However, the two viruses do not only differ by the biochemical nature of their genetic material. Herpes simplex virus genome presents notably a very high density of PQS with approximately 500 potentially stable G4 forming motifs (Table VI). In contrast, both positive- and negative-sense RNA strands of Influenza A genomic material are devoid of potentially stable G4 forming sequences.

Table VI | **Computational analysis of potential intramolecular G4-forming sequences (PQS) in human Herpes simplex 1 (HSV-1) and Influenzy A (H1N1) viruses.** The $N_{G4}(2)$–$N_{G4}(5)$ fields denote the number of predicted G4 composed of 2, 3, 4 or 5 successive G-quartets, respectively.

		HSV-1		Influenza A (H1N1)	
		Temp. strand	Comp. strand	Temp. strand	Comp. strand
Non-overlapping PQS	$N_{G4}(2)$	1'678	1'914	49	12
	$N_{G4}(3)$	210	227	0	0
	$N_{G4}(4)$	32	27	0	0
	$N_{G4}(5)$	1	2	0	0
G+C content		68 %		43 %	
Genome		152'261 bp (dsDNA)		13'588 nt (8× neg.-sense ssRNAs)	

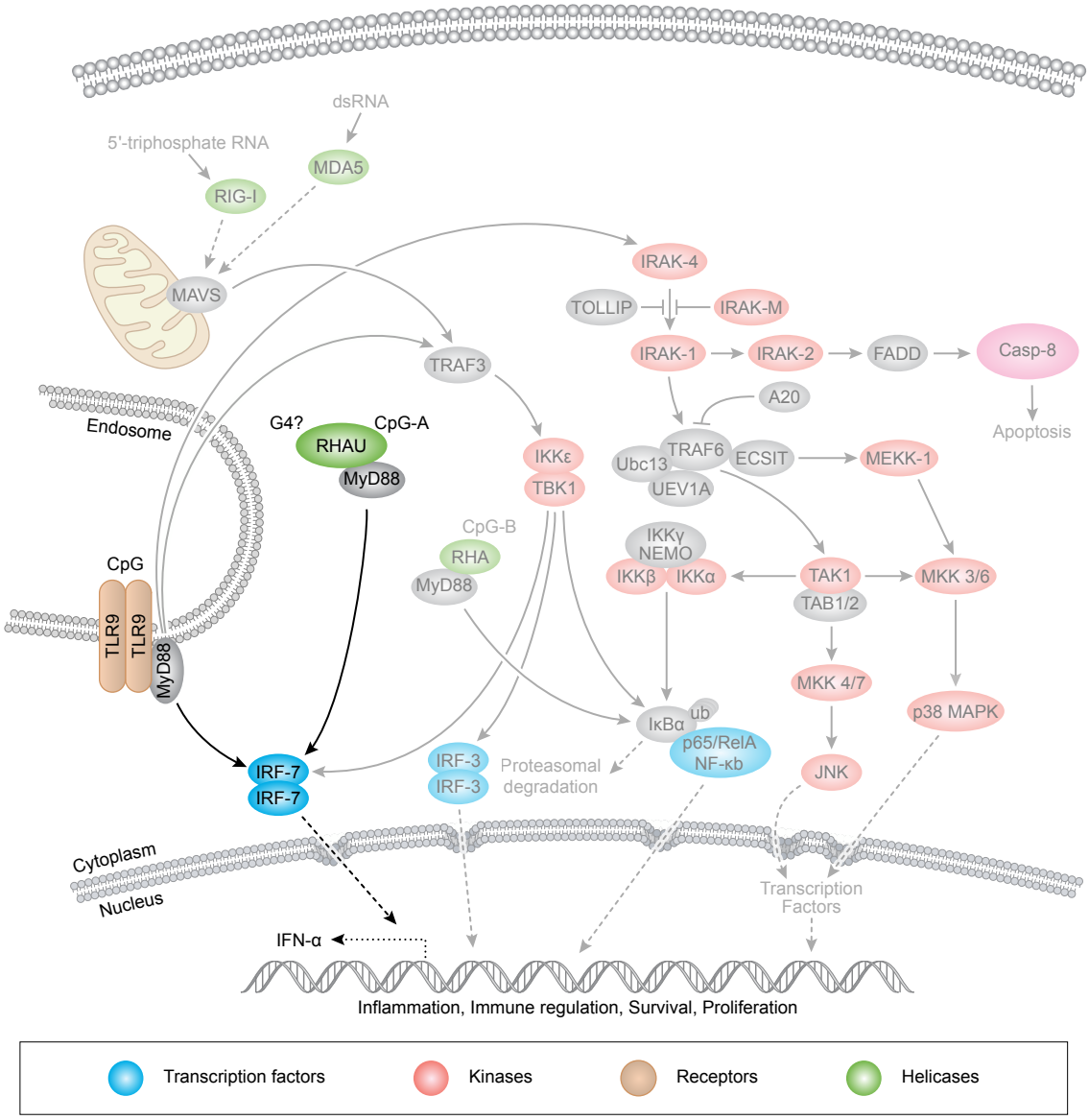


Figure 41 | **TLR9-dependant and -independent detection of microbial CpG motifs**. According to the classical pathway, microbial DNA containing unmethylated CpG motifs are specifically recognised in endosomes by Toll-like receptor 9 (TLR9). Activation of TLR9 pathway induces the secretion of the antiproliferative cytokines type I interferons (IFN-α and IFN-β). In plasmacytoid dentritic cells, TLR9-independent recognition of viral cytoplasmic CpG motifs may occur through the DEAH-box proteins RHAU and RHA (281). RHA specifically recognises class B CpG oligodeoxynucleotides (CpG-B) and promotes the production of both TNF-α and IL-6 through the activation of the NF-κB pathway. RHAU instead senses class A CpGs (CpG-A) and triggers the expression of IFN-α via the MyD88/IRF-7 signalling pathway. Adapted from the 'Toll-like Receptor Signaling' pathway (source: Cell Singaling Technology, Inc., version Nov. 2010).

Although speculative at this point, the idea that RHAU functions as a cytoplasmic sensor of microbial G4 structures may be of relevance given its very high specificity for G4 structures. Implicitly, this would also indicate that microbial G4 structures define a novel category of pathogen-associated molecular patterns (PAMPs). Cells recognise PAMPs by means of the so-called 'pattern recognition receptors' (PRRs, e.g. Toll-like receptors) which are primitive and evolutionary conserved factors of innate cell immunity. We have previously shown that RHAU is conserved from choanoflagelates to humans (268) and that it retains the ability to bind G4 structures in higher eukaryotes. Thus, one may speculate that RHAU shares a similar G4 sensing function in other organisms, a property which in turn would be consistent with the evolutionary conserved aspect of PRRs. In support of our model, several helicases such as those from the RIG-I-like family (RIG-I, MDA-5, LGP2 and Dicer) have been shown to be

implicated in sensing pathogenic nucleic acids in cells (ref. 285,286; Figure 41). Thus, additional analyses are necessary to investigate the potential role of RHAU in detecting pathogen-associated G4 structures. However, RHAU could emerge as a promising novel component of the innate immunity.

Current status and future prospects

Structural insight into the recognition of G4 stuctures by RHAU

G4 structures constitute a family with extremely heterogeneous structural variations regarding the length and the conformation of the quadruplex stem (intra-, bi- or tetramolecular structures; parallel, antiparallel or mixed configurations; *syn* or *anti* glycosidic conformations of guanines), as well as the sequences, length and orientation of the intervening loops. Computational surveys identified about 300'000 and 200'000 PQS in human genome and transcriptome, respectively (186,188,217). Besides, over the past two decades, several proteins have been shown to bind to various synthetic and *bona fide* G4 nucleic acid structures (260,348,349). However, although most of these proteins execute specific tasks in cells, several of them can bind a broad array of G4 conformations when analysed *in vitro*. Actually, not much is know regarding the molecular details about G4–protein interactions and specificity. Currently, only three experimental structures of proteins bound to G4 in a complex are publicly accessible (350-353). Intriguingly, in all three structures the protein moiety makes few, if any, contacts with the quadruplex core. Instead, most of the G4–protein interactions implicate the connecting loops and/or the protruding strands of the G4. It was suggested that from a thermodynamic standpoint, the interaction of proteins with constrained loops, rather than random coils, is favoured due to the reduced entropic binding cost (176,354). However, given the current small size of the structural data, it is somewhat premature to establish common structural principles on G4 binding specificity. Likewise, comparative analyses of various G4 recognition motifs in proteins show that it is currently not possible to establish consensual sequences or motifs involved specifically in the recognition of G4 structures. Notably, some G4 binding proteins bind their targets through canonical RNA-binding motifs such as RGG (FMRP, ref. 254,353) or RRM (nucleolin, ref. 355) motifs, while other proteins resort to more family-restricted motifs such as the RQC domain (RecQ-like helicases, ref. 302) or the RSM (RHAU).

The data presented here together with previous studies show that *in vitro*, RHAU can bind and unwind with almost the same efficiency many different types of tetra- (TP, Z33, rD4 and AGA) and intra- (c-Myc, Zic1, YY1 and TERC) molecular G4 structures (33,39,89,343). This suggests that, at least *in vitro*, no particular sequence element in the overhang strand and/or in the connecting loops is essential for the recognition of G4 structures by RHAU. It emerges also that the connecting loops are likely to play an incidental role since RHAU efficiently binds tetramolecular G4 structures which are devoid of intervening sequence elements. The presence of a minimal 9-nt 3'-overhang strand is currently the only known structural requirement for binding and resolving G4 structures by RHAU. Similar requirements hold for WRN and BLM helicases (356), and it is assumed that the single-stranded extremity is bound by the helicase core region and permits the 3'-to-5' tranlocation of the catalytic unit towards the G4 structure. In this regard, both experimental and computational data are consistent and show that the helicase core region alone cannot account for the recognition of G4 structures by RHAU. Besides, we could reproducibly show both *in vitro* and *in vivo* that the N-terminal RSM domain makes a decisive contribution in the G4 binding process. However,

further structural and biochemical experiments are necessary to provide a more detailed model of the binding mechanism. Indeed, our mutational analysis of the RSM domain did not permit to truly pinpoint the residues interacting with the nucleic acid structure since the G4 binding deficiency for the RSM mutated forms of RHAU was systematically less severe than for the RSM deleted form. Thus, it might be possible that some of the contacts to the nucleic acid require only the peptide backbone of the RSM or that other less conserved residues take part in the interaction. Furthermore, truncation of the catalytic unit showed that although necessary, the N-terminal region alone does not fully account for the high affinity for G4 structure of the whole protein. We had previously suggested that substantial additional binding activity could also stem from the helicase core region through its interaction with the 3'-overhang strand (88). However, it is also conceivable that the N-terminal region can form a functional and high affinity G4 binding domain only in the presence of the adjacent catalytic core region. In any cases, this is an important point that needs to be considered for the design of future experimental structural analyses of RHAU.

Targetting RHAU and TERC as therapeutic approaches against cancers

Following the discovery that telomerase activity *in vitro* could be hindered by formation of G4 structures (184), stabilisation of G4 by means of small molecule ligands has emerged as a promising anticancer chemotherapeutic approach. Indeed, a large majority (80–90%) of tumor cells overexpress this enzyme while most of the somatic cells harbour low, if any, levels of telomerase activity (357,358). Since this finding, considerable efforts have been investigated in the design and development of small molecule compounds selective for G4 structures and presenting minimal affinity for dsDNA (359-361). Although several G4 ligands demonstrated efficient antitumor activities, it appeared actually that only a few compounds truly interfered with telomerase function. In fact, most of these ligands caused unspecific short-term cellular cytotoxicity by altering the integrity of the telomere capping complex or interfered with other unrelated cellular processes (193,362). To date, none of these compounds has yet proven to be therapeutically effective and most of them did not progress beyond the experimental stage. Currently, one of the main challenges lies in developing G4 ligand compounds presenting high selectivity for a particular G4 conformation in order to minimise off-target effects on other biologically relevant G4 structures. One such approach was recently undertaken by Mergny and co-workers that screened a library of known G4 ligands for compounds presenting high selectivity for the G4 structure formed at the 5'-extremity of human TERC (363). Part of the rationale was that extra-stabilisation of the G4 conformation of TERC should hinder telomerase activity by preventing P1 helix formation (Figure 37). Although the authors of this study identified a few ligands manifesting higher selectivity for TERC G4 structure over various telomeric and promoter G4-DNAs, it turned out that the differential affinity was not sufficient to harness them as potentially selective chemotherapeutic agents. Thus much effort needs to be expended in developing even more selective G4 ligands. However this is not a trivial task considering the incidence and the broad polymorphism of G4 structures in human genome.

Alternatively, G4-based aptamers may offer a potentially attractive approach to interfere with some biological processes involving G4 binding proteins. G4-based aptamers may be employed as competitive reversible inhibitors of G4 binding and resolving proteins. Indeed, several *in vitro* selected G-rich oligonucleotides folds into G4 structures and harbour promising antiproliferative properties on various cancer cells (364,365). Thus, the selective

inhibition of RHAU G4 binding and resolving activities by means of *in vitro* selected G4 aptamers might be an appealing alternative solution to prevent P1 helix formation in TERC. However, for now, it is above all fundamental to clarify the modalities that caused excessive telomere elongation in cells deficient for RHAU. Otherwise, inhibition of RHAU activity may not induce the expected antiproliferative effect on cancer cells.

It might be also worth to further explore the functional crosstalk between RHAU and TLR9 pathways that both lead to the production of the antiproliferative IFN-α cytokine (Figure 41; ref. 281). TLR9 agonists have indeed demonstrated potential as anticancer chemotherapeutics and are also used as vaccine adjuvants (366). As antineoplastic agents, TLR9 agonists are assumed to promote the activation and maturation of plasmacytoid dendritic cells and enhance the differentiation of B cells into antibody-secreting plasma cells, potentially promoting Ig-dependent cellular cytotoxicity against tumor cells. To date, several pharmaceutical and biotechnology companies undertook to develop synthetic CpG ODN agonists for TLR9 as a novel immunotherapeutic approach to the treatment of cancers. However, synthetic ODNs need to be chemically modified or encapsulated into virus-like particles to limit their degradation by endonucleases. Assuming that RHAU support sensing of G4-DNA in cytoplasm, one may count on comparable antiproliferative effects by treating plasmacytoid dendritic cells or B cells with *in vitro* selected G4 ODNs. Compared to classical CpG ODNs, G4-based nucleic acids are usually much more resistant against endonucleolytic cleavage which may improve the intrinsic bioavailability. Thus, targeting RHAU by means of G4 ODNs may open new perspectives for the treatment of infectious diseases, allergies and possibly cancers.

Together with more recent findings that RHAU is required for the survival of *ras*-transformed cells (Lai, J.C., unpublished data), the observations presented here suggest that RHAU may represent a potentially interesting target for cancer therapy. However, one has to keep in mind that altered expression of RHAU in healthy tissues can also produce severe phenotypes such as the strong anaemia (90) or the neuromuscular paralytic disorders (Lai, J.C. and Wu, P., unpublished data) observed in RHAU knockout mice. Thus, further experimentations are necessary to establish whether molecular targeted treatments specific to RHAU offer a sufficiently wide therapeutic window with both specific and efficient curative effects.

Concluding remarks

Together with our previous findings that RHAU binds and exhibits robust resolvase activity on various types of G4 structures, the data presented here bring forth the idea that G4-RNAs and especially intramolecular G4-RNAs serve as physiologically relevant targets for RHAU in cells. We notably showed that RHAU binds a G4 structure in the 5'-region of TERC and suggest that RHAU catalyses the conversion of the G4 scaffold to ssRNA, thus allowing formation of helix P1. In addition, we established that the N-terminal RSM domain is a major affinity and specificity determinant for the recognition of G4-RNA structures by RHAU. This finding may open up new avenues for exploring more extensively the concrete contribution of RHAU in G4 nucleic acid metabolism. The development of a knock-in mouse model of RHAU carrier of a non-functional RSM domain would be one way to study its G4 binding and resolving properties in a more physiologically relevant context and may allow to identify potentially new RNA targets. Identification of naturally occurring substrates of RNA helicases is a crucial step to any further investigation of their biochemical

properties *in vitro*. However, although most of the RNA helicases achieve highly specific tasks *in vivo*, they often show only little or even no nucleic acid binding specificity *in vitro*. Currently, our major limitation towards understanding the mechanisms whereby RNA helicases melt nucleic acid structures stem from the difficulty of identifying such physiologically relevant substrates. Hence, few detailed molecular models available for studying RNA helicases *in vitro*. However, with the finding that the 5' region of TERC constitutes a biologically relevant substrate, RHAU emerges as a novel and promising prototype of DEAH-box protein and deserves more investigations to explore its functions as a G4 resolving enzyme.

CHAPTER 5

Acknowledgments

I first would like to thank the principal investigator Dr. Yoshikuni Nagamine for giving me the opportunity to work in his lab and for all the freedom he granted me on the project. I would like also to acknowledge all the members of my thesis committee: Prof. Susan Gasser, Prof. Christoph Moroni and Prof. Dagmar Klostermeier for their availability for discussion and their valuable suggestions. I am grateful to Prof. Ruth Chiquet-Ehrismann for chairing the thesis committee.

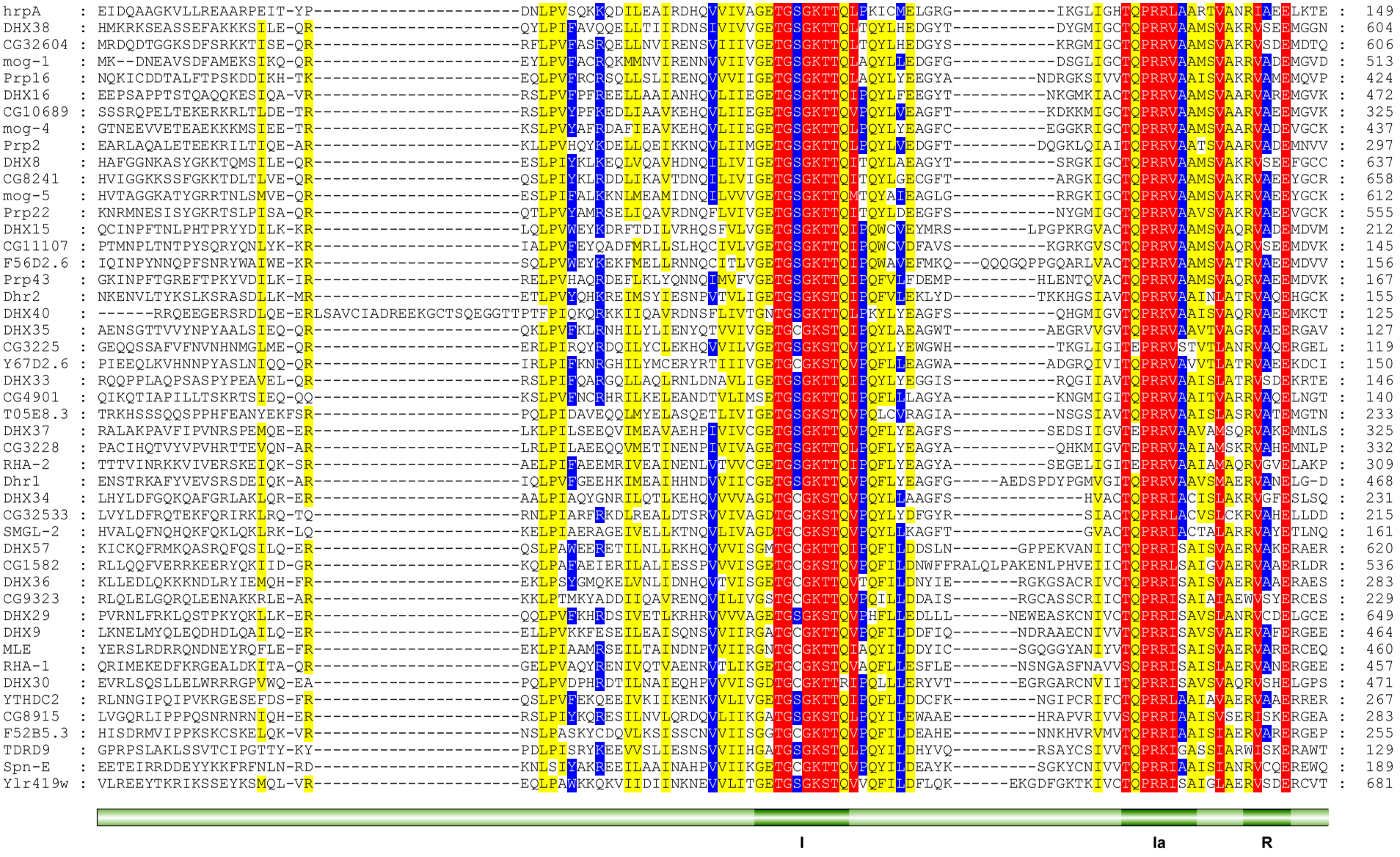
I would like also to acknowledge all the former members of the Nagamine's lab: Fumiko, Janice, Katerina, Nives, Sandra and Stéphane for their precious help and for the great working environment. I am also grateful to several members of the technology platforms at the FMI for their contributions to the project: Michael Stadler (bioinformatics facility), Hubertus Kohler (FACS facility), Edward Oakeley (functional genomics facility) and Dominique Klein (protein analysis facility).

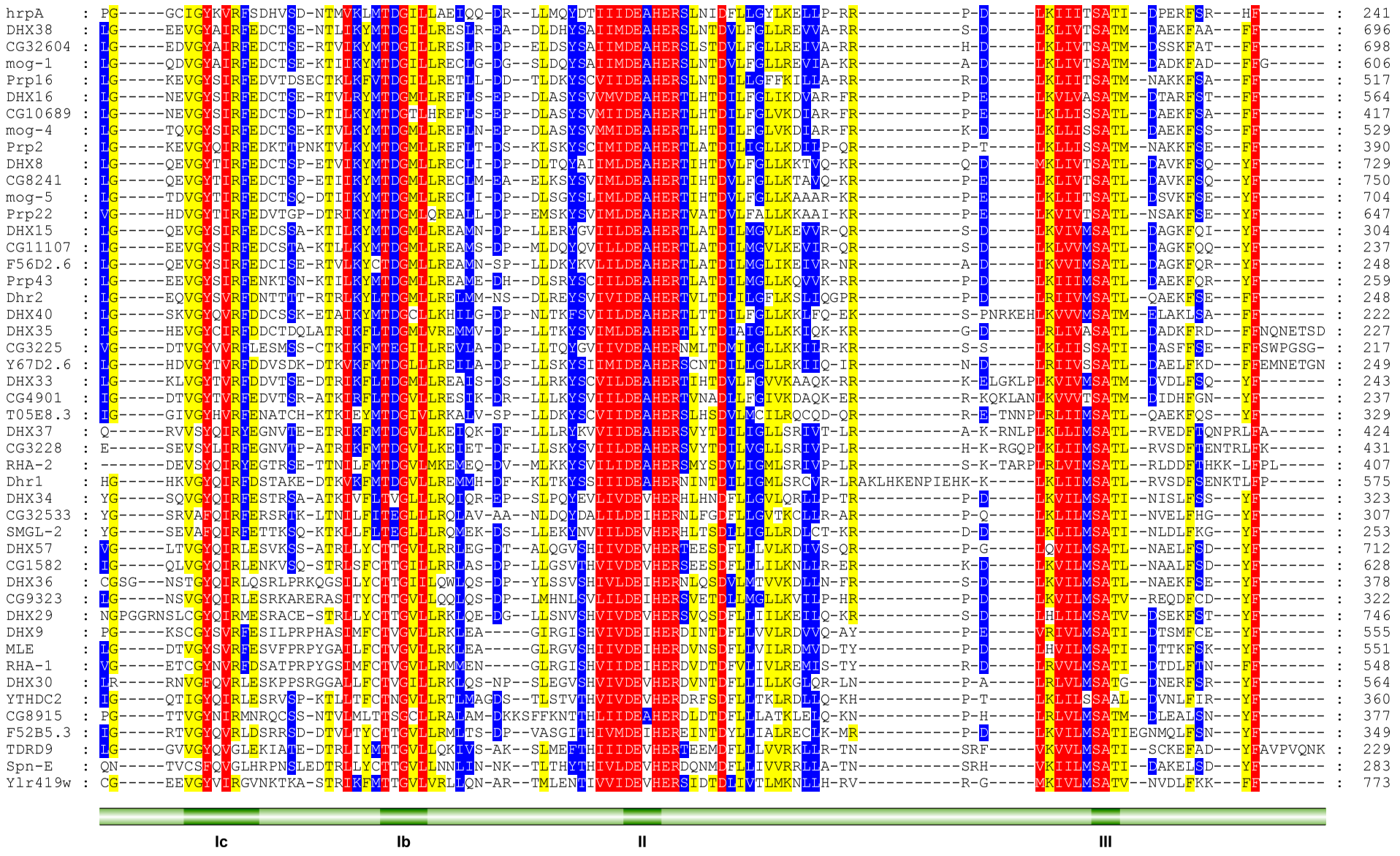
Finally, I would like to specially thank Fanny, Stéphane, Séverine and Fabien for their friendship, their never ending encouragements as well as for the nice moments spent together.

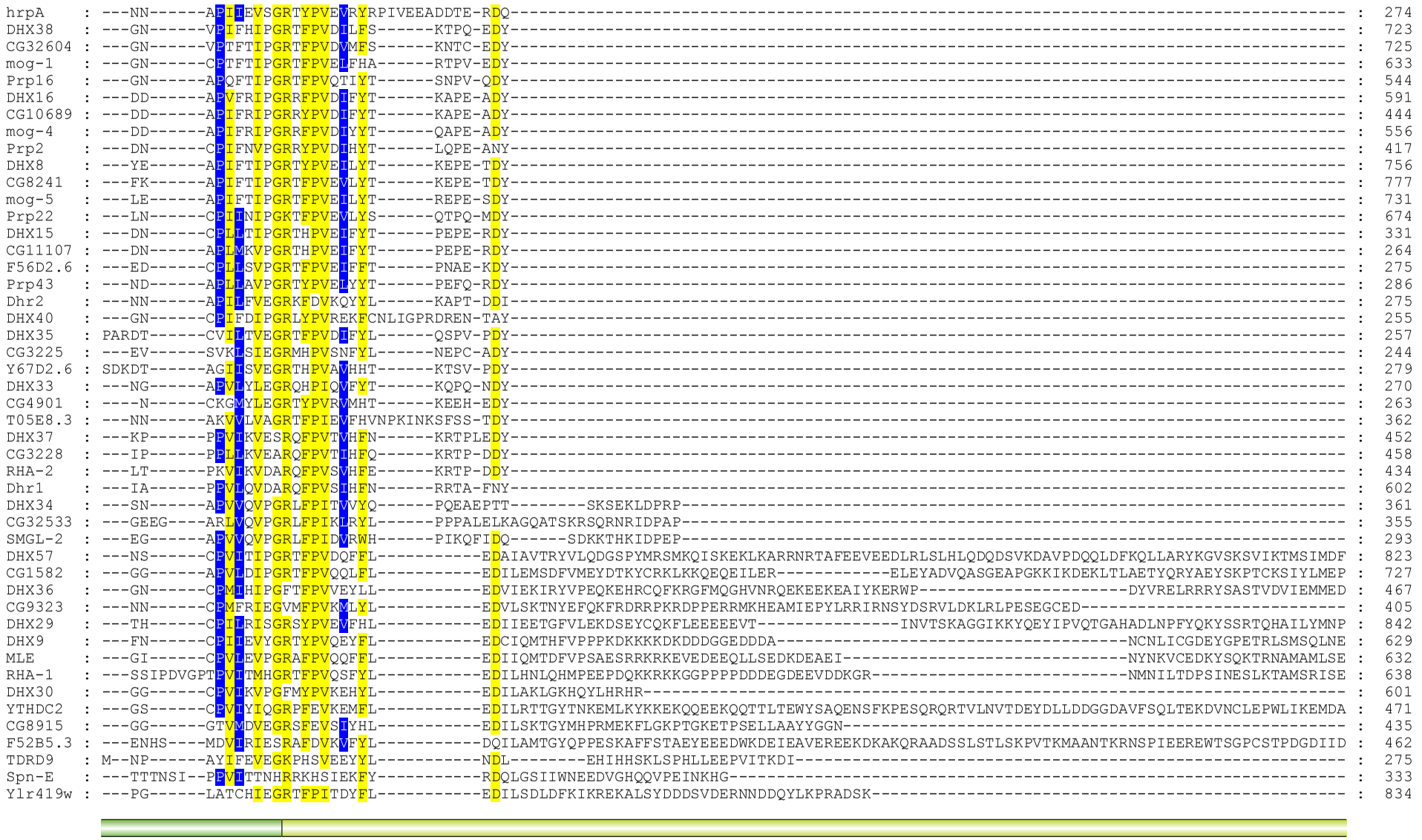
CHAPTER 6

Supplementary material

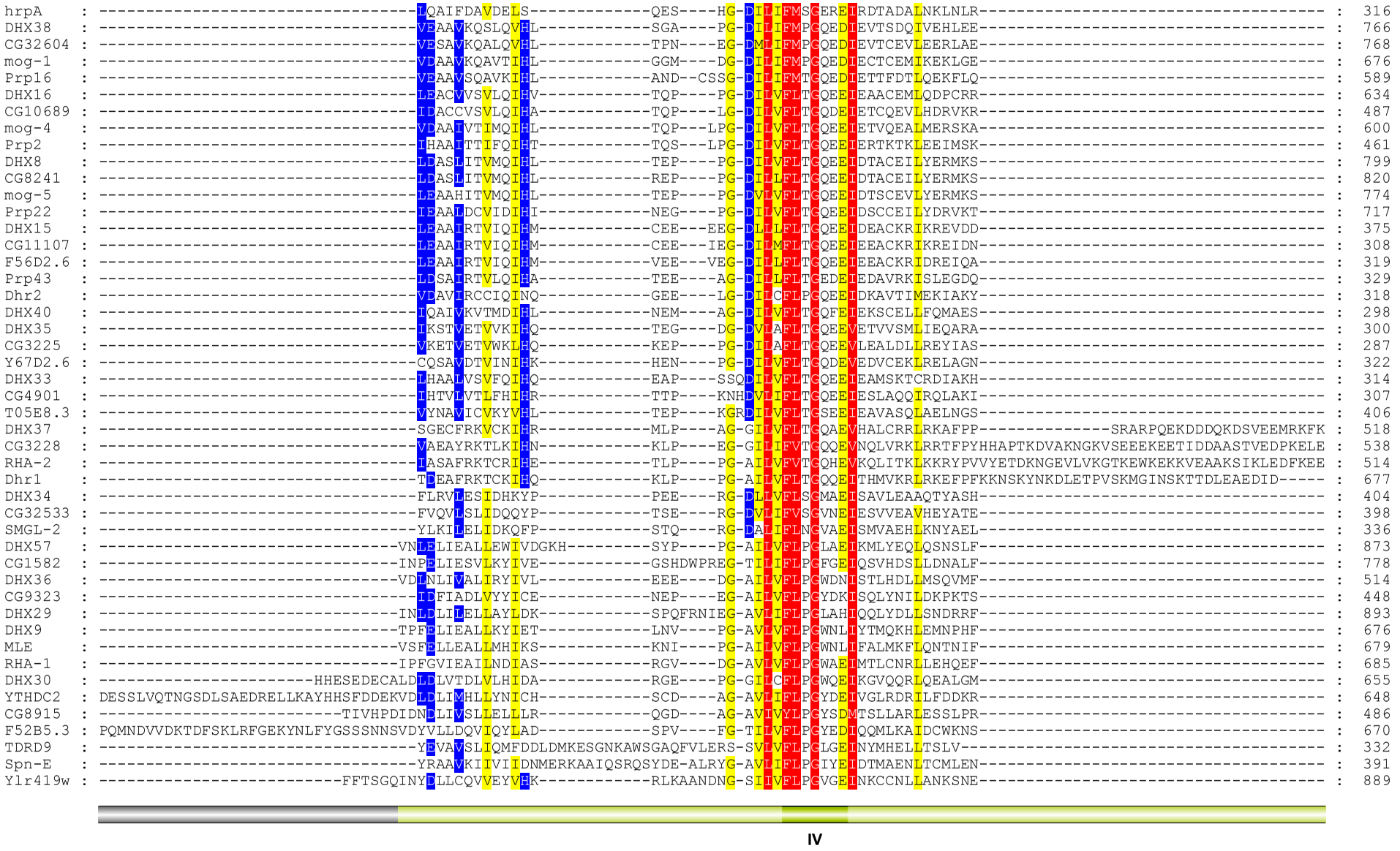
Supplementary Figure 1 ► | **Amino acid sequence conservation among paralogous DEAH-box proteins** from *E. coli*, *S. cerevisiae*, *C. elegans*, *D. melanogaster* and *H. sapiens*. Sequence alignment was carried out with MAFFT (version 6, ref. 313). Similarity analysis was performed with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Similarity is shown in red for 100 %, yellow for 99–80 % and blue for 79–60 %. The position of five structural domains (RecA1, RecA2, WH, Ratchet and OB-fold) as well as the conserved ATPase/helicase motifs I–VI is shown as coloured boxes beneath the sequence alignment.


I
Ia
R

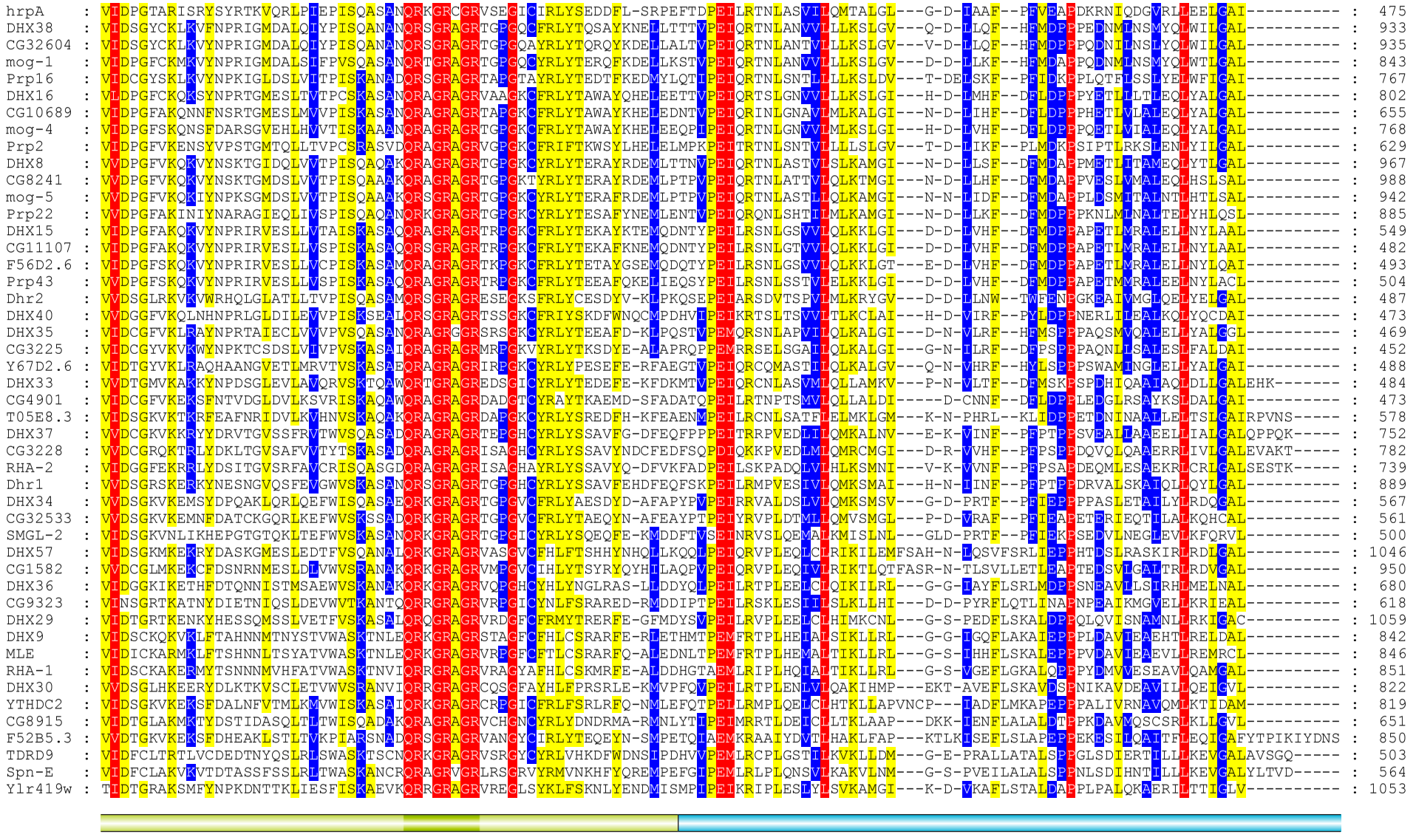

hrpA : PG-----GCIGYKVRFSDHVSD-NTMVKLMTDGILLAEIQQ-DR--LLMQYDTIIIDEAHERSLNIDFLLGYLKELLP-RR----------P-D-----LKIITSATI--DPERFSR---HF------- : 241
DHX38 : LG-----EEVGYAIRFEDCTSE-NTLIKYMTDGILLRESLR-EA--DLDHYSAIIMDEAHERSLNTDVLFGLLREVVA-RR----------S-D-----LKLIVTSATM--DAEKFAA---FF------- : 696
CG32604 : LG-----EDVGYAIRFEDCTSE-RTVIKYMTDGILLRESLR-DP--ELDSYSAIIMDEAHERSLSTDVLFGLLREIVA-RR----------H-D-----LKLIVTSATM--DSSKFAT---FF------- : 698
mog-1 : LG-----QDVGYAIRFEDCTSE-KTIIKYMTDGILLRECLG-DG--SLDQYSAIIMDEAHERSLNTDVLFGLLREVIA-KR----------A-D-----LKLIVTSATM--DADKFAD---FFG------- : 606
Prp16 : LG-----KEVGYSIRFEDVTDSECTKLKFVTDGILLRETLL-DD--TLDKYSCVIIDEAHERSLNTDILLGFFKILLA-RR----------R-D-----LKLIITSATM--NAKKFSA---FF------- : 517
DHX16 : LG-----NEVGYSIRFEDCTSE-RTVLRYMTDGMLLREFLS-EP--DLASYSVVMVDEAHERTLHTDILFGLIKDVAR-FR----------P-E-----LKVLVASATM--DTARFST---FF------- : 564
CG10689 : LG-----NEVGYSIRFEDCTSD-RTILKYMTDGTLHREFLS-EP--DLASYSVMIIDEAHERTLHTDILFGLVKDIAR-FR----------P-E-----LKLLISSATL--DAEKFSA---FF------- : 417
mog-4 : LG-----TQVGYSIRFEDCTSE-KTVLKYMTDGMLLREFLN-EP--DLASYSVMMIDEAHERTLHTDILFGLVKDIAR-FR----------K-D-----LKLLISSATL--DAEKFSS---FF------- : 529
Prp2 : LG-----KEVGYQIRFEDKTTPNKTVLKYMTDGMLLREFLT-DS--KLSKYSCIMIDEAHERTLATDILIGLLKDILP-QR----------P-T-----LKLLISSATM--NAKKFSE---FF------- : 390
DHX8 : LG-----QEVGYTIRFEDCTSP-ETVIKYMTDGMLLRECLI-DP--DLTQYAIIMLDEAHERTIHTDVLFGLLKKTVQ-KR----------Q-D-----MKLIVTSATL--DAVKFSQ---YF------- : 729
CG8241 : LG-----QEVGYTIRFEDCTSP-ETIIKYMTDGMLLRECLM-EA--ELKSYSVIMLDEAHERTIHTDVLFGLLKTAVQ-KR----------P-E-----LKLIVTSATL--DAVKFSQ---YF------- : 750
mog-5 : LG-----TDVGYTIRFEDCTSQ-DTIIKYMTDGMLLRECLI-DP--DLSGYSLIMLDEAHERTIHTDVLFGLLKAAAR-KR----------P-E-----LKLITTSATL--DSVKFSE---YF------- : 704
Prp22 : VG-----HDVGYTIRFEDVTGP-DTRIKYMTDGMLQREALL-DP--EMSKYSVIMLDEAHERTVATDVLFALLKKAAI-KR----------P-E-----LKVIVTSATL--NSAKFSE---YF------- : 647
DHX15 : LG-----QEVGYSIRFEDCSSA-KTILKYMTDGMLLREAMN-DP--LLERYGVIILDEAHERTLATDILMGVLKEVVR-QR----------S-D-----LKVIVMSATL--DAGKFQI---YF------- : 304
CG11107 : LG-----EEVGYSIRFEDCSTA-KTLLKYMTDGMLLREAMS-DP--MLDQYQVIILDEAHERTLATDILMGVLKEVIR-QR----------S-D-----LKLVVMSATL--DAGKFQQ---YF------- : 237
F56D2.6 : LG-----QEVGYSIRFEDCISE-RTVLKYCTDGMLLREAMN-SP--LLDKYKVLILDEAHERTLATDILMGLIKEIVR-NR----------A-D-----IKVVIMSATL--DAGKFQR---YF------- : 248
Prp43 : LG-----EEVGYSIRFENKTSN-KTILKYMTDGMLLREAME-DH--DLSRYSCIILDEAHERTLATDILMGLLKQVVK-RR----------P-D-----LKIIIMSATL--DAEKFQR---YF------- : 259
Dhr2 : LG-----EQVGYSVRFDNTTTT-RTRLKYLTDGMLLRELMM-NS--DLREYSVIVIDEAHERTVLTDILLGFLKSLIQGPR----------P-D-----LRIIVMSATL--QAEKFSE---FF------- : 248
DHX40 : LG-----SKVGYQVRFDDCSSK-ETAIKYMTDGCLLKHILG-DP--NLTKFSVIILDEAHERTLTTDILFGLLKKLFQ-EK----------S-PNRKEHLKVVVMSATM--ELAKLSA---FF------- : 222
DHX35 : LG-----HEVGYCIRFDDCTDQLATRIKFLTDGMLVREMMV-DP--LLTKYSVIMLDEAHERTLYTDIAIGLLKKIQK-KR----------G-D-----LRLIVASATL--DADKFRD---FFNQNETSD : 227
CG3225 : VG-----DTVGYVVRFLESMSS-CTKIKFMTEGILLREVLA-DP--LLTQYGVIIVDEAHERNMLTDVLIGLLKKIIR-KR----------S-S-----LKLIISSATI--DASFFSE---FFSWPGSG- : 217
Y67D2.6 : LG-----HDVGYTVRFDDVSDK-DTKVKFMTDGLLLRETLA-DP--LLSKYSIIMIDEAHERSCNTDILLGLLRKIIQ-IR----------N-D-----LRIIVSSATL--DAELFKD---FFEMNETGN : 249
DHX33 : LG-----KLVGYTVRFDDVTSE-DTRIKFLTDGMLLREAIS-DS--LLRKYSCVILDEAHERTIHTDVLFGVVKAAQK-RR----------K-ELGKLPLKVIVMSATM--DVDLFSQ---YF------- : 243
CG4901 : IG-----DTVGYTVRFEDVTSR-ATKIRFLTDGVLLRESIK-DR--LLLKYSVIILDEAHERTVNADLLFGIVKDAQK-ER----------R-KQKLANLKVVVTSATM--DIDHFGN---YF------- : 237
T05E8.3 : LG-----GIVGYHVRFENATCH-KTKIEYMTDGIVLRKALV-SP--LLDKYSCVIIDEAHERSLHSDVLMCILRQCQD-QR----------R-E-TNNPLRLIIMSATL--QAEKFQS---YF------- : 329
DHX37 : Q------RVVSYQIRYEGNVTE-ETRIKFMTDGVLLKEIQK-DF--LLLRYKVVIIDEAHERSVYTDILIGLLSRIVT-LR----------A-K-RNLLPLKLLIMSATL--RVEDFTQNPRLFA------ : 424
CG3228 : E------SEVSYLIRFEGNVTP-ATRIKFMTDGVLLKEIET-DF--LLSKYSVIILDEAHERSVYTDILVGLLSRIVP-LR----------H-K-RGQPLKLLIMSATL--RVSDFTENTRLFK------ : 431
RHA-2 : -------DEVSYQIRYEGTRSE-TTNILFMTDGVLMKEMEQ-DV--MLKKYSVILIDEAHERSMYSDVLIGMLSRIVP-LR----------S-K-TARPLRLVIMSATL--RLDDFTHKK-LFPL----- : 407
Dhr1 : HG-----HKVGYQIRFDSTAKE-DTKVKFMTDGVLLREMMH-DF--KLTKYSSIIIDEAHERNINTDILIGMLSRCVR-LRAKLHKENPIEHK-K-----LKLIIMSATL--RVSDFSENKTLFP------ : 575
DHX34 : YG-----SQVGYQIRFESTRSA-ATKIVFLTVGLLLRQIQR-EP--SLPQYEVLIVDEVHERHLHNDFLLGVLQRLLP-TR----------P-D-----LKVILMSATI--NISLFSS---YF------- : 323
CG32533 : YG-----SRVAFQIRFERSRTK-LTNILFITEGLLLRQLAV-AA--NLDQYDALILDEIHERNLFGDFLLGVTKCLLR-AR----------P-Q-----LKLILMSATI--NVELFHG---YF------- : 307
SMGL-2 : YG-----SEVAFQIRFETTKSQ-KTKLLFLTEGLLLRQMEK-DS--LLEKYNVIILDEVHERHLTSDILIGLLRDLCT-KR----------D-D-----LKLILMSATI--NLDLFKG---YF------- : 253
DHX57 : VG-----LTVGYQIRLESVKSS-ATRLLYCTTGVLLRRLEG-DT--ALQGVSHIIVDEVHERTEESDFLLLVLKDIVS-QR----------P-G-----LQVILMSATL--NAELFSD---YF------- : 712
CG1582 : IG-----QLVGYQIRLENKVSQ-STRLSFCTTGILLRRLAS-DP--LLGSVTHVIVDEVHERSEESDFLLLILKNLLR-ER----------K-D-----LKVILMSATL--NAALFSD---YF------- : 628
DHX36 : CGSG---NSTGYQIRLQSRLPRKQGSILYCTTGIILQWLQS-DP--YLSSVSHIVLDEIHERNLQSDVLMTVVKDLLN-FR----------S-D-----LKVILMSATL--NAEKFSE---YF------- : 378
CG9323 : LG-----NSVGYQIRLESRKARERASITYCTTGVLLQQLQS-DP--LMHNLSVLILDEIHERSVETDLLMGLLKVILP-HR----------P-D-----LKVILMSATV--REQDFCD---YF------- : 322
DHX29 : NGPGGRNSLCGYQIRMESRACE-STRLLYCTTGVLLRKLQE-DG--LLSNVSHVIVDEVHERSVQSDFLLIILKEILQ-KR----------S-D-----LHLILMSATV--DSEKFST---YF------- : 746
DHX9 : PG-----KSCGYSVRFESILPRPHASIMFCTVGVLLRKLEA-----GIRGISHVIVDEIHERDINTDFLLVVLRDVVQ-AY----------P-E-----VRIVLMSATI--DTSMFCE---YF------- : 555
MLE : LG-----DTVGYSVRFESVFPRPYGAILFCTVGVLLRKLEA-----GLRGVSHIIVDEIHERDVNSDFLLVILRDMVD-TY----------P-D-----LHVILMSATI--DTTKFSK---YF------- : 551
RHA-1 : VG-----ETCGYNVRFDSATPRPYGSIMFCTVGVLLRMMEN-----GLRGISHVIIDEIHERDVDTDFVLIVLREMIS-TY----------R-D-----LRVVLMSATI--DTDLFTN---FF------- : 548
DHX30 : LR-----RNVGFQVRLESKPPSRGGALLFCTVGILLRKLQS-NP--SLEGVSHVIVDEVHERDVNTDFLLILLKGLQR-LN----------P-A-----LRLVLMSATG--DNERFSR---YF------- : 564
YTHDC2 : LG-----QTIGYQIRLESRVSP-KTLLTFCTNGVLLRTLMAGDS--TLSTVTHVIVDEVHERDRFSDFLLTKLRDLLQ-KH----------P-T-----LKLIISSAAL--DVNLFIR---YF------- : 360
CG8915 : PG-----TTVGYNIRMNRQCSS-NTVLMLTTSGCLLRALAM-DKKSFFKNTTHLIIDEAHERDLDTDFLLLATKLELQ-KN----------P-H-----LRLVLMSATM--DLEALSN---YF------- : 377
F52B5.3 : IG-----RTVGYQVRLDSRRSD-DTVLTYCTTGVLLRMLTS-DP--VASGITHIVMDEIHEREINTDYLLIALRECLK-MR----------P-D-----LKVILMSATIEGNMQLFSN---YF------- : 349
TDRD9 : LG-----GVVGYQVGLEKIATE-DTRLIYMTTGVLLQKIVS-AK--SLMEFTHIIIDEVHERTEEMDFLLLVVRKLLR-TN----------SRF-----VKVVLMSATI--SCKEFAD---YFAVPVQNK : 229
Spn-E : QN-----TVCSFQVGLHRPNSLEDTRLLYCTTGVLLNNLIN-NK--TLTHYTHIVLDEVHERDQNMDFLLIVVRRLLA-TN----------SRH-----VKIILMSATI--DAKELSD---YF------- : 283
Ylr419w : CG-----EEVGYVIRGVNKTKA-STRIKFMTTGVLVRLLQN-AR--TMLENTIVVIDEVHERSIDTDLIVTLMKNLLH-RV----------R-G-----MKIVLMSATV--NVDLFKK---FF------- : 773
Ic
Ib
II
III

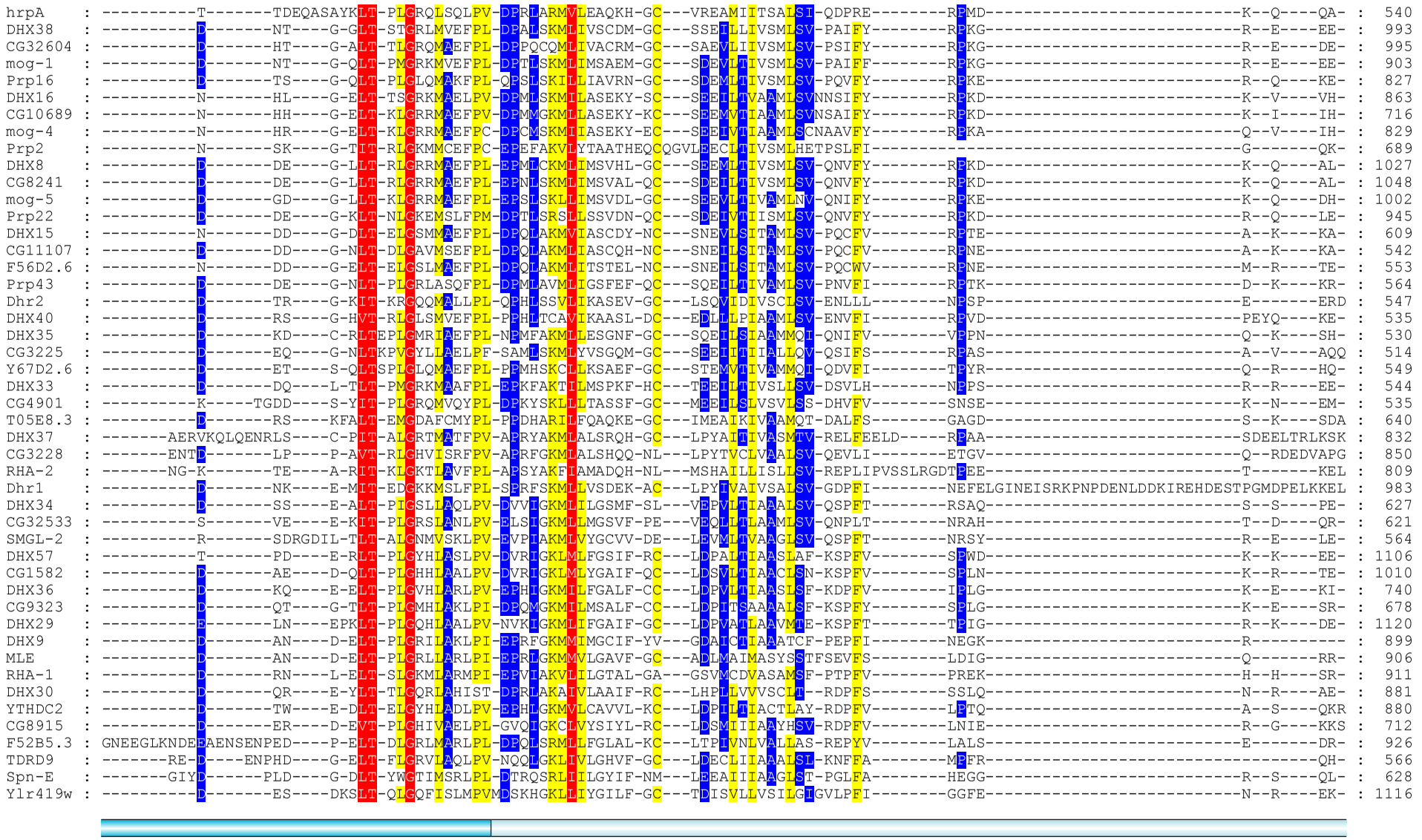

hrpA : ---NN------APIIEVSGRTYPVEVRYRPIVEEADDTE-RDQ---------- : 274
DHX38 : ---GN------VPIFHIPGRTFPVDILFS------KTPQ-EDY---------- : 723
CG32604 : ---GN------VPTFTIPGRTFPVDVMFS------KNTC-EDY---------- : 725
mog-1 : ---GN------CPTFTIPGRTFPVELFHA------RTPV-EDY---------- : 633
Prp16 : ---GN------APQFTIPGRTFPVQTIYT------SNPV-QDY---------- : 544
DHX16 : ---DD------APVFRIPGRRFPVDIFYT------KAPE-ADY---------- : 591
CG10689 : ---DD------APIFRIPGRRYPVDIFYT------KAPE-ADY---------- : 444
mog-4 : ---DD------APIFRIPGRRFPVDIYYT------QAPE-ADY---------- : 556
Prp2 : ---DN------CPIFNVPGRRYPVDIHYT------LQPE-ANY---------- : 417
DHX8 : ---YE------APIFTIPGRTYPVEILYT------KEPE-TDY---------- : 756
CG8241 : ---FK------APIFTIPGRTFPVEVLYT------KEPE-TDY---------- : 777
mog-5 : ---LE------APIFTIPGRTFPVEILYT------REPE-SDY---------- : 731
Prp22 : ---LN------CPILNIPGKTFPVEVLYS------QTPQ-MDY---------- : 674
DHX15 : ---DN------CPLLTIPGRTHPVEIFYT------PEPE-RDY---------- : 331
CG11107 : ---DN------APLMKVPGRTHPVEIFYT------PEPE-RDY---------- : 264
F56D2.6 : ---ED------CPLLSVPGRTFPVEIFFT------PNAE-KDY---------- : 275
Prp43 : ---ND------APLLAVPGRTYPVELYYT------PEFQ-RDY---------- : 286
Dhr2 : ---NN------APILFVEGRKFDVKQYYL------KAPT-DDI---------- : 275
DHX40 : ---GN------CPIFDIPGRLYPVREKFCNLIGPRDREN-TAY---------- : 255
DHX35 : PARDT------CVILTVEGRTFPVDIFYL------QSPV-PDY---------- : 257
CG3225 : ---EV------SVKLSIEGRMHPVSNFYL------NEPC-ADY---------- : 244
Y67D2.6 : SDKDT------AGITSVEGRTHPVAVHHT------KTSV-PDY---------- : 279
DHX33 : ---NG------APVLYLEGRQHPIQVFYT------KQPQ-NDY---------- : 270
CG4901 : ----N------CKGMYLEGRTYPVRVMHT------KEEH-EDY---------- : 263
T05E8.3 : ---NN------AKVVLVAGRTFPIEVFHVNPKINKSFSS-TDY---------- : 362
DHX37 : ---KP------PRVIKVESRQFPVTVHFN------KRTPLEDY---------- : 452
CG3228 : ---IP------PELLKVEARQFPVTIHFQ------KRTP-DDY---------- : 458
RHA-2 : ---LT------PKVIKVDARQFPVSVHFE------KRTP-DDY---------- : 434
Dhr1 : ---IA------PRVLQVDARQFPVSIHFN------RRTA-FNY---------- : 602
DHX34 : ---SN------APVVQVPGRLFPITVVYQ------PQEAEPTT--------SKSEKLDPRP---------- : 361
CG32533 : ---GEEG----ARLVQVPGRLFPIKIRYL------PPPALELKAGQATSKRSQRNRIDPAP---------- : 355
SMGL-2 : ---EG------APVVQVPGRLFPIDVRWH------PIKQFIDQ------SDKKTHKIDPEP---------- : 293
DHX57 : ---NS------CPVTIPGRTFPVDQFFL----------EDAIAVTRYVLQDGSPYMRSMKQISKEKLKARRNRTAFEEVEEDLRLSLHLQDQDSVKDAVPDQQLDFKQLLARYKGVSKSVIKTMSIMDF : 823
CG1582 : ---GG------APVLDIPGRTFPVQQLFL----------EDILEMSDFVMEYDTKYCRKLKKQEQEILER----------ELEYADVQASGEAPGKKIKDEKLTLAETYQRYAEYSKPTCKSIYLMEP : 727
DHX36 : ---GN------CPMTHIPGFTFPVVEYLL----------EDVIEKIRYVPEQKEHRCQFKRGFMQGHVNRQEKEEKEAIYKERWP----------DYVRELRRRYSASTVDVIEMMED : 467
CG9323 : ---NN------CPMFRIEGVMFPVKMLYL----------EDVLSKTNYEFQKFRDRRPKRDPPERRMKHEAMIEPYLRRIRNSYDSRVLDKLRLPESEGCED---------- : 405
DHX29 : ---TH------CPILRISGRSYPVEVFHL----------EDIIEETGFVLEKDSEYCQKFLEEEEEVT----------INVTSKAGGIKKYQEYIPVQTGAHADLNPFYQKYSSRTQHAILYMNP : 842
DHX9 : ---FN------CPIIEVYGRTYPVQEYFL----------EDCIQMTHFVPPPKDKKKKDKDDDGGEDDDA----------NCNLICGDEYGPETRLSMSQLNE : 629
MLE : ---GI------CPVLEVPGRAFPVQQFFL----------EDIIQMTDFVPSAESRRKRKEVEDEEQLLSEDKDEAEI----------NYNKVCEDKYSQKTRNAMAMLSE : 632
RHA-1 : ---SSIPDVGPTPVLTMHGRTFPVQSFYL----------EDILHNLQHMPEEPDQKKRKKGGPPPPDDDEGDEEVDDKGR----------NMNILTDPSINESLKTAMSRISE : 638
DHX30 : ---GG------CPVIKVPGFMYPVKEHYL----------EDILAKLGKHQYLHRHR---------- : 601
YTHDC2 : ---GS------CPVIYIQGRPFEVKEMFL----------EDILRTTGYTNKEMLKYKKEKQQEEKQQTTLTEWYSAQENSFKPESQRQRTVLNVTDEYDLLDDGGDAVFSQLTEKDVNCLEPWLIKEMDA : 471
CG8915 : ---GG------GTVMDVEGRSFEVSIYHL----------EDILSKTGYMHPRMEKFLGKPTGKETPSELLAAYYGGN---------- : 435
F52B5.3 : ---ENHS----MDVIRIESRAFDVKVFYL----------DQILAMTGYQPPESKAFFSTAEYEEEDWKDEIEAVEREEKDKAKQRAADSSLSTLSKPVTKMAANTKRNSPIEEREWTSGPCSTPDGDIID : 462
TDRD9 : M--NP------AYIFEVEGKPHSVEEYYL----------NDL--------EHIHHSKLSPHLLEEPVITKDI---------- : 275
Spn-E : ---TTTNSI--PEVITTNHRRKHSIEKFY----------RDQLGSIIWNEEDVGHQQVPEINKHG---------- : 333
Ylr419w : ---PG------LATCHIEGRTFPITDYFL----------EDILSDLDFKIKREKALSYDDDSVDERNNDDQYLKPRADSK---------- : 834

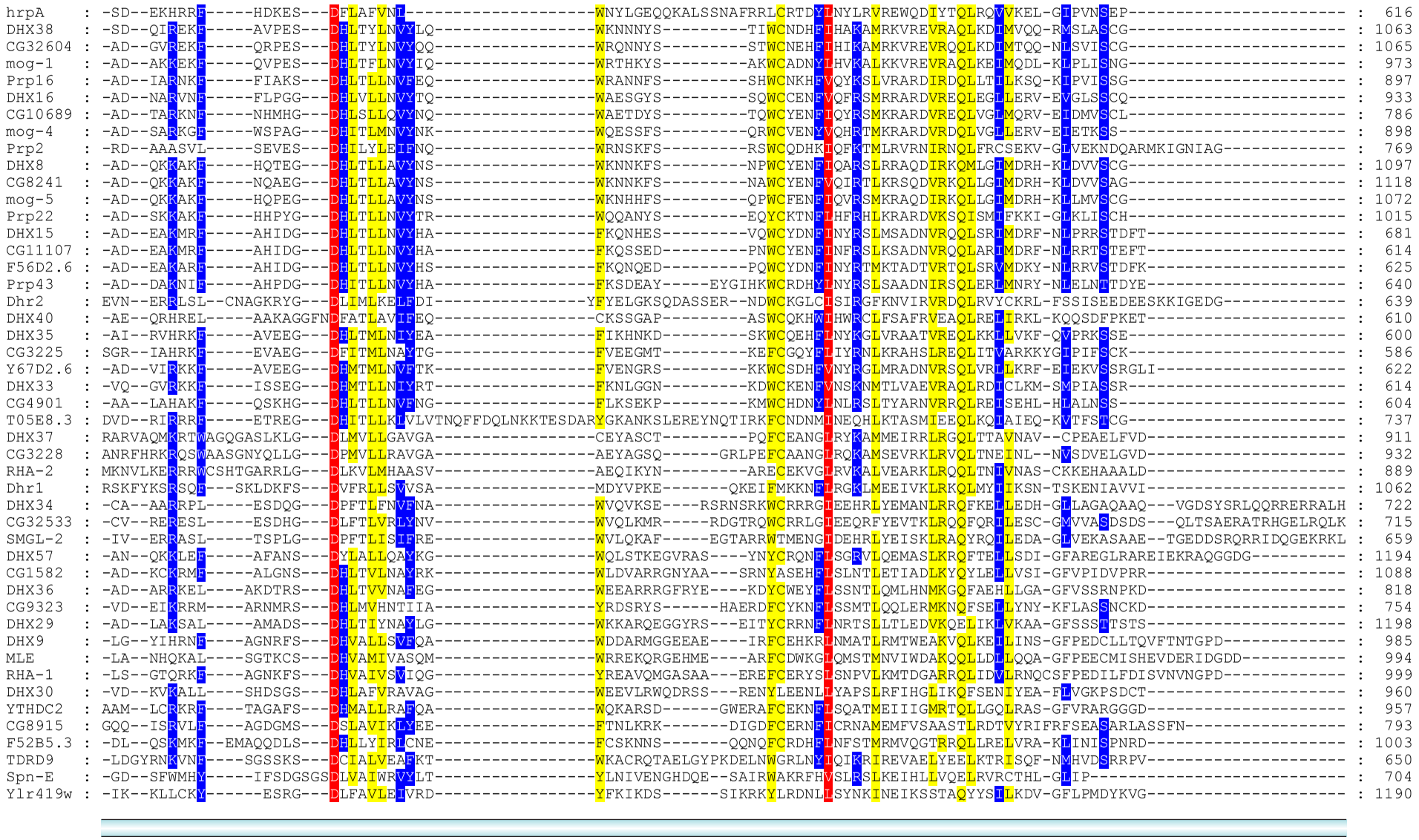
```
hrpA    : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX38   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG32604 : ------------------------------------------------------------------------------------------------------------------------ :    -
mog-1   : ------------------------------------------------------------------------------------------------------------------------ :    -
Prp16   : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX16   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG10689 : ------------------------------------------------------------------------------------------------------------------------ :    -
mog-4   : ------------------------------------------------------------------------------------------------------------------------ :    -
Prp2    : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX8    : ------------------------------------------------------------------------------------------------------------------------ :    -
CG8241  : ------------------------------------------------------------------------------------------------------------------------ :    -
mog-5   : ------------------------------------------------------------------------------------------------------------------------ :    -
Prp22   : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX15   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG11107 : ------------------------------------------------------------------------------------------------------------------------ :    -
F56D2.6 : ------------------------------------------------------------------------------------------------------------------------ :    -
Prp43   : ------------------------------------------------------------------------------------------------------------------------ :    -
Dhr2    : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX40   : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX35   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG3225  : ------------------------------------------------------------------------------------------------------------------------ :    -
Y67D2.6 : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX33   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG4901  : ------------------------------------------------------------------------------------------------------------------------ :    -
T05E8.3 : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX37   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG3228  : ------------------------------------------------------------------------------------------------------------------------ :    -
RHA-2   : ------------------------------------------------------------------------------------------------------------------------ :    -
Dhr1    : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX34   : ------------------------------------------------------------------------------------------------------------------------ :    -
CG32533 : ------------------------------------------------------------------------------------------------------------------------ :    -
SMGL-2  : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX57   : EK---------------------------------------------------------------------------------------------------------------------- :  825
CG1582  : MT---------------------------------------------------------------------------------------------------------------------- :  729
DHX36   : DK---------------------------------------------------------------------------------------------------------------------- :  469
CG9323  : ------------------------------------------------------------------------------------------------------------------------ :    -
DHX29   : HK---------------------------------------------------------------------------------------------------------------------- :  844
DHX9    : KE---------------------------------------------------------------------------------------------------------------------- :  631
MLE     : SD---------------------------------------------------------------------------------------------------------------------- :  634
RHA-1   : KD---------------------------------------------------------------------------------------------------------------------- :  640
DHX30   : ------------------------------------------------------------------------------------------------------------------------ :    -
YTHDC2  : CLSDIWLHKDI-----------------------------DAFAQVFHLILTENVSVDYRHSETSATALMVAAGRGFASQVEQLISMGANVHSKASNGWMALDWAKHFGQTEIVDLLESYSASLEFGNL :  571
CG8915  : ------------------------------------------------------------------------------------------------------------------------ :    -
F52B5.3 : ERAMRQYMNIIGNPDIRPVITVEFRPTTPNSVFINDVKYTTNPNQSAQALLQQKRERCLMPRGPLTFTCENPASTTKSPPSPTPTLGISSGLAKFTPFGGFGEKQRISTLLEKKFKEELKTALPDRTVYYH :  593
TDRD9   : ------------------------------------------------------------------------------------------------------------------------ :    -
Spn-E   : ------------------------------------------------------------------------------------------------------------------------ :    -
Ylr419w : ------------------------------------------------------------------------------------------------------------------------ :    -
```

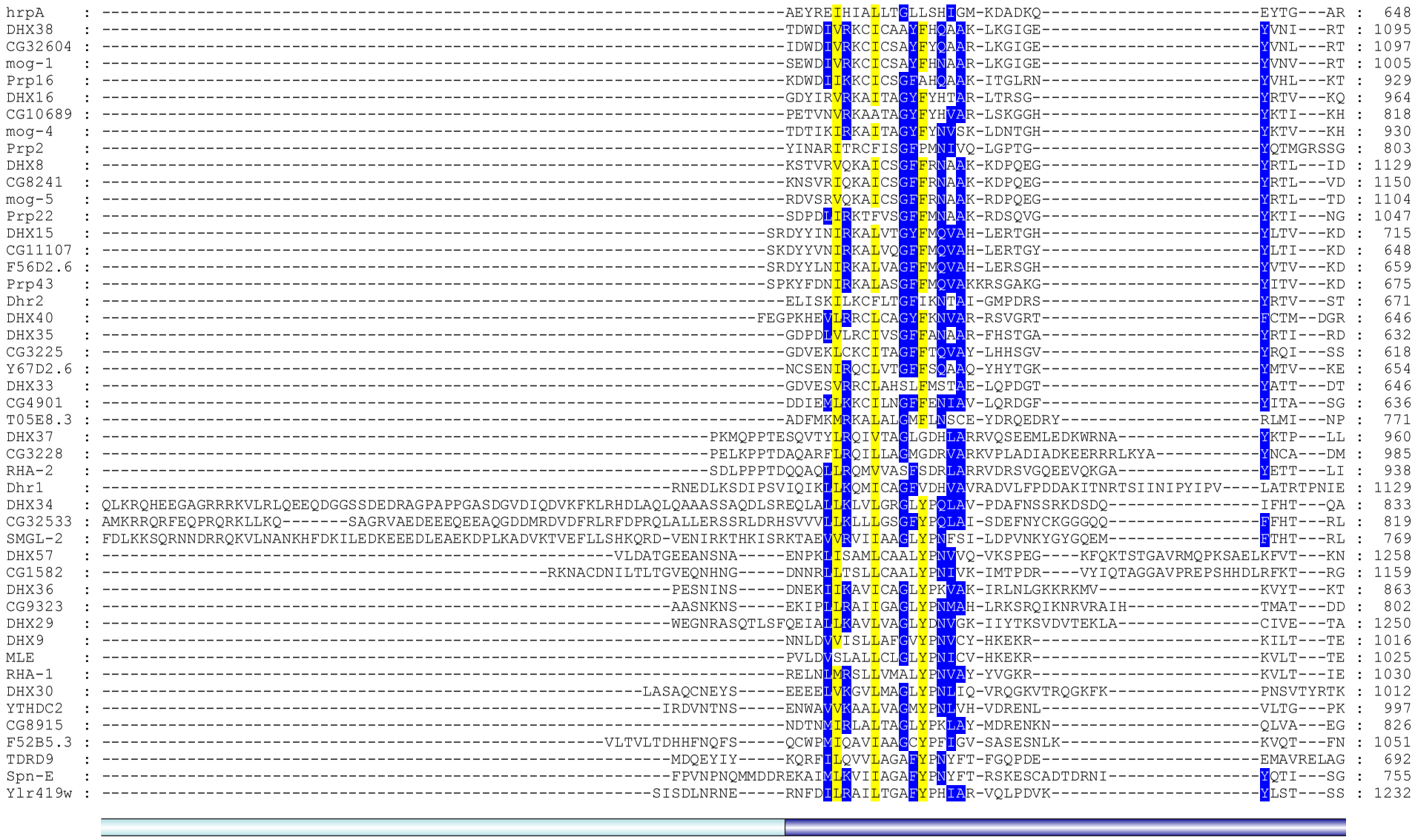

hrpA : -----LQAIFDAVDELS-----QES----HG-DILIFMSGEREIRDTADALNKLNLR----- : 316
DHX38 : -----VEAAVKQSLQVHL-----SGA----PG-DILIFMPGQEDIEVTSDQIVEHLEE----- : 766
CG32604 : -----VESAVKQALQVHL-----TPN----EG-DMLIFMPGQEDIEVTCEVLEERLAE----- : 768
mog-1 : -----VDAAVKQAVTIHL-----GGM----DG-DILIFMPGQEDIECTCEMIKEKLGE----- : 676
Prp16 : -----VEAAVSQAVKIHL-----AND--CSSG-DILIFMTGQEDIETTFDTLQEKFLQ----- : 589
DHX16 : -----LEACVVSVLQIHV-----TQP----PG-DILVFLTGQEEIEAACEMLQDPCRR----- : 634
CG10689 : -----IDACCVSVLQIHA-----TQP----LG-DILVFLTGQDEIETCQEVLHDRVKR----- : 487
mog-4 : -----VDAAIVTIMQIHL-----TQP---LPG-DILVFLTGQEEIETVQEALMERSKA----- : 600
Prp2 : -----IHAAITTIFQIHT-----TQS---LPG-DILVFLTGQEEIERTKTKLEEIMSK----- : 461
DHX8 : -----LDASLITVMQIHL-----TEP----PG-DILVFLTGQEEIDTACEILYERMKS----- : 799
CG8241 : -----LDASLITVMQIHL-----REP----PG-DILLFLTGQEEIDTACEILYERMKS----- : 820
mog-5 : -----LEAAHITVMQIHL-----TEP----PG-DVLVFLTGQEEIDTSCEVLYERMKS----- : 774
Prp22 : -----IEAALDCVIDIHI-----NEG----PG-DILVFLTGQEEIDSCCEILYDRVKT----- : 717
DHX15 : -----LEAAIRTVIQIHM-----CEE---EEG-DLLLFLTGQEEIDEACKRIKREVDD----- : 375
CG11107 : -----LEAAIRTVIQIHM-----CEE---IEG-DILMFLTGQEEIEEACKRIKREIDN----- : 308
F56D2.6 : -----LEAAIRTVIQIHM-----VEE---VEG-DILLFLTGQEEIEEACKRIDREIQA----- : 319
Prp43 : -----LDSAIRTVLQIHA-----TEE----AG-DILLFLTGEDEIEDAVRKISLEGDQ----- : 329
Dhr2 : -----VDAVIRCCIQINQ-----GEE----LG-DILCFLPGQEEIDKAVTIMEKIAKY----- : 318
DHX40 : -----IQAIVKVTMDIHL-----NEM----AG-DILVFLTGQFEIEKSCELLFQMAES----- : 298
DHX35 : -----IKSTVETVVKIHQ-----TEG----DG-DVLAFLTGQEEVETVVSMLIEQARA----- : 300
CG3225 : -----VKETVETVWKLHQ-----KEP----PG-DILAFLTGQEEVLEALDLLREYIAS----- : 287
Y67D2.6 : -----CQSAVDTVINIHK-----HEN----PG-DILVFLTGQDEVEDVCEKLRELAGN----- : 322
DHX33 : -----IHAALVSVFQIHQ-----EAP----SSQDILVFLTGQEEIEAMSKTCRDIAKH----- : 314
CG4901 : -----IHTVLVTLFHIHR-----TTP----KNHDVLIFLTGQEEIESLAQQIRQLAKI----- : 307
T05E8.3 : -----VYNAVICVKYVHL-----TEP----KGRDILVFLTGSEEIEAVASQLAELNGS----- : 406
DHX37 : -----SGECFRKVCKIHR-----MLP----AG-GILVFLTGQAEVHALCRRLRKAFPP----------SRARPQEKDDDQKDSVEEMRKFK : 518
CG3228 : -----VAEAYRKTLKIHN-----KLP----EG-GILIFVTGQQEVNQLVRKLRRTFPYHHAPTKDVAKNGKVSEEEKEETIDDAASTVEDPKELE : 538
RHA-2 : -----IASAFRKTCRIHE-----TLP----PG-AILVFVTGQHEVKQLITKLKKRYPVVYETDKNGEVLVKGTKEWKEKKVEAAKSIKLEDFKEE : 514
Dhr1 : -----TDEAFRKTCKIHQ-----KLP----PG-AILVFLTGQQEITHMVKRLRKEFPFKKNSKYNKDLETPVSKMGINSKTTDLEAEDID----- : 677
DHX34 : -----FLRVLESIDHKYP-----PEE----RG-DLLVFLSGMAEISAVLEAAQTYASH----- : 404
CG32533 : -----FVQVLSLIDQQYP-----TSE----RG-DVLIFVSGVNEIESVVEAVHEYATE----- : 398
SMGL-2 : -----YLKILELIDKQFP-----STQ----RG-DALIFLNGVAEISMVAEHLKNYAEL----- : 336
DHX57 : -----VNLELIEALLEWIVDGKH-----SYP----PG-AILVFLPGLAEIKMLYEQLQSNSLF----- : 873
CG1582 : -----INPELIESVLKYIVE-----GSHDWPREG-TILIFLPGFGEIQSVHDSLLDNALF----- : 778
DHX36 : -----VDLNLIVALIRYIVL-----EEE----DG-AILVFLPGWDNISTLHDLLMSQVMF----- : 514
CG9323 : -----LDFIADLVYYICE-----NEP----EG-AILVFLPGYDKISQLYNILDKPKTS----- : 448
DHX29 : -----INLDLILELLAYLDK-----SPQFRNIEG-AVLIFLPGLAHIQQLYDLLSNDRRF----- : 893
DHX9 : -----TPFELIEALLKYIET-----LNV----PG-AVLVFLPGWNLIYTMQKHLEMNPHF----- : 676
MLE : -----VSFELLEALLMHIKS-----KNI----PG-AILVFLPGWNLITFALMKFLQNTNIF----- : 679
RHA-1 : -----IPFGVIEAILNDIAS-----RGV----DG-AVLVFLPGWAEITMTLCNRLLEHQEF----- : 685
DHX30 : ----------HHESEDECALDLDLVTDLVLHIDA-----RGE----PG-GILCFLPGWQEIKGVQQRLQEALGM----- : 655
YTHDC2 : DESSLVQTNGSDLSAEDRELLKAYHHSFDDEKVDLLLIMHLLYNICH-----SCD----AG-AVLIFLPGYDEIVGLRDRILFDDKR----- : 648
CG8915 : -----TIVHPDIDNLLIVSLLELLLR-----QGD----AG-AVIVYLPGYSDVTSLLARLESSLPR----- : 486
F52B5.3 : PQMNDVVDKTDFSKLRFGEKYNLFYGSSSNNSVDYVLLDQVIQYLAD-----SPV----FG-TILVFLPGYEDIQQMLKAIDCWKNS----- : 670
TDRD9 : -----YEVAVSLIQMFDDLDMKESGNKAWSGAQFVLERS-SVLVFLPGLGEINYMHELLTSLV----- : 332
Spn-E : -----YRAAVKIIVIIDNMERKAAIQSRQSYDE-ALRYG-AVLIFLPGIYEIDTMAENLTCMLEN----- : 391
Ylr419w : -----FFTSGQINYDLLCQVVEYVHK-----RLKAANDNG-SIIVFLPGVGEINKCCNLLANKSNE----- : 889
IV

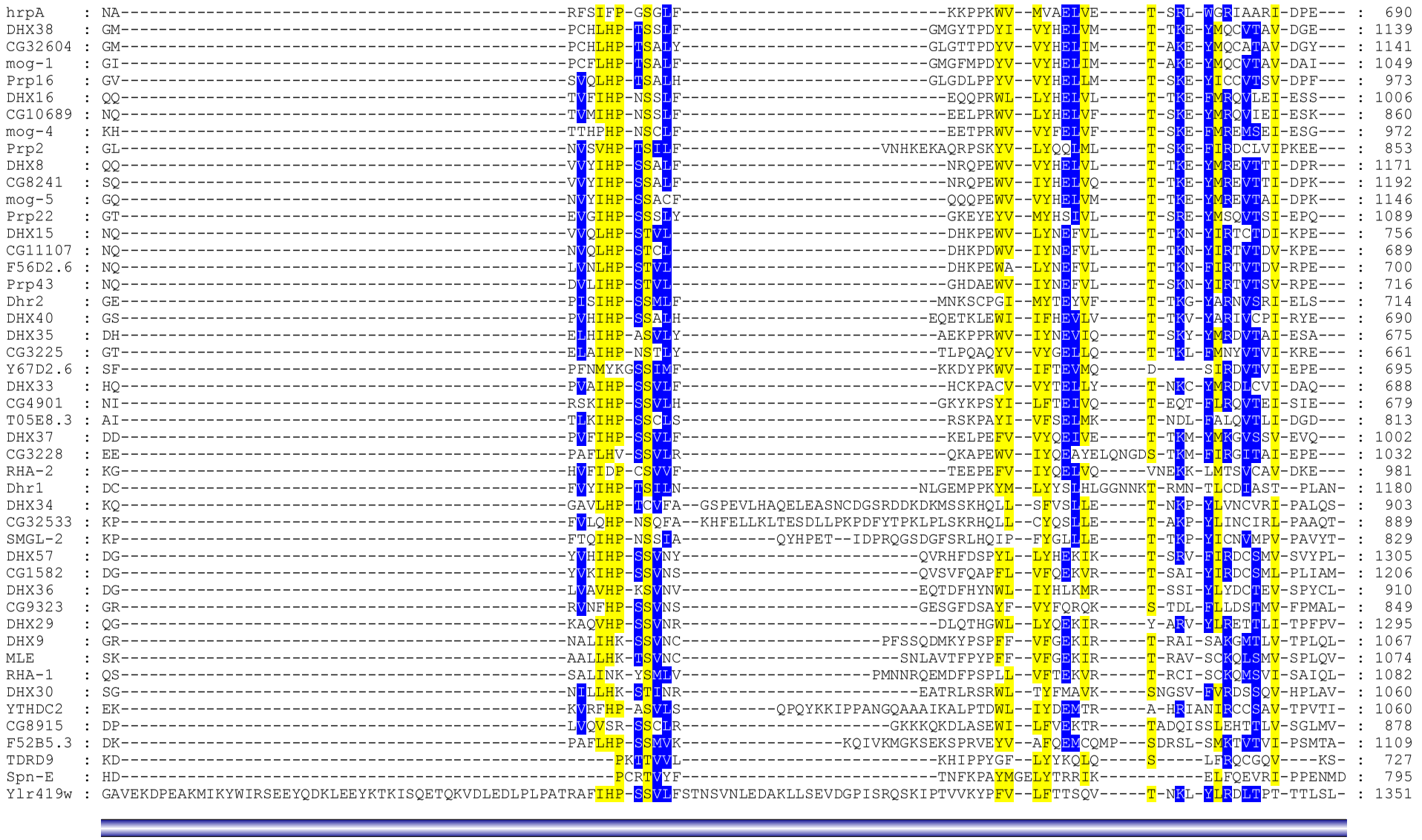

IVa
F
V

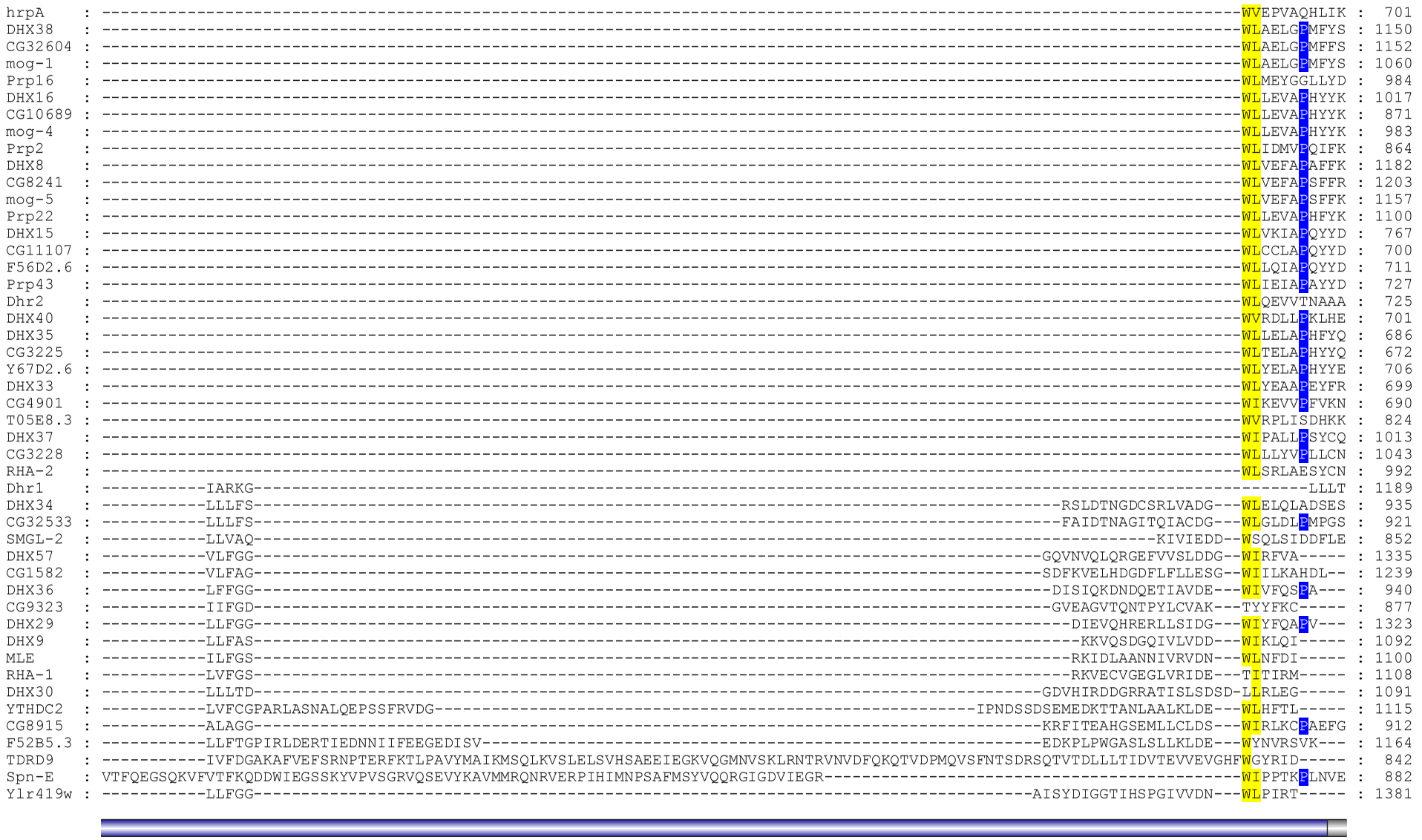

VI

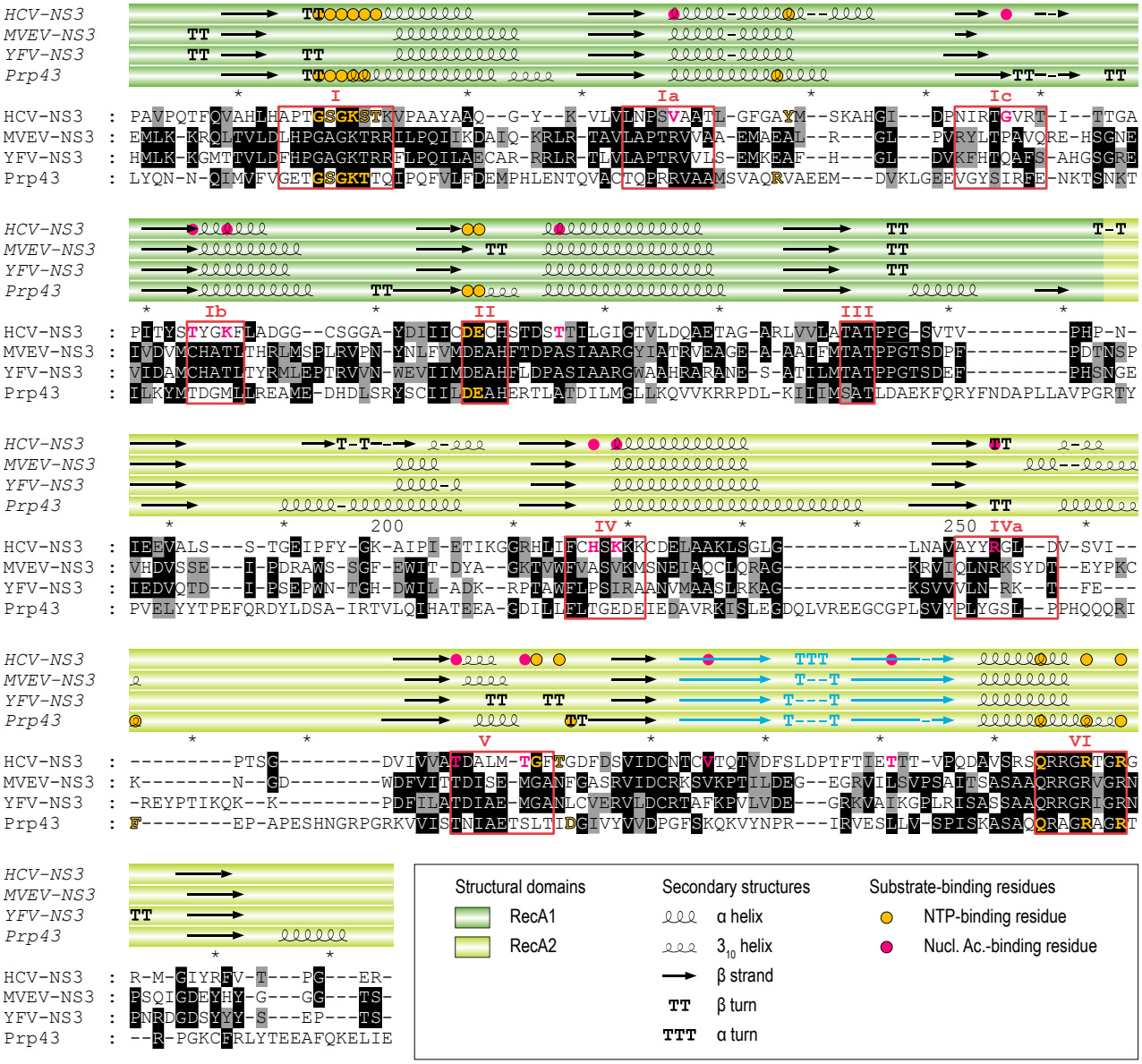

hrpA
DHX38
CG32604
mog-1
Prp16
DHX16
CG10689
mog-4
Prp2
DHX8
CG8241
mog-5
Prp22
DHX15
CG11107
F56D2.6
Prp43
Dhr2
DHX40
DHX35
CG3225
Y67D2.6
DHX33
CG4901
T05E8.3
DHX37
CG3228
RHA-2
Dhr1
DHX34
CG32533
SMGL-2
DHX57
CG1582
DHX36
CG9323
DHX29
DHX9
MLE
RHA-1
DHX30
YTHDC2
CG8915
F52B5.3
TDRD9
Spn-E
Ylr419w



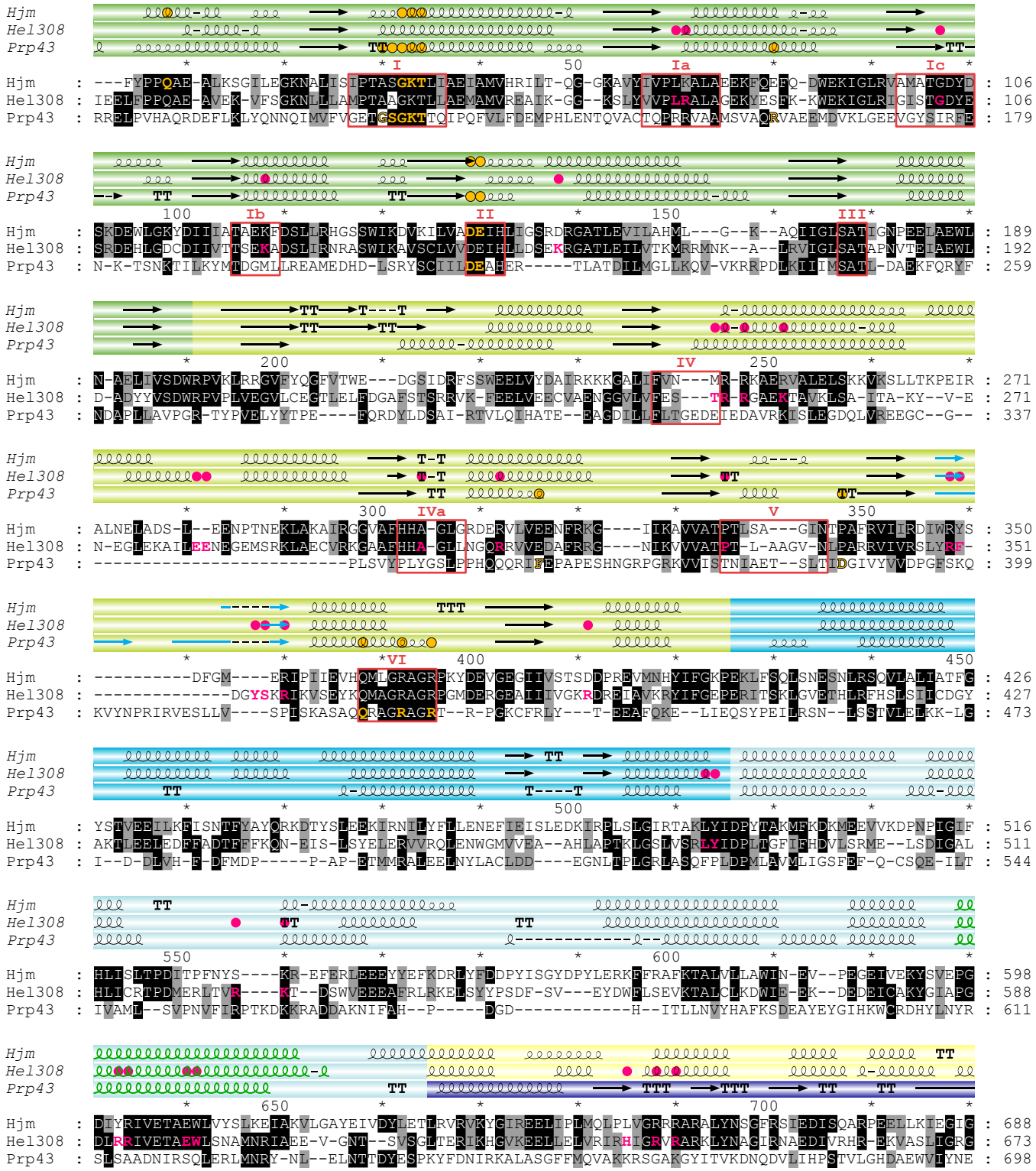

hrpA : -----------AEYREIHIALLTGLLSHIGM-KDADKQ-----------EYTG---AR : 648
DHX38 : -----------TDWDIVRKCICAAYFHQAAK-LKGIGE-----------YVNI---RT : 1095
CG32604 : -----------IDWDIVRKCICSAYFYQAAR-LKGIGE-----------YVNL---RT : 1097
mog-1 : -----------SEWDIVRKCICSAYFHNAAR-LKGIGE-----------YVNV---RT : 1005
Prp16 : -----------KDWDIIKKCICSGFAHQAAK-ITGLRN-----------YVHL---KT : 929
DHX16 : -----------GDYIRVRKAITAGYFYHTAR-LTRSG-----------YRTV---KQ : 964
CG10689 : -----------PETVNVRKAATAGYFYHVAR-LSKGGH-----------YKTI---KH : 818
mog-4 : -----------TDTIKIRKAITAGYFYNVSK-LDNTGH-----------YKTV---KH : 930
Prp2 : -----------YINARITRCFISGFPMNIVQ-LGPTG-----------YQTMGRSSG : 803
DHX8 : -----------KSTVRVQKAICSGFFRNAAK-KDPQEG-----------YRTL---ID : 1129
CG8241 : -----------KNSVRIQKAICSGFFRNAAK-KDPQEG-----------YRTL---VD : 1150
mog-5 : -----------RDVSRVQKAICSGFFRNAAK-RDPQEG-----------YRTL---TD : 1104
Prp22 : -----------SDPDLIRKTFVSGFFMNAAK-RDSQVG-----------YKTI---NG : 1047
DHX15 : -----------SRDYYINIRKALVTGYFMQVAH-LERTGH-----------YLTV---KD : 715
CG11107 : -----------SKDYYVNIRKALVQGFFMQVAH-LERTGY-----------YLTI---KD : 648
F56D2.6 : -----------SRDYYLNIRKALVAGFFMQVAH-LERSGH-----------YVTV---KD : 659
Prp43 : -----------SPKYFDNIRKALASGFFMQVAKKRSGAKG-----------YITV---KD : 675
Dhr2 : -----------ELISKILKCFLTGFIKNTAI-GMPDRS-----------YRTV---ST : 671
DHX40 : -----------FEGPKHEVLRRCLCAGYFKNVAR-RSVGRT-----------FCTM--DGR : 646
DHX35 : -----------GDPDLVLRCIVSGFFANAAR-FHSTGA-----------YRTI---RD : 632
CG3225 : -----------GDVEKLCKCITAGFFTQVAY-LHHSGV-----------YRQI---SS : 618
Y67D2.6 : -----------NCSENIRQCLVTGFFSQAAQ-YHYTGK-----------YMTV---KE : 654
DHX33 : -----------GDVESVRRCLAHSLFMSTAE-LQPDGT-----------YATT---DT : 646
CG4901 : -----------DDIEMLKKCILNGFFENIAV-LQRDGF-----------YITA---SG : 636
T05E8.3 : -----------ADFMKMRKALALGMFLNSCE-YDRQEDRY-----------RLMI---NP : 771
DHX37 : -----------PKMQPPTESQVTYLRQIVTAGLGDHLARRVQSEEMLEDKWRNA-----------YKTP---LL : 960
CG3228 : -----------PELKPPTDAQARFLRQILLAGMGDRVARKVPLADIADKEERRRLKYA-----------YNCA---DM : 985
RHA-2 : -----------SDLPPPTDQQAQLLRQMVVASFSDRLARRVDRSVGQEEVQKGA-----------YETT---LI : 938
Dhr1 : -----------RNEDLKSDIPSVIQIKLLKQMICAGFVDHVAVRADVLFPDDAKITNRTSIINIPYIPV----LATRTPNIE : 1129
DHX34 : QLKRQHEEGAGRRRKVLRLQEEQDGGSSDEDRAGPAPPGASDGVDIQDVKFKLRHDLAQLQAAASSAQDLSREQLALLKLVLGRGLYPQLAV-PDAFNSSRKDSDQ-----------IFHT---QA : 833
CG32533 : AMKRRQRFEQPRQRKLLKQ-------SAGRVAEDEEEQEEAQGDDMRDVDFRLRFDPRQLALLERSSRLDRHSVVVLLKLLLGSGFYPQLAI-SDEFNYCKGGGQQ-----------FFHT---RL : 819
SMGL-2 : FDLKKSQRNNDRRQKVLNANKHFDKILEDKEEEDLEAEKDPLKADVKTVEFLLSHKQRD-VENIRKTHKISRKTAEVVRVIIAAGLYPNFSI-LDPVNKYGYGQEM-----------FTHT---RL : 769
DHX57 : -----------VLDATGEEANSNA-----ENPKLISAMLCAALYPNVVQ-VKSPEG----KFQKTSTGAVRMQPKSAELKFVT---KN : 1258
CG1582 : -----------RKNACDNILTLTGVEQNHNG-----DNNRLLTSLLCAALYPNIVK-IMTPDR----VYIQTAGGAVPREPSHHDLRFKT---RG : 1159
DHX36 : -----------PESNINS-----DNEKIIKAVICAGLYPKVAK-IRLNLGKKRKMV-----------KVYT---KT : 863
CG9323 : -----------AASNKNS-----EKIPLLRAIIGAGLYPNMAH-LRKSRQIKNRVRAIH-----------TMAT---DD : 802
DHX29 : -----------WEGNRASQTLSFQEIALLKAVLVAGLYDNVGK-IIYTKSVDVTEKLA-----------CIVE---TA : 1250
DHX9 : -----------NNLDVVISLLAFGVYPNVCY-HKEKR-----------KILT---TE : 1016
MLE : -----------PVLDVSLALLCLGLYPNICV-HKEKR-----------KVLT---TE : 1025
RHA-1 : -----------RELNLMRSLLVMALYPNVAY-YVGKR-----------KVLT---IE : 1030
DHX30 : -----------LASAQCNEYS-----EEEELVKGVLMAGLYPNLIQ-VRQGKVTRQGKFK-----------PNSVTYRTK : 1012
YTHDC2 : -----------IRDVNTNS-----ENWAVVKAALVAGMYPNLVH-VDRENL-----------VLTG---PK : 997
CG8915 : -----------NDTNMIRLALTAGLYPKLAY-MDRENKN-----------QLVA---EG : 826
F52B5.3 : -----------VLTVLTDHHFNQFS-----QCWPMIQAVIAAGCYPFIGV-SASESNLK-----------KVQT---FN : 1051
TDRD9 : -----------MDQEYIY-----KQRFILQVVLAGAFYPNYFT-FGQPDE-----------EMAVRELAG : 692
Spn-E : -----------FPVNPNQMMDDREKAIMLKVIIAGAFYPNYFT-RSKESCADTDRNI-----------YQTI---SG : 755
Ylr419w : -----------SISDLNRNE-----RNFDILRAILTGAFYPHIAR-VQLPDVK-----------YLST---SS : 1232



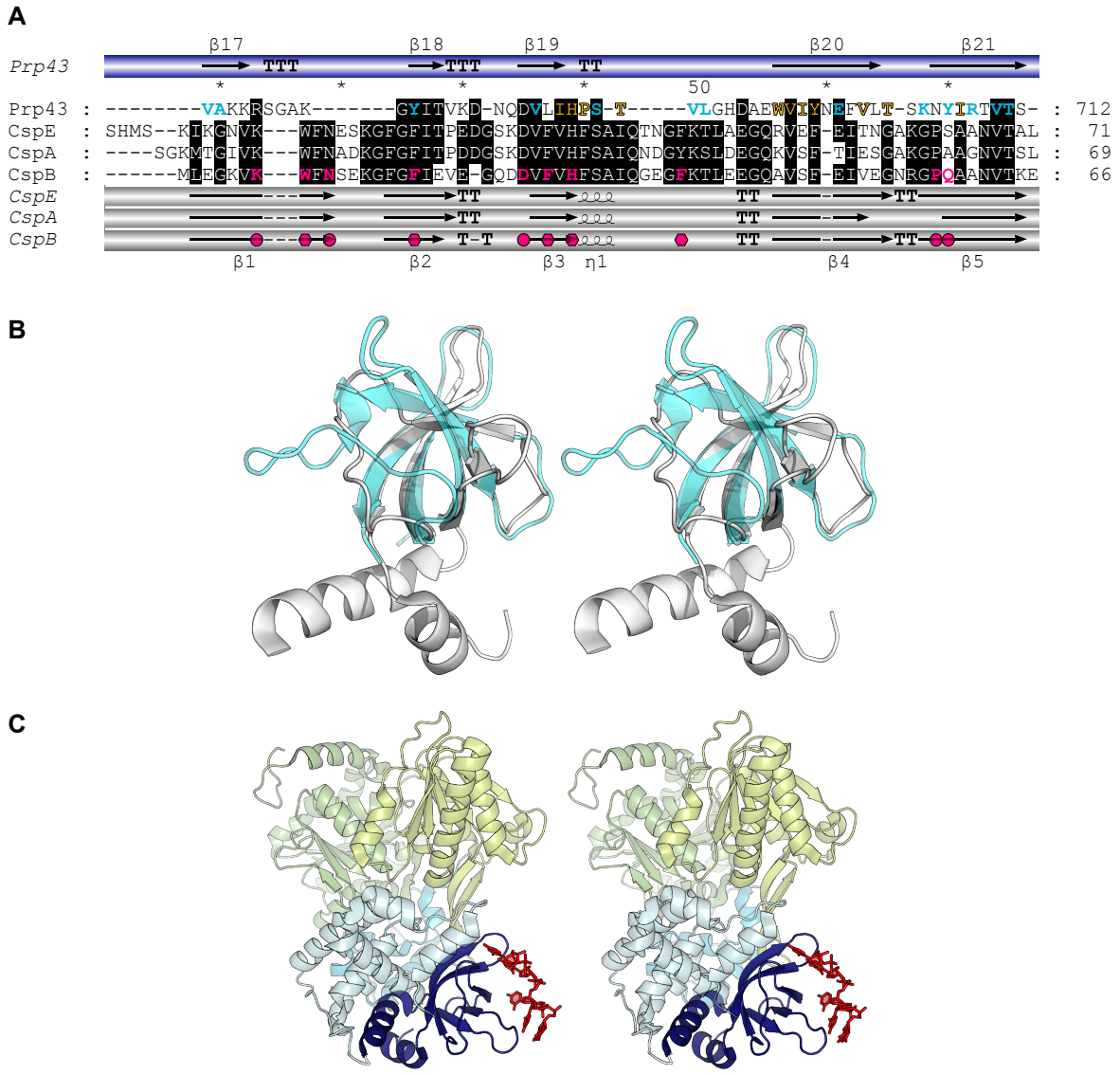

hrpA : ----------WVEPVAQHLIK : 701
DHX38 : ----------WLAELGPMFYS : 1150
CG32604 : ----------WLAELGPMFFS : 1152
mog-1 : ----------WLAELGPMFYS : 1060
Prp16 : ----------WLMEYGGLLYD : 984
DHX16 : ----------WLLEVAPHYYK : 1017
CG10689 : ----------WLLEVAPHYYK : 871
mog-4 : ----------WLLEVAPHYYK : 983
Prp2 : ----------WLIDMVPQIFK : 864
DHX8 : ----------WLVEFAPAFFK : 1182
CG8241 : ----------WLVEFAPSFFR : 1203
mog-5 : ----------WLVEFAPSFFK : 1157
Prp22 : ----------WLLEVAPHFYK : 1100
DHX15 : ----------WLVKIAPQYYD : 767
CG11107 : ----------WLCCLAPQYYD : 700
F56D2.6 : ----------WLLQIAPQYYD : 711
Prp43 : ----------WLIEIAPAYYD : 727
Dhr2 : ----------WLQEVVTNAAA : 725
DHX40 : ----------WVRDLLPKLHE : 701
DHX35 : ----------WLLELAPHFYQ : 686
CG3225 : ----------WLTELAPHYYQ : 672
Y67D2.6 : ----------WLYELAPHYYE : 706
DHX33 : ----------WLYEAAPEYFR : 699
CG4901 : ----------WIKEVVPFVKN : 690
T05E8.3 : ----------WVRPLISDHKK : 824
DHX37 : ----------WIPALLPSYCQ : 1013
CG3228 : ----------WLLLYVPLLCN : 1043
RHA-2 : ----------WLSRLAESYCN : 992
Dhr1 : ----------IARKG----------LLLT : 1189
DHX34 : ----------LLLFS----------RSLDTNGDCSRLVADG---WLELQLADSES : 935
CG32533 : ----------LLLFS----------FAIDTNAGITQIACDG---WLGLDLPMPGS : 921
SMGL-2 : ----------LLVAQ----------KIVIEDD--WSQLSIDDFLE : 852
DHX57 : ----------VLFGG----------GQVNVQLQRGEFVVSLDDG--WIRFVA----- : 1335
CG1582 : ----------VLFAG----------SDFKVELHDGDFLFLLESG--WIILKAHDL-- : 1239
DHX36 : ----------LFFGG----------DISIQKDNDQETIAVDE---WIVFQSPA--- : 940
CG9323 : ----------IIFGD----------GVEAGVTQNTPYLCVAK---TYYFKC----- : 877
DHX29 : ----------LLFGG----------DIEVQHRERLLSIDG---WIYFQAPV--- : 1323
DHX9 : ----------LLFAS----------KKVQSDGQIVLVDD---WIKLQI----- : 1092
MLE : ----------ILFGS----------RKIDLAANNIVRVDN---WLNFDI----- : 1100
RHA-1 : ----------LVFGS----------RKVECVGEGLVRIDE---TITIRM----- : 1108
DHX30 : ----------LLLTD----------GDVHIRDDGRRATISLSDSD-LLRLEG----- : 1091
YTHDC2 : ----------LVFCGPARLASNALQEPSSFRVDG----------IPNDSSDSEMEDKTTANLAALKLDE---WLHFTL----- : 1115
CG8915 : ----------ALAGG----------KRFITEAHGSEMLLCLDS---WIRLKCPAEFG : 912
F52B5.3 : ----------LLFTGPIRLDERTIEDNNIIFEEGEDISV----------EDKPLPWGASLSLLKLDE---WYNVRSVK--- : 1164
TDRD9 : ----------IVFDGAKAFVEFSRNPTERFKTLPAVYMAIKMSQLKVSLELSVHSAEEIEGKVQGMNVSKLRNTRVNVDFQKQTVDPMQVSFNTSDRSQTVTDLLLTIDVTEVVEVGHFWGYRID----- : 842
Spn-E : VTFQEGSQKVFVTFKQDDWIEGSSKYVPVSGRVQSEVYKAVMMRQNRVERPIHIMNPSAFMSYVQQRGIGDVIEGR----------WIPPTKPLNVE : 882
Ylr419w : ----------LLFGG----------AISYDIGGTIHSPGIVVDN---WLPIRT----- : 1381



Supplementary Figure 2 | **Sequence and structural similarities between Prp43 and viral NS3 DExH-box helicases**. The structure based sequences alignment of the helicase core region of Prp43 (RCSB PDB ID: 2XAU) with three flavivirin NS3 related helicase structures (RCSB PDB IDs: 3KQN, 2WV9, 1YKS) was performed with STRAP application (368,369). Similarity analysis was performed with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Identical and similar amino acids are shaded in black and grey, respectively. Residues implicated in NTP-binding or nucleic acid binding are depicted in yellow and pink, respectively. The RecA2 β-hairpin is depicted in blue. HCV: hepatitis C virus, MVEV: Murray Valley encephalitis virus, YFV: yellow fever virus.

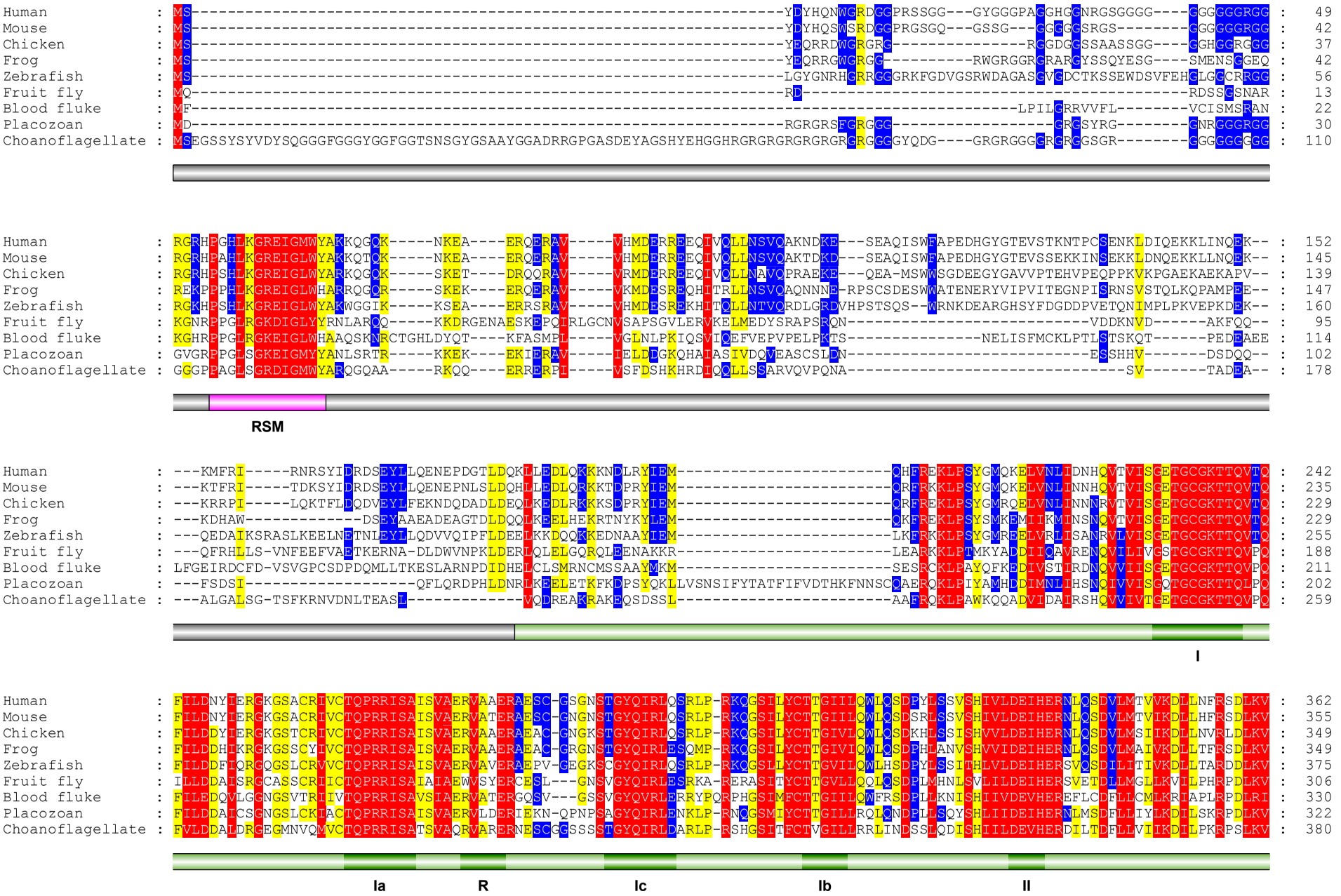


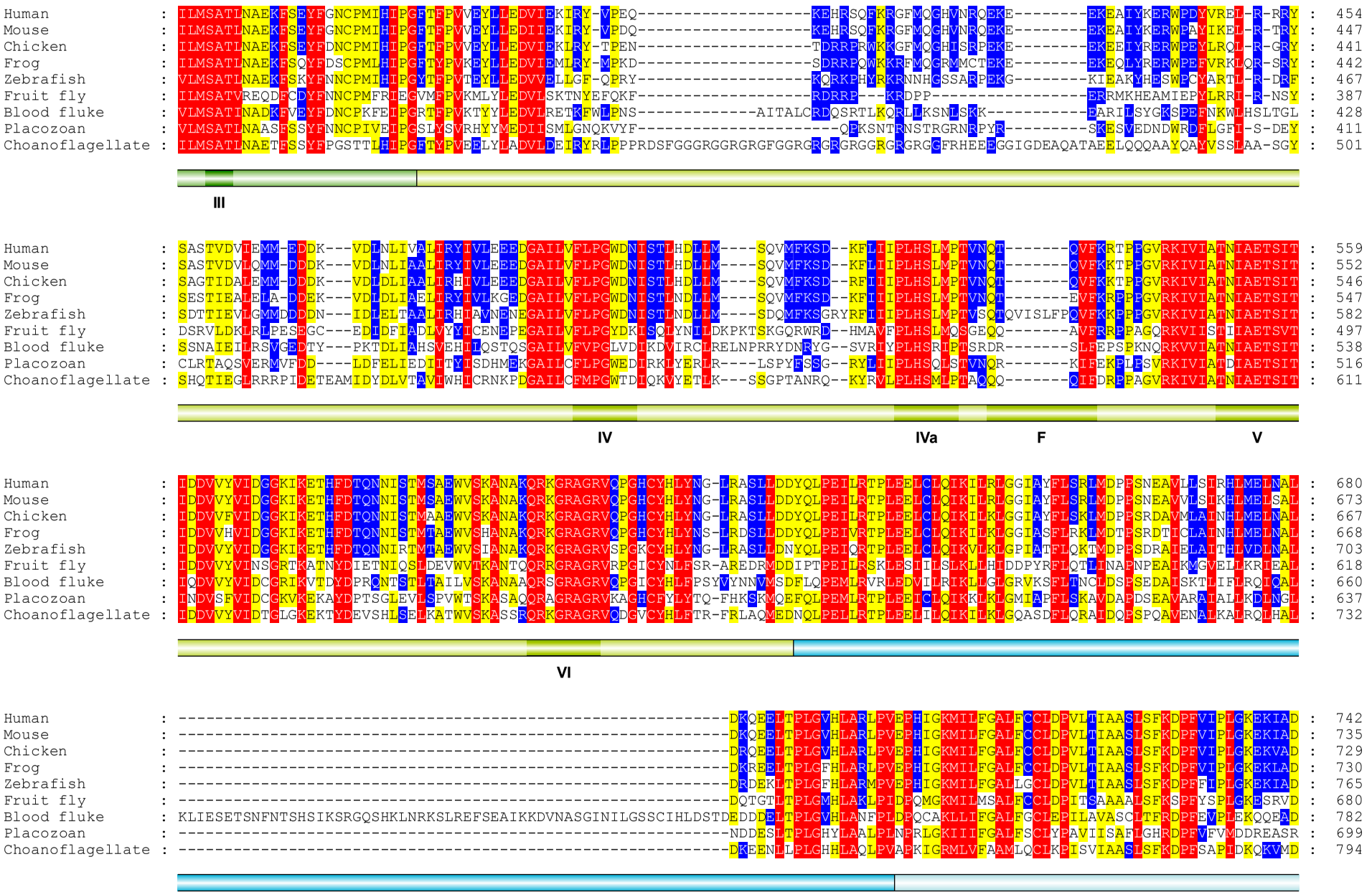
Supplementary Figure 3 | **Sequence and structural similarities between Prp43 and Ski2-related helicases**. The structure based sequences alignment of the helicase core and HA regions of Prp43 (RCSB PDB ID: 2XAU) with two Ski2-related helicases [RCSB PDB IDs: 2ZJA (Hjm), 2P6U (Hel308)] was performed with STRAP (368,369). Similarity analysis was performed with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Identical and similar amino acids are shaded in black and grey, respectively. Residues implicated in NTP-binding or nucleic acid binding are depicted in yellow and pink, respectively. The RecA2 β-hairpin and the ratchet helix are depicted in blue and green, respectively. The helix-loop-helix (HLH) domain 5 in Hjm and Hel308 is depicted in yellow.

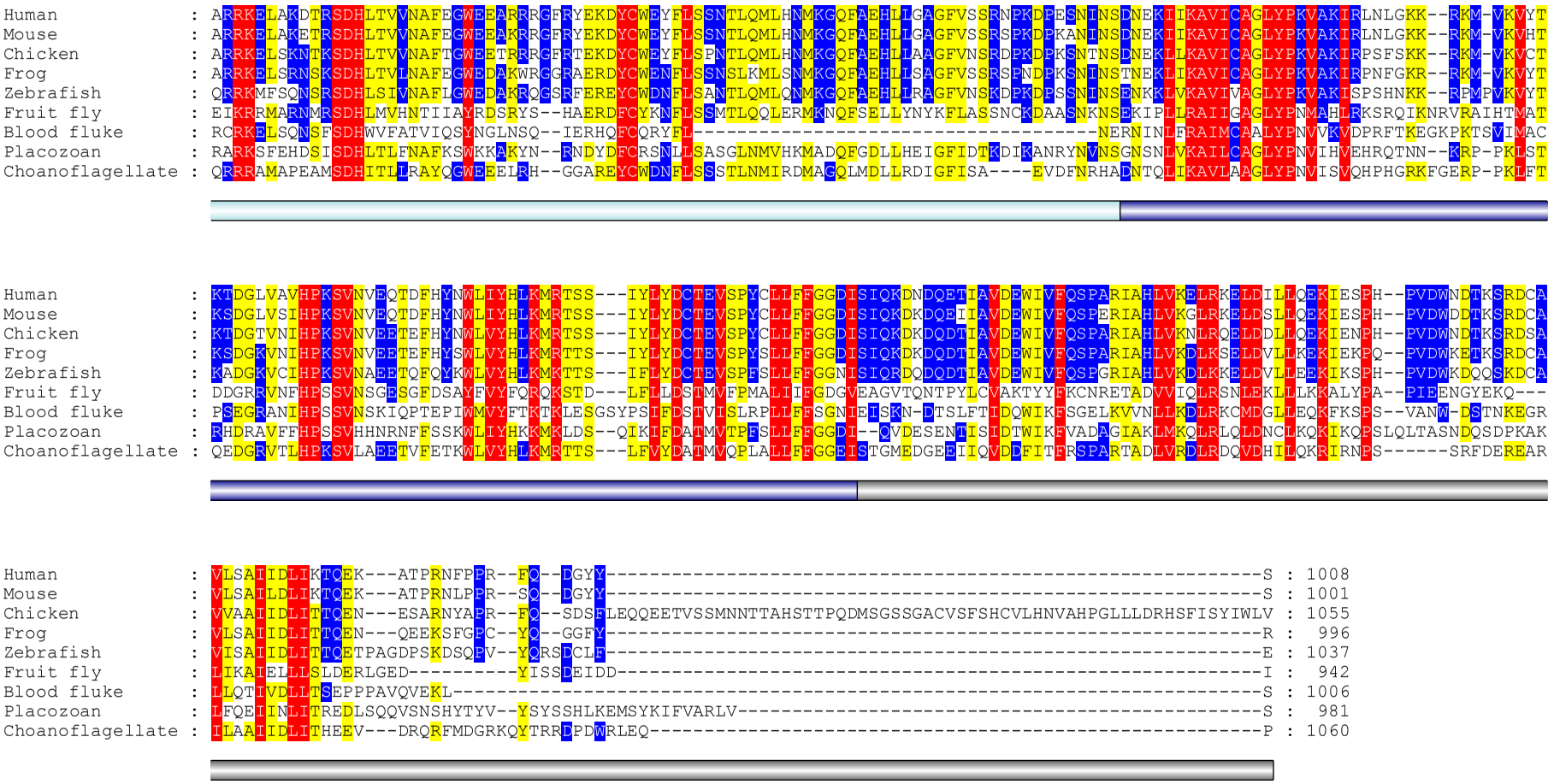


Supplementary Figure 4 | **Structural similarities between Prp43 OB-fold domain and prokaryotic cold shock proteins (Csp)**. (**A**) Structure based sequence alignment between the OB-fold domain of Prp43 (RCSB PDB ID: 3KX2, aa. 660–712) and three bacterial Csps [RCSB PDB IDs: 3I2Z (*Salmonella Typhimurium* CspE), 1MJC (*E. coli* CspA), 3PF5 (*Bacillus subtilis* CspB)]. The alignment was performed with STRAP (368,369). Similarity analysis was carried out with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Coloured amino acids in Prp43 sequence denote reciprocal conserved residues in paralogous DEAH-box proteins (yellow, 99–80 % similarity; blue, 79–60 % similarity; cf. Supplementary Figure 1). For CspB, residues implicated in nucleic acid binding are depicted in pink. Pink hexagons and circles denote respectively base stacking and polar interactions betwen CspB and the oligoribonucleotide. (**B**) Three-dimensional protein structure superposition of Prp43 OB-fold domain (white, RCSB PDB ID: 3KX2, aa. 643–725) with *E. coli* CspA (transparent light blue, RCSB PDB ID: 1MJC, aa. 2–70). (**C**) Model of Prp43 OB-fold domain in complex with an rU_6 oligoribonucleotide. The model was constructed by simple fitting of *B. subtilis* CspB/RNA complex to the OB-fold domain of Prp43 and masking CspB protein structure.

Supplementary Figure 5 ► | **Amino acid sequence conservation within RHAU and orthologous proteins**. Sequence alignment was carried out with MAFFT (version 6, ref. 313). Similarity analysis was performed with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Similarity is shown in red for 100 %, yellow for 99–80 % and blue for 79–60 %. Species and accession numbers of RHAU orthologues listed are: human (*Homo sapiens*, NP_065916), mouse (*Mus musculus*, NP_082412), chicken (*Gallus gallus*, XP_422834), frog (*Xenopus tropicalis*, ENSXETP00000016958), zebrafish (*Danio rerio*, NP_001122016), fruit fly (*Drosophila melanogaster*, NP_610056), blood fluke (*Schistosoma mansoni*, XP_002577014), placozoan (*Trichoplax adhaerens*, XP_002110272), choanoflagellate (*Monosiga brevicollis*, XP_001747335). The position of the RSM, the five structural domains (RecA1, RecA2, WH, Ratchet and OB-fold) as well as the conserved ATPase/helicase motifs I–VI is shown as coloured boxes beneath the sequence alignment.


Human
Mouse
Chicken
Frog
Zebrafish
Fruit fly
Blood fluke
Placozoan
Choanoflagellate
RSM
I
Ia
R
Ic
Ib
II


Human
Mouse
Chicken
Frog
Zebrafish
Fruit fly
Blood fluke
Placozoan
Choanoflagellate
III
IV
IVa
F
V
VI

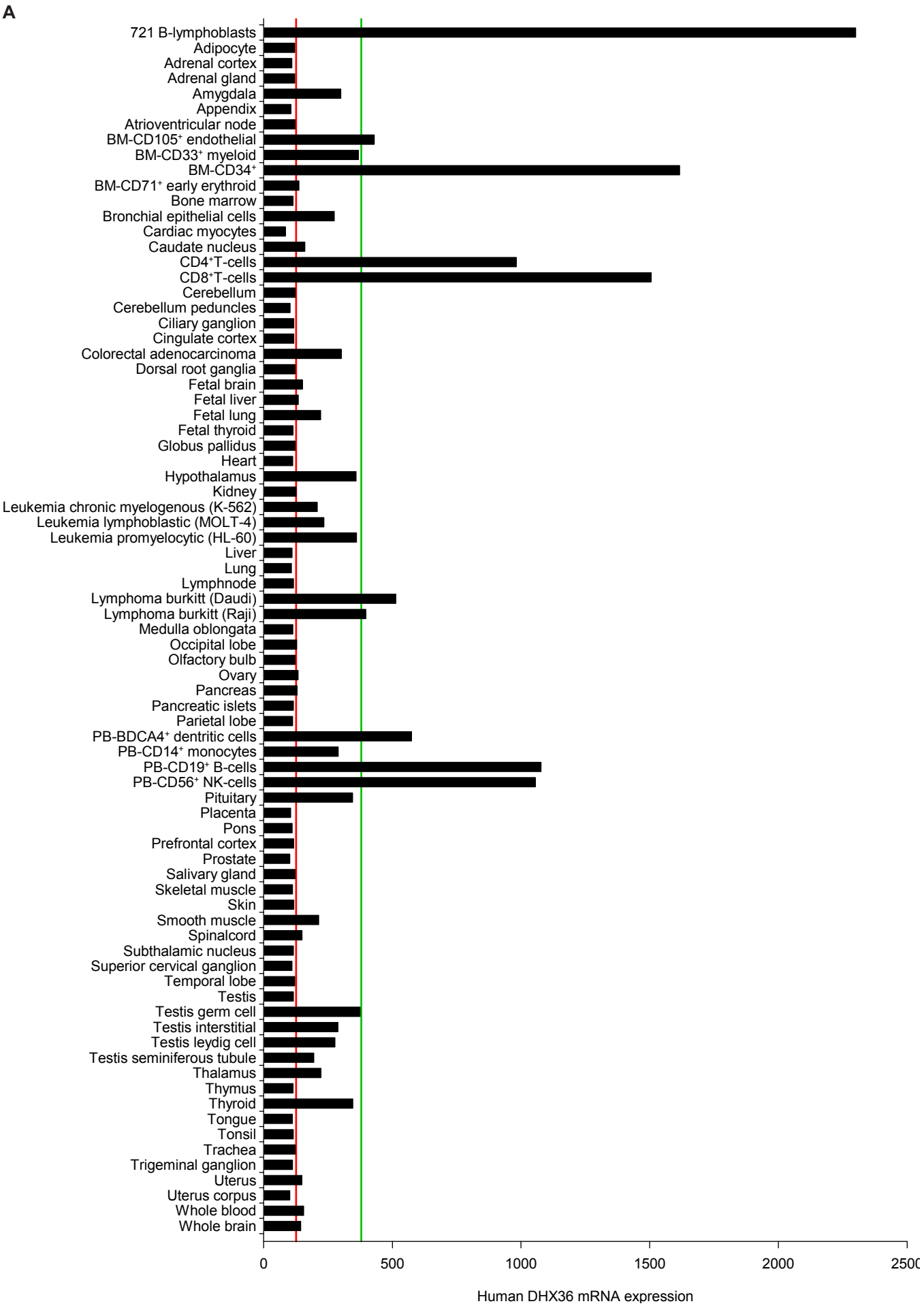

Human : ARRKELAKDTRSDHLTVVNAFEGWEEARRRGFRYEKDYCWEYFLSSNTLQMLHNMKGQFAEHLLGAGFVSSRNPKDPESNINSDNEKIIKAVICAGLYPKVAKIRLNLGKK--RKM-VKVYT : 861
Mouse : ARRKELAKETRSDHLTVVNAFEGWEEAKRRGFRYEKDYCWEYFLSSNTLQMLHNMKGQFAEHLLGAGFVSSRSPKDPKANINSDNEKIIKAVICAGLYPKVAKIRLNLGKK--RKM-VKVHT : 854
Chicken : ARRKELSKNTKSDHLTVVNAFTGWEETRRRGFRTEKDYCWEYFLSPNTLQMLHNMKGQFAEHLLAAGFVNSRDPKDPKSNTNSDNEKLLKAVICAGLYPKVAKIRPSFSKK--RKM-VKVCT : 848
Frog : ARRKELSRNSKSDHLTVLNAFEGWEDAKWRGGRAERDYCWENFLSSNSLKMLSNMKGQFAEHLLSAGFVSSRSENDPKSNINSTNEKLIKAVICAGLYPKVAKIRPNFGKR--RKM-VKVYT : 849
Zebrafish : QRRKMFSQNSRSDHLSIVNAFLGWEDAKRQGSRFEREYCWDNFLSANTLQMLQNMKGQFAEHLLRAGFVNSKDPKDPSSNINSENKKLVKAVIVAGLYPKVAKISPSHNKK--RPMPVKVYT : 885
Fruit fly : EIKRRMARNMRSDHLMVHNTIIAYRDSRYS--HAERDFCYKNFLSSMTLQQLERMKNQFSELLYNYKFLASSNCKDAASNKNSEKIPLLRAIIGAGLYPNMAHLRKSRQIKNRVRAIHTMAT : 800
Blood fluke : RCRKELSQNSFSDHWVFATVIQSYNGLNSQ--IERHQFCQRYFL----------NERNINLFRAIMCAALYPNVVKVDPRFTKEGKPKTSVIVAC : 865
Placozoan : RARKSFEHDSISDHLTLFNAFKSWKKAKYN--RNDYDFCRSNLLSASGLNMVHKMADQFGDLLHEIGFIDTKDIKANRYNVNSGNSNLVKAILCAGLYPNVIHVEHRQTNN--KRP-PKLST : 816
Choanoflagellate : QRRRAMAPEAMSDHITLLRAYQGWEEELRH--GGAREYCWDNFLSSSTLNMIRDMAGQLMDLLRDIGFISA----EVDFNRHADNTQLIKAVLAAGLYPNVISVQHPHGRKFGERP-PKLFT : 909

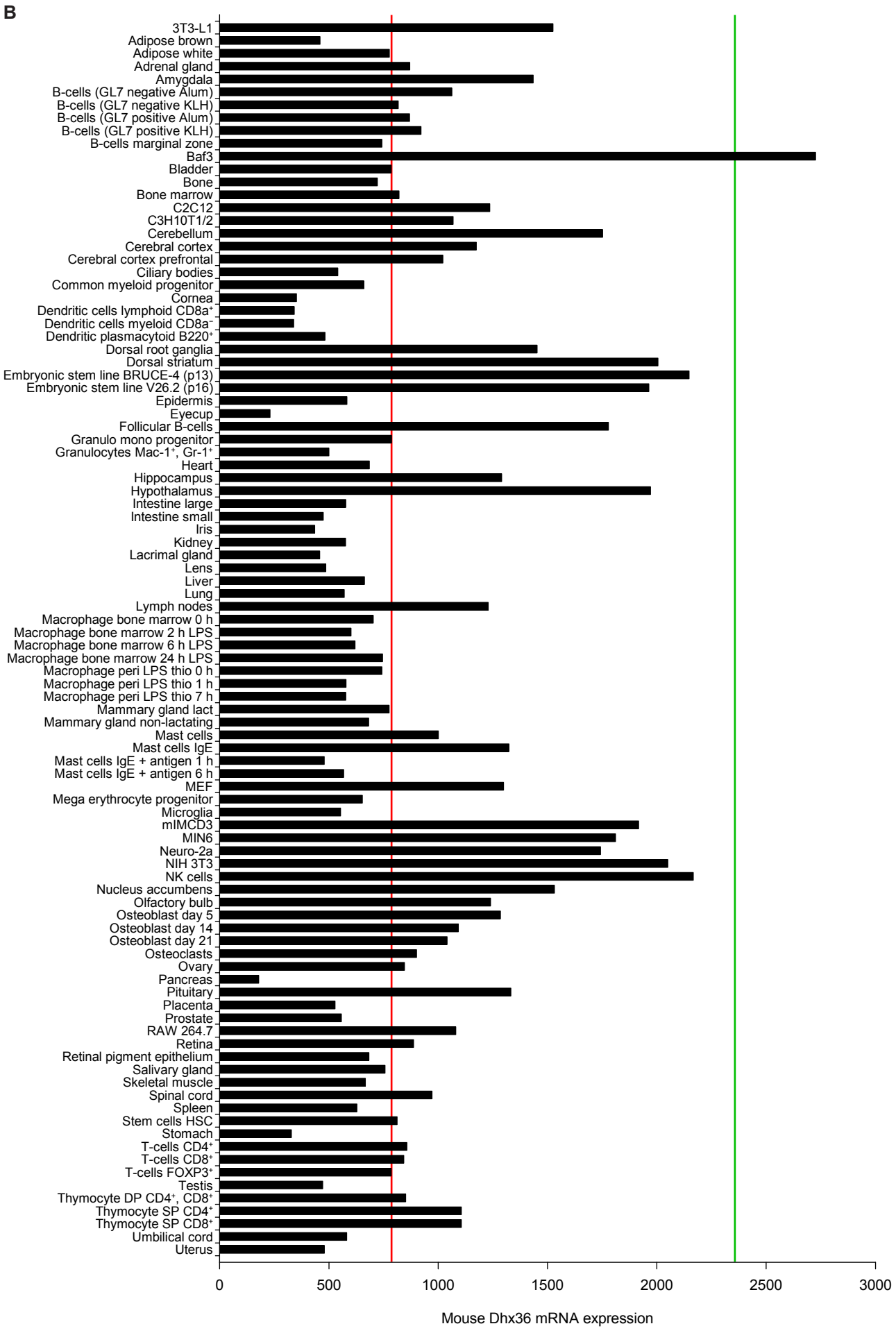
Human : KTDGLVAVHPKSVNVEQTDFHYNWLIYHLKMRTSS---IYLYDCTEVSPYCLLFFGGDISIQKDNDQETIAVDEWIVFQSPARIAHLVKELRKELDILLQEKIESPH--PVDWNDTKSRDCA : 978
Mouse : KSDGLVSIHPKSVNVEQTDFHYNWLIYHLKMRTSS---IYLYDCTEVSPYCLLFFGGDISIQKDKDQEIIAVDEWIVFQSPERIAHLVKGLRKELDSLLQEKIESPH--PVDWDDTKSRDCA : 971
Chicken : KTDGTVNIHPKSVNVEETEFHYNWLVYHLKMRTSS---IYLYDCTEVSPYCLLFFGGDISIQKDKDQDTIAVDEWIVFQSPARIAHLVKNLRQELDDLLQEKIENPH--PVDWNDTKSRDSA : 965
Frog : KSDGKVNIHPKSVNVEETEFHYSWLVYHLKMRTTS---IYLYDCTEVSPYSLLFFGGDISIQKDKDQDTIAVDEWIVFQSPARIAHLVKDLKSELDVLLKEKIEKPQ--PVDWKETKSRDCA : 966
Zebrafish : KADGKVCIHPKSVNAEETQFQYKWLVYHLKMKTTS---IFLYDCTEVSPESLLFFGGNISIQRDQDQDTIAVDEWIVFQSPGRIAHLVKDLKKELDVLLEEKIKSPH--PVDWKDQQSKDCA : 1002
Fruit fly : DDGRRVNFHPSSVNSGESGFDSAYFVYFQRQKSTD---LFLLDSTMVFPMALIIFGDGVEAGVTQNTPYLCVAKTYYFKCNRETADVVIQLRSNLEKLLLKKALYPA--PIEENGYEKQ--- : 914
Blood fluke : PSEGRANIHPSSVNSKIQPTEPIWMVYFTKTKLESGSYPSIFDSTVISLRPLLFFSGNIEISKN-DTSLFTIDQWIKFSGELKVVNLLKDLRKCMDGLLEQKFKSPS--VANW-DSTNKEGR : 983
Placozoan : RHDRAVFFHPSSVHHNRNFFSSKWLIYHKKMKLDS--QIKIFDATMVTPESLLFFGGDI--QVDESENTISIDTWIKFVADAGIAKLMKQLRLQLDNCLKQKIKQPSLQLTASNDQSDPKAK : 934
Choanoflagellate : QEDGRVTLHPKSVLAEETVFETKWLVYHLKMRTTS---LFVYDATMVQPLALLFFGGEISTGMEDGEEIIQVDEFITFRSPARTADLVRDLRDQVDHILQKRIRNPS------SRFDEREAR : 1022

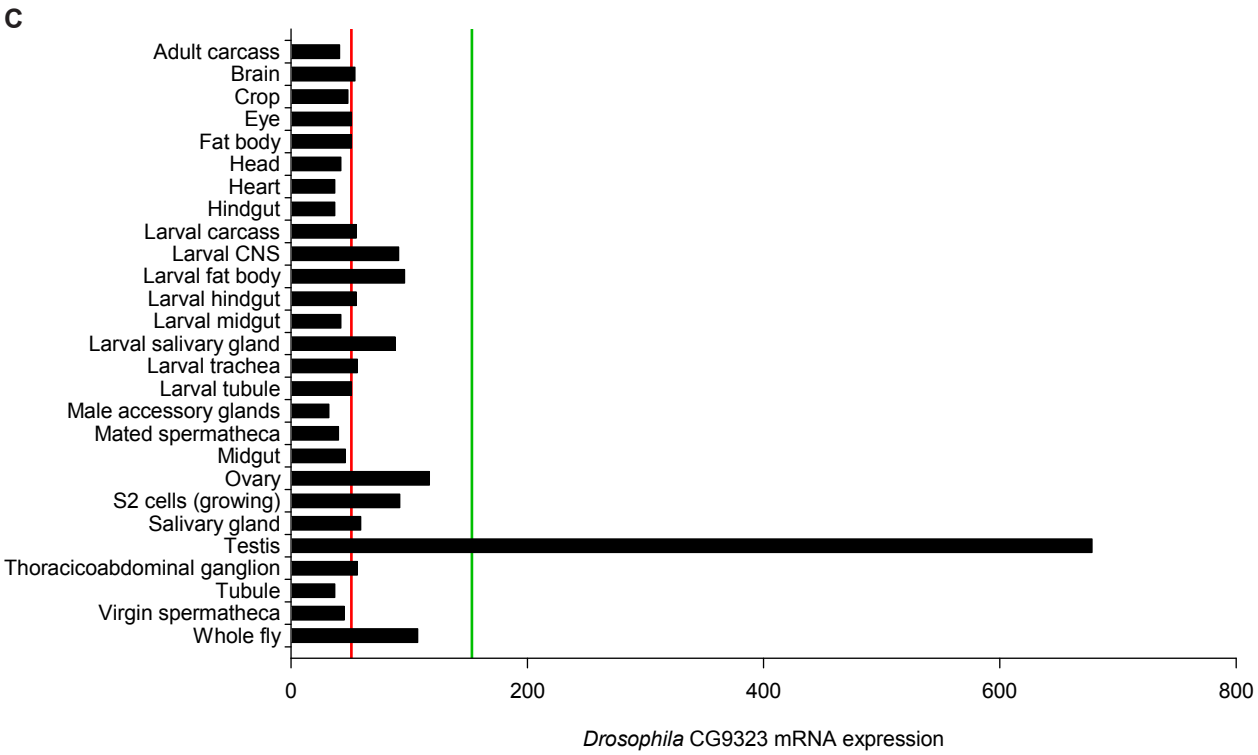
Human : VLSAIIDLIKTQEK---ATPRNFPRR--FQ--DGYY----------S : 1008
Mouse : VLSAILDLIKTQEK---ATPRNLPRR--SQ--DGYY----------S : 1001
Chicken : VVAAIIDLITTQEN---ESARNYAPR--FQ--SDSFLEQQEETVSSMNNTTAHSTTPQDMSGSSGACVSFSHCVLHNVAHPGLLLDRHSFISYIWLV : 1055
Frog : VLSAIIDLITTQEN---QEEKSFGPC--YQ--GGFY----------R : 996
Zebrafish : VISAIIDLITTQETPAGDPSKDSQPV--YQRSDCLF----------E : 1037
Fruit fly : LIKAIELLLSLDERLGED----------YISSDEIDD----------I : 942
Blood fluke : LLQTIVDLLTSEPPPAVQVEKL----------S : 1006
Placozoan : LFQEIINLITREDLSQQVSNSHYTYV--YSYSSHLKEMSYKIFVARLV----------S : 981
Choanoflagellate : ILAAIIDLITHEEV---DRQRFMDGRKQYTRRDPDWRLEQ----------P : 1060

Supplementary Figure 6 ► | **RHAU mRNA expression in various human, mouse and *Drosophila* tissues**. (**A**) Human, (**B**) mouse and (**C**) *Drosophila*. The red line represents the median expression value. The green line is equivalent to 3-fold the median expression value. Sources: BioGPS (273) and FlyAtlas (370) databases.

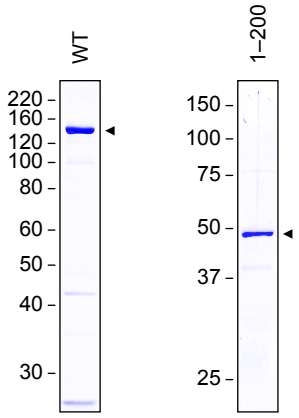
A


721 B-lymphoblasts
Adipocyte
Adrenal cortex
Adrenal gland
Amygdala
Appendix
Atrioventricular node
BM-CD105+ endothelial
BM-CD33+ myeloid
BM-CD34+
BM-CD71+ early erythroid
Bone marrow
Bronchial epithelial cells
Cardiac myocytes
Caudate nucleus
CD4+T-cells
CD8+T-cells
Cerebellum
Cerebellum peduncles
Ciliary ganglion
Cingulate cortex
Colorectal adenocarcinoma
Dorsal root ganglia
Fetal brain
Fetal liver
Fetal lung
Fetal thyroid
Globus pallidus
Heart
Hypothalamus
Kidney
Leukemia chronic myelogenous (K-562)
Leukemia lymphoblastic (MOLT-4)
Leukemia promyelocytic (HL-60)
Liver
Lung
Lymphnode
Lymphoma burkitt (Daudi)
Lymphoma burkitt (Raji)
Medulla oblongata
Occipital lobe
Olfactory bulb
Ovary
Pancreas
Pancreatic islets
Parietal lobe
PB-BDCA4+ dentritic cells
PB-CD14+ monocytes
PB-CD19+ B-cells
PB-CD56+ NK-cells
Pituitary
Placenta
Pons
Prefrontal cortex
Prostate
Salivary gland
Skeletal muscle
Skin
Smooth muscle
Spinalcord
Subthalamic nucleus
Superior cervical ganglion
Temporal lobe
Testis
Testis germ cell
Testis interstitial
Testis leydig cell
Testis seminiferous tubule
Thalamus
Thymus
Thyroid
Tongue
Tonsil
Trachea
Trigeminal ganglion
Uterus
Uterus corpus
Whole blood
Whole brain
0
500
1000
1500
2000
2500
Human DHX36 mRNA expression


B
3T3-L1
Adipose brown
Adipose white
Adrenal gland
Amygdala
B-cells (GL7 negative Alum)
B-cells (GL7 negative KLH)
B-cells (GL7 positive Alum)
B-cells (GL7 positive KLH)
B-cells marginal zone
Baf3
Bladder
Bone
Bone marrow
C2C12
C3H10T1/2
Cerebellum
Cerebral cortex
Cerebral cortex prefrontal
Ciliary bodies
Common myeloid progenitor
Cornea
Dendritic cells lymphoid CD8a+
Dendritic cells myeloid CD8a−
Dendritic plasmacytoid B220+
Dorsal root ganglia
Dorsal striatum
Embryonic stem line BRUCE-4 (p13)
Embryonic stem line V26.2 (p16)
Epidermis
Eyecup
Follicular B-cells
Granulo mono progenitor
Granulocytes Mac-1+, Gr-1+
Heart
Hippocampus
Hypothalamus
Intestine large
Intestine small
Iris
Kidney
Lacrimal gland
Lens
Liver
Lung
Lymph nodes
Macrophage bone marrow 0 h
Macrophage bone marrow 2 h LPS
Macrophage bone marrow 6 h LPS
Macrophage bone marrow 24 h LPS
Macrophage peri LPS thio 0 h
Macrophage peri LPS thio 1 h
Macrophage peri LPS thio 7 h
Mammary gland lact
Mammary gland non-lactating
Mast cells
Mast cells IgE
Mast cells IgE + antigen 1 h
Mast cells IgE + antigen 6 h
MEF
Mega erythrocyte progenitor
Microglia
mIMCD3
MIN6
Neuro-2a
NIH 3T3
NK cells
Nucleus accumbens
Olfactory bulb
Osteoblast day 5
Osteoblast day 14
Osteoblast day 21
Osteoclasts
Ovary
Pancreas
Pituitary
Placenta
Prostate
RAW 264.7
Retina
Retinal pigment epithelium
Salivary gland
Skeletal muscle
Spinal cord
Spleen
Stem cells HSC
Stomach
T-cells CD4+
T-cells CD8+
T-cells FOXP3+
Testis
Thymocyte DP CD4+, CD8+
Thymocyte SP CD4+
Thymocyte SP CD8+
Umbilical cord
Uterus
0
500
1000
1500
2000
2500
3000
Mouse Dhx36 mRNA expression

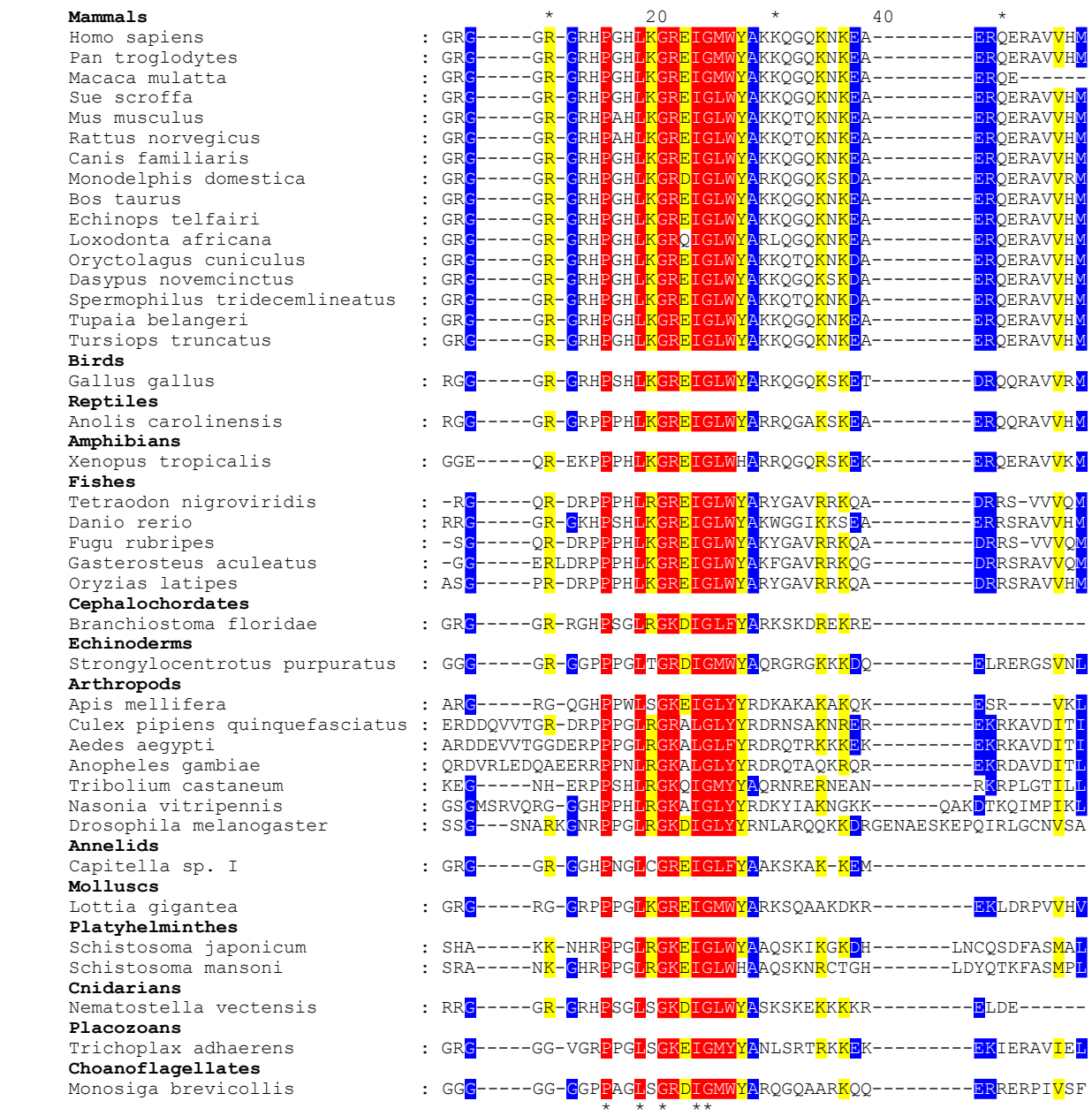

C
Adult carcass
Brain
Crop
Eye
Fat body
Head
Heart
Hindgut
Larval carcass
Larval CNS
Larval fat body
Larval hindgut
Larval midgut
Larval salivary gland
Larval trachea
Larval tubule
Male accessory glands
Mated spermatheca
Midgut
Ovary
S2 cells (growing)
Salivary gland
Testis
Thoracicoabdominal ganglion
Tubule
Virgin spermatheca
Whole fly
0
200
400
600
800
Drosophila CG9323 mRNA expression

Supplementary Figure 7 | **Pairwise alignment of human RHAU and its mouse orthologue**. Sequence alignment was carried out with MAFFT (version 6, ref. 313). Similarity analysis was performed with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Identical and similar amino acids are shaded in black and grey, respectively.

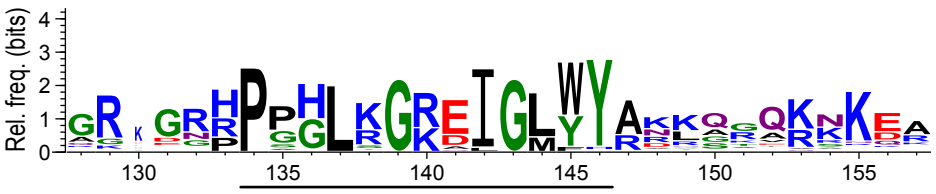


Supplementary Figure 8 | **Recombinant RHAU protein purification**. SDS-PAGE separation and Coomassie staining of purified GST-tagged recombinant wild-type (WT) and N-terminal fragment (1–200) of RHAU (2 µg protein per lane). The positions and sizes (kDa) of marker proteins are indicated at the left. The mark (◀) on the right indicates the position of the purified proteins.

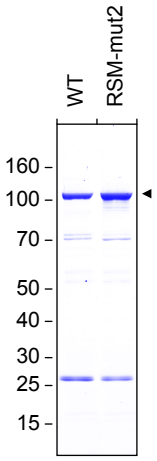
A



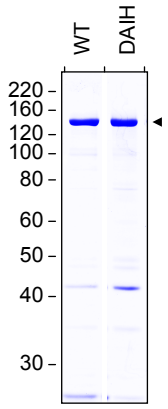
B



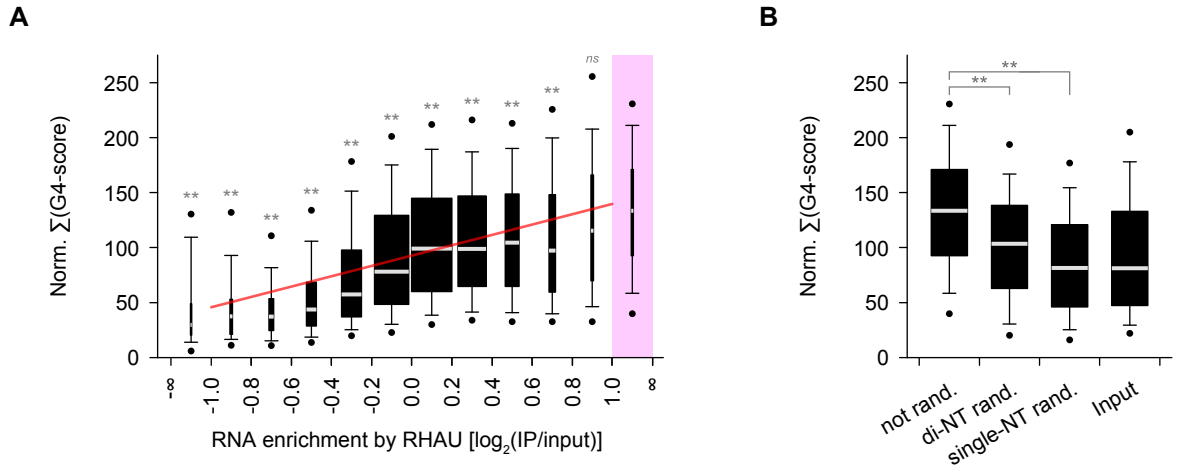
Supplementary Figure 9 | **Conserved residues within the RSM region**. (**A**) Sequence alignment of the RSM-containing region of 40 different RHAU orthologues. Multiple sequence alignment was carried out with MAFFT (version 6, ref. 313). Similarity analysis was made by GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Similarity is shown in red for 100 %, yellow for 99–80 % and blue for 79–60 %. Amino acids that are identical among all sequences are indicated by asterisks below the alignment. The RSM is underlined. (**B**) Sequence logo (371) derived from the multiple sequence alignment in (A). Amino acids are coloured according to their biochemical properties: green for polar (G, S, T, Y, C, Q, N), blue for basic (K, R, H), red for acidic (D, E) and black for hydrophobic (A, V, L, I, P, W, F, M). The RSM is underlined.



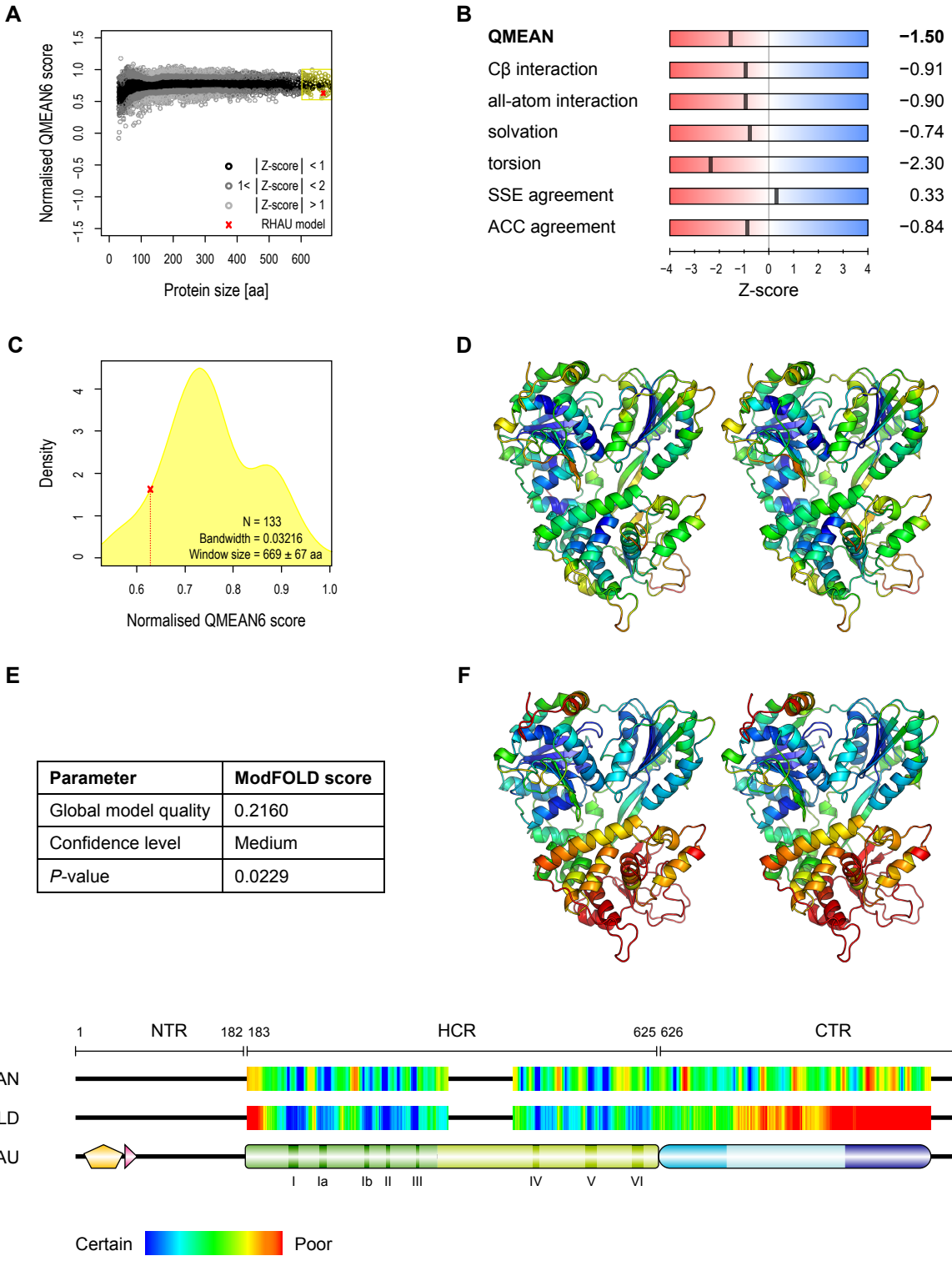
Supplementary Figure 10 | **Recombinant CG9323 protein purification**. SDS-PAGE separation and Coomassie staining of purified FLAG-tagged recombinant wild-type (WT) and RSM mutant (RSM-mut2) CG9323 protein (2 µg protein per lane). The positions and sizes (kDa) of marker proteins are indicated at the left. The mark (◀) on the right indicates the position of the purified proteins.



Supplementary Figure 11 | **Recombinant RHAU protein purification**. SDS-PAGE separation and Coomassie staining of purified GST-tagged recombinant wild-type (WT) and ATPase-deficient mutant (DAIH) RHAU proteins (2 µg protein per lane). The positions and sizes (kDa) of marker proteins are indicated at the left. The mark (◀) on the right indicates the position of the purified protein.



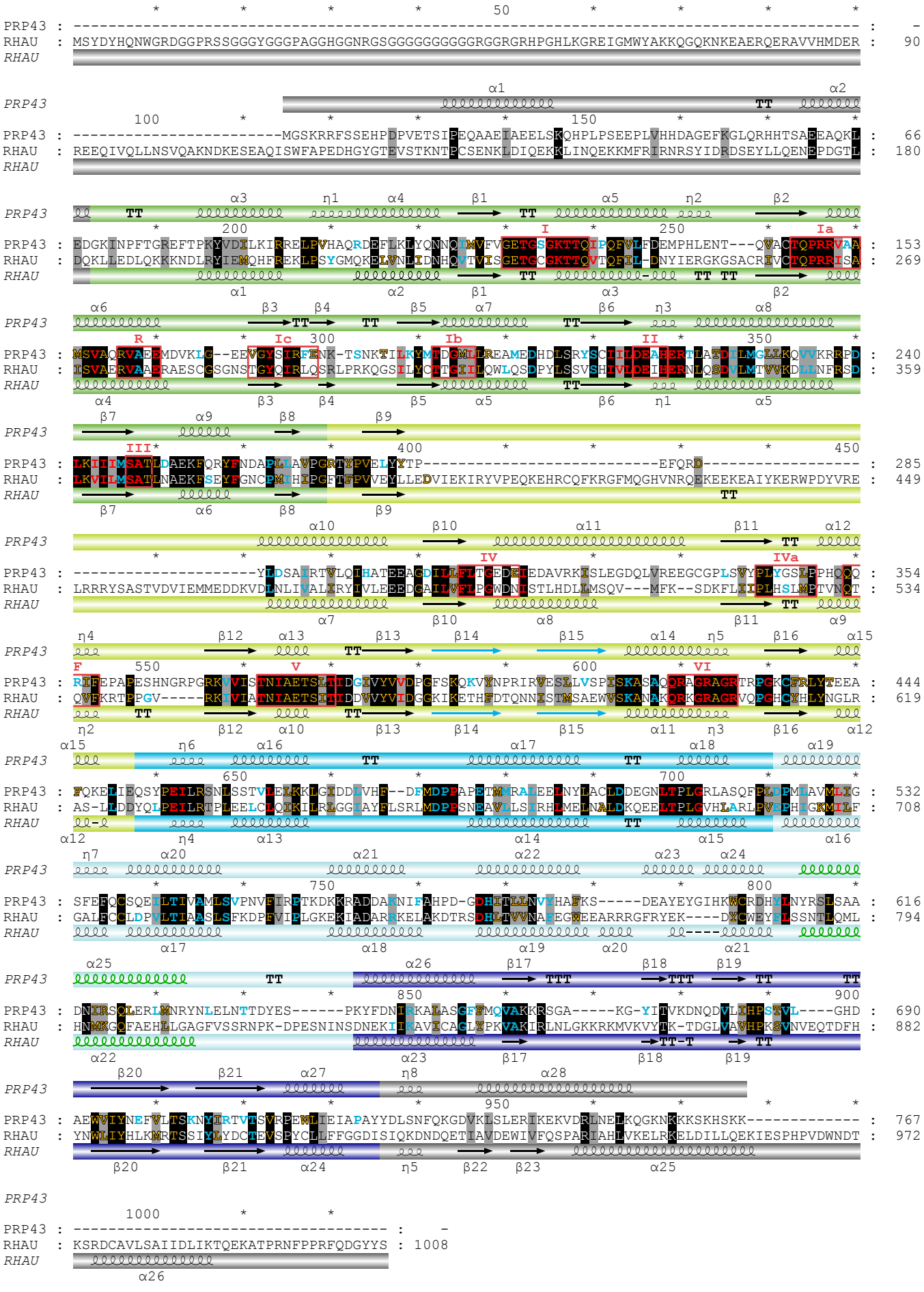
Supplementary Figure 12 | **Computational analysis of potential intramolecular G4-forming sequences (PQS) among RNAs enriched by RHAU**. (**A**) Box plot representation of the normalised ∑(G4-score) value per transcript as a function of their level of enrichment by RHAU. The pink rectangle refers to the group of 106 RNAs specifically enriched by RHAU and serves as a control group for multiple comparisons analysis. Statistical significance was determined by Kruskal-Wallis one-way ANOVA on ranks and the Dunn's test. $^{ns}P \geq 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$. The correlation between the two variables was estimated by a linear regression model ($y = 47x + 93$, $R^2 = 0.10$, $P < 0.001$) and is indicated as a red line on the graph. (**B**) PQS analysis among randomised RHAU target sequences. The sequences of the 106 transcripts specifically enriched by RHAU were shuffled, retaining single- or di-nucleotide base composition of the original transcripts. The box plot represents the normalised ∑(G4-score) value per transcript among the original (not rand.), di-nucleotide shuffled (di-NT rand.) and single-nucleotide shuffled (single-NT rand.) sequences. Statistical significance was determined by Friedman repeated measures ANOVA on ranks and the Dunnett's test.



Parameter	ModFOLD score
Global model quality	0.2160
Confidence level	Medium
P-value	0.0229

Supplementary Figure 13 | **Quality prediction for RHAU 3-D model**. Quality assessment of the modelled RHAU structure by QMEAN6 (**A–D**, **G**) and ModFOLD (**E–G**) algorithms. (**A**) Comparison of the normalised QMEAN score value for RHAU structural model with non-redundant set of PDB structures. (**B**) Absolute QMEAN score value together with the six scoring function terms contributing to the QMEAN score. A distance-dependent pairwise Cβ potential as well as an all-atom potential with 167 atom types are used to assess long-range interactions. The solvation potential describes the burial status of the residues. The local geometry is analysed by a torsion angle potential over three consecutive amino acids. 'SSE agreement' and 'ACC agreement' scores describe respectively the agreement of predicted and calculated secondary structure and solvent accessibility. High-resolution X-ray structures on average have a QMEAN score equal to zero. (**C**) Density plot visualising the normalised QMEAN6 score distribution of PDB structures ranging from 602 to 736 amino acids. The window size (660 ± 67 amino acids, N = 133) corresponds to the yellow rectangle depicted in (**A**). The normalised QMEAN6 score for RHAU structural model is indicated in red. (**D**) RHAU structural model coloured according to local error prediction by QMEAN algorithm. (**E**) Estimation of RHAU structural quality by ModFOLD algorithm. (**F**) RHAU structural model coloured according to local error prediction by ModFOLD algorithm. (**G**) Linear and coloured representation of the local error prediction for RHAU structural model.

Supplementary Figure 14 ► | **Sequence and structural similarities between Prp43 and RHAU structural model**. The structure based sequence alignment of the helicase core and HA regions of Prp43 (RCSB PDB ID: 2XAU) with RHAU structural model (Figure 38) was performed with STRAP (368,369). Similarity analysis was performed with GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Identical and similar amino acids are shaded in black and grey, respectively. Coloured amino acids in Prp43 and RHAU sequences denote reciprocal conserved residues in paralogous DEAH-box proteins (red, 100 % similarity; yellow, 99–80 % similarity; blue, 79–60 % similarity; cf. Supplementary Figure 1). The RecA2 β-hairpin and the ratchet helix are depicted in blue and green, respectively.


PRP43
RHAU
MSYDYHQNWGRDGGPRSSGGGYGGGPAGGHGGNRGSGGGGGGGGGGRGGRGRHPGHLKGREIGMWYAKKQGQKNKEAERQERAVVHMDER
REEQIVQLLNSVQAKNDKESEAQISWFAPEDHGYGTEVSTKNTPCSENKLDIQEKKLINQEKKMFRIRNRSYIDRDSEYLLQENEPDGTL
MGSKRRFSSEHPDPVETSIPEQAAETAEELSKQHPLPSEEPLVHHDAGEFKGLQRHHTSAEEAQKL
EDGKINPFTGREFTPKYVDILKIRRELPVHAQRDEFLKLYQNNQIMVFVGETCSGKTTQIPQFVLEDEMPHLENT---QVACTQPRRVAA
DQKLLEDLQKKKNDLRYIEMQHFREKLPSYGMQKELVNLIDNHQVTVISGETCCGKTTQVTQFILLDNYIERGKGSACRIVCTQPRRISA
MSVAQRVAEEMDVKLG--EEVGYSIRFENK-TSNKTILKYMTDGMLLREAMEDHDLSRYSCIILDEAHERTLATDVLMGLLKQVVKKRPD
ISVAERVAAERAESCGSNSTGYQIRLQSRLPRKQGSILYCTTGILLQWLQSDPYLSSVSHIVLDEIHERNLQSDVLMTVVKDLLNFRSD
LKIILMSATLDAEKFQRYFNDAPLLAVPGRTYPVELYYTP----------EFQRD
LKVILMSATLNAEKFSEYFGNCPMIHIPGFTFPVVEYLLEDVIEKIRYVPEQKEHRCQFKRGFMQGHVNRQEKEEKEAIYKERWPDYVRE
YLDSATRTVLQIHATEEAGDILLFLTGEDEIEDAVRKISLEGDQLVREEGCGPLSVYPLYGSLPPHQQQ
LRRRYSASTVDVIEMMEDDKVDLNLIVALIRYIVLEEEDGAIIVFLPGWDNISTLHDLLMSQV---MFK--SDKFLIIPLHSLMPTVNQT
RIFEPAPESHNGRPGRKVVISTNIAETSLTIDGIVYVVDPGFSKQKVYNPRIRVESLLVSPISKASAQQRAGRAGRTRPGKCFRLYTEEA
QVFKRTPPGV-----RKTVIATNIAETSITIDDVVYVIDGGKIKETHFDTQNNISTMSAEWVSKANAKQRKGRAGRVQPGHCYHLYNGLR
FQKELIEQSYPEILRSNLSSTVLELKKLGIDDLVHF--DFMDPPAPETMMRALEELNYLACLDDEGNLTPLGRLASQFPLDPMLAVMLIG
AS-LLDDYQLPEILRTPLEELCLQIKILRLGGIAYFLSRLMDPPSNEAVLLSIRHLMELNALDKQEELTPLGVHLARLPVEPHIGKMILF
SFEFQCSQEILTIVAMLSVPNVFIRPTKDKKRADDAKNIFAHPD-GDHITLLNVVHAFKS-----DEAYEYGIHKWCRDHYLNYRSLSAA
GALFCCLDPVLTIAASLSFKDPFVIPLGKEKIADARKELAKDTRSDHLTVVNAFEGWEEARRRGFRYEK----DYCWEYFLSSNTLQML
DNIRSQLERLMNRYNLELNTTDYES------PKYFDNIRKALASGFFMQVAKKRSGA-----KG-YITVKDNQDVLIHPSTVL----GHD
HNMKCQFAEHLLGAGFVSSRNPK-DPESNINSDNEKIIKAVICAGLYPKVAKIRLNLGKKRKMVKVYTK-TDGLVAVHPKSVNVEQTDFH
AEWVIYNEFVLTSKNYIRTVTSVRPEWLIEIAPAYYDLSNFQKGDVKLSLERIKEKVDRLNELKQGKNKKKSKHSKK
YNWLIYHLKMRTSSIYLYDCTEVSPYCLLFFGGDISIQKDNDQETIAVDEWIVFQSPARIAHLVKELRKELDILLQEKIESPHPVDWNDT
PRP43
KSRDCAVLSAIIDLIKTQEKATPRNFPPRFQDGYYS
1000
1008

Supplementary Table I | **List of oligodeoxynucleotides employed for RT-PCR and TRAP assays**.

Gene name	Oligo ID	Oligo name	Sequence (5' - 3')
ACTB	166	ACTB_Fw	CAC CCC GTG CTG CTG ACC
	167	ACTB_Rv	CCA GAG GCG TAC AGG GAT AGC
ARL3	244	ARL3_Fw	GAA GAG ACG GGT CAG GAA CTA GC
	245	ARL3_Rv	GAC TCG GTC GCG GAT GGT ATG
CEP57	246	CEP57_Fw	CAT GGT TCG GCA TTC TTC ATC TCC
	247	CEP57_Rv	TCT GCC TGA ATC CTC TCA AGT TCC
DYNLL2	248	DYNLL2_Fw	ACA CAC GAG ACA AAG CAC TTC ATC
	249	DYNLL2_Rv	AGT AGC ACA GCC ATT GGA GAG G
GAPDH	162	GAPDH_Fw	CGC TCT CTG CTC CTC CTG TTC
	163	GAPDH_Rv	CGC CCA ATA CGA CCA AAT CCG
HBXIP	200	HBXIP_Fw	CCT CTG ACC CCA CTG ATA TTC CTG
	201	HBXIP_Rv	CAC CGT GAT GCC ATC GTG TTT C
IGFBP6	230	IGFBP6_Fw	GGA ATC CAG GCA CCT CTA CCA C
	231	IGFBP6_Rv	TGA GTC CAG ATG TCT ACG GCA TG
MAGOH	224	MAGOH_Fw	TGG GGC ACA AGG GCA AGT TC
	225	MAGOH_Rv	CTT CCC GTC CGG TCG AAA CTC
PAAF1	234	PAAF1_Fw	CCA AGG GCT CCG GTA CAA GTC
	235	PAAF1_Rv	ATC TGC TTC TCC CAT GTG GCT AC
PABPC1	250	PABPC1_Fw	AGA ACC GTG CTG CAT ACT ATC CTC
	251	PABPC1_Rv	TGA CTC GTG GAA CCT GTG AAG AAG
TERC	266	TERC_Fw	GGC GAG GGC GAG GTT CAG
	267	TERC_Rv	GCA CGT CCC ACA GCT CAG G
TMEM55B	238	TMEM55B_Fw	CCT TGC CTT TGG CAC ATG GAA G
	239	TMEM55B_Rv	TCA TCA GGC TCA GGA GAA GTT CTG
U2AF1	260	U2AF1_Fw	AAG ATG CGG AAA AGG CTG TGA TTG
	261	U2AF1_Rv	TCT CTG GAA ATG GGC TTC AAA TGC
WBP4	240	WBP4_Fw	GTA ACC AGC ACT ATC CCA CCT ACC
	241	WBP4_Rv	TTT CTC CCA CTG AGA TGC TCC TG
TRAP assay	270	TS	AAT CCG TCG AGC AGA GTT
	271	Cy5-TS	*Cy5*-AAT CCG TCG AGC AGA GTT
	273	ACX	GCG CGG CTT ACC CTT ACC CTT ACC CTA ACC

Supplementary Table II | **List of RNAs significantly enriched by RHAU in HeLa cells**. The enrichement value is calculated as the $\log_2$ of the ratio $IP_{FLAG}/input_{FLAG}$. The $N_{G4}(2)$, $N_{G4}(3)$ and $N_{G4}(4)$ fields correspond to the number of predicted G4 composed of 2, 3 or 4 succesive G-quartets, respectively. For RNAs bearing $N_{G4}(3)$ and $N_{G4}(4)$, their location in the transcript is indicated (5'-UTR, CDS, 3'-UTR). Transcripts whose enrichment by RHAU has been validated by RT-PCR are coloured in red. Transcripts whose enrichment by RHAU has been observed in both HeLa and HEK293T cell lines are preceded by a full black dot.

Rk	Accession	Gene	Definition	Enrichment ($\log_2$)	G4 density [PQS·kb^{-1}]	Normalised ∑(G4-score)	∑(G4-score)	$N_{G4}(2)$	$N_{G4}(3)$	$N_{G4}(4)$	5'-UTR	CDS	3'-UTR
1	NR_001566	• TERC	Telomerase RNA component	1.930	11.1	252.8	114	4	1	0			
2	NM_001100880	• ST20	Suppressor of tumorigenicity 20	1.870	10.9	192.2	123	7	0	0			
3	NM_025155	• PAAF1	Proteasomal ATPase-associated factor 1	1.756	8.2	126.8	201	13	0	0			
4	NM_018320	RNF121	Ring finger protein 121	1.706	8.0	130.8	343	19	2	0			2
5	NM_001017979	RAB28	RAB28, member RAS oncogene family	1.628	4.0	76.8	133	6	1	0	1		
6	NM_018669	WDR4	WD repeat domain 4	1.625	8.4	141.1	304	17	1	0	1		
7	NM_006402	• HBXIP	Hepatitis B virus x interacting protein	1.586	7.0	90.1	77	6	0	0			
8	NM_017410	HOXC13	Homeobox C13	1.583	7.8	137.2	334	18	1	0	1		
9	NM_031209	QTRT1	Queuine tRNA-ribosyltransferase 1	1.581	10.4	166.4	225	13	1	0			1
10	XM_497450	LOC441722	PREDICTED: similar to U2 small nuclear RNA auxillary factor 1	1.555	8.9	168.9	152	7	1	0	1		
11	NM_001040056	MAPK3	Mitogen-activated protein kinase 3	1.548	10.0	232.9	467	15	4	1	2	2	1
12	NM_001025204	U2AF1	U2 small nuclear RNA auxiliary factor 1	1.544	7.7	132.0	137	8	0	0			
13	NM_004317	• ASNA1	ArsA arsenite transporter, ATP-binding, homolog 1	1.540	7.7	151.8	197	8	2	0		1	1
14	NM_020395	INTS12	Integrator complex subunit 12	1.528	2.8	50.3	89	5	0	0			
15	NM_203472	• SELS	Selenoprotein S	1.515	10.1	172.6	223	13	0	0			
16	NM_020179	• C11orf75	Chromosome 11 open reading frame 75	1.507	9.2	138.9	136	9	0	0			
17	NM_001077268	ZFYVE19	Zinc finger, FYVE domain containing 19	1.475	10.0	174.4	400	21	2	0		2	
18	NM_005452	• WDR46	WD repeat domain 46	1.474	10.2	164.6	389	24	0	0			
19	NM_004311	ARL3	ADP-ribosylation factor-like 3	1.465	6.2	107.7	419	22	2	0			2
20	NM_007187	WBP4	WW domain binding protein 4 (formin binding protein 21)	1.458	3.4	54.4	128	7	1	0	1		
21	NM_005308	GRK5	G protein-coupled receptor kinase 5	1.442	8.2	142.5	367	20	1	0			1
22	NM_016586	MBIP	MAP3K12 binding inhibitory protein 1	1.435	1.8	33.1	55	3	0	0			
23	BC024237	• C22orf32	Chromosome 22 open reading frame 32	1.421	4.4	98.0	155	6	1	0	1		
24	NM_133491	• SAT2	Spermidine/spermine N1-acetyltransferase family member 2	1.417	11.2	170.9	168	11	0	0			
25	NM_022044	• SDF2L1	Stromal cell-derived factor 2-like 1	1.376	16.0	294.5	258	13	1	0	1		
26	NM_014184	CNIH4	Cornichon homolog 4	1.350	3.1	64.5	42	2	0	0			
27	NM_004450	ERH	Enhancer of rudimentary homolog	1.337	5.0	118.6	95	3	1	0			1
28	NM_033518	SLC38A5	Solute carrier family 38, member 5	1.328	8.0	145.6	292	15	1	0		1	
29	NM_001861	COX4I1	Cytochrome c oxidase subunit IV isoform 1	1.326	8.7	135.9	109	7	0	0			
30	NM_013320	HCFC2	Host cell factor C2	1.320	2.6	36.0	207	14	1	0	1		
31	NM_004231	ATP6V1F	ATPase, H+ transporting, lysosomal 14kDa, V1 subunit F	1.318	11.6	204.1	141	8	0	0			
32	NM_001350	DAXX	Death-domain associated protein	1.310	7.2	100.4	264	19	0	0			
33	NM_014408	TRAPPC3	Trafficking protein particle complex 3	1.293	7.0	118.9	152	8	1	0			1
34	NM_001100814	• TMEM55B	Transmembrane protein 55B	1.292	9.2	202.3	350	13	3	0		1	2
35	NM_013330	• NME7	Non-metastatic cells 7	1.279	3.0	51.6	85	5	0	0			
36	NM_015670	SENP3	SUMO1/sentrin/SMT3 specific peptidase 3	1.273	9.6	176.1	440	23	1	0		1	
37	NM_003768	PEA15	Phosphoprotein enriched in astrocytes 15	1.273	6.1	139.0	344	12	2	1			3

Rk	Accession	Gene	Definition	Enrichment (log_2)	G4 density [PQS·kb^{-1}]	Normalised ∑(G4-score)	∑(G4-score)	$N_{G4}(2)$	$N_{G4}(3)$	$N_{G4}(4)$	5'-UTR	CDS	3'-UTR
38	NM_022916	VPS33A	Vacuolar protein sorting 33 homolog A	1.269	4.2	65.4	172	11	0	0			
39	NM_016020	TFB1M	Transcription factor B1, mitochondrial	1.249	3.9	59.0	76	5	0	0			
40	NM_005466	• MED6	Mediator complex subunit 6	1.247	3.7	72.9	98	5	0	0			
41	NM_002640	SERPINB8	Serpin peptidase inhibitor, clade B (ovalbumin), member 8	1.245	3.2	46.9	159	11	0	0			
42	NM_024671	ZNF768	Zinc finger protein 768	1.237	9.9	218.7	510	18	4	1	1	1	3
43	NM_030815	• PDRG1	P53 and DNA-damage regulated 1	1.234	8.7	135.1	187	12	0	0			
44	NM_004374	• COX6C	Cytochrome c oxidase subunit VIc	1.225	4.3	69.5	64	4	0	0			
45	NM_007241	• SNF8	SNF8, ESCRT-II complex subunit, homolog	1.220	9.0	166.4	204	10	1	0	1		
46	NM_002178	IGFBP6	Insulin-like growth factor binding protein 6	1.218	10.2	194.9	191	9	1	0		1	
47	NM_006693	CPSF4	Cleavage and polyadenylation specific factor 4, 30kDa	1.212	7.7	134.0	245	13	1	0			1
48	NM_003341	UBE2E1	Ubiquitin-conjugating enzyme E2E 1 (UBC4/5 homolog, yeast)	1.208	3.4	94.0	139	3	1	1	1		1
49	NM_173852	KRTCAP2	Keratinocyte associated protein 2	1.207	5.2	114.4	66	2	1	0		1	
50	NM_014017	ROBLD3	Roadblock domain containing 3	1.207	11.4	220.3	154	8	0	0			
51	NM_018622	PARL	Presenilin associated, rhomboid-like	1.204	7.0	116.0	165	10	0	0			
52	NM_003977	• AIP	Aryl hydrocarbon receptor interacting protein	1.200	9.6	139.2	174	12	0	0			
53	BC035511		Family with sequence similarity 100, member B, mRNA	1.193	9.8	189.3	270	12	2	0			2
54	NM_001261	CDK9	Cyclin-dependent kinase 9	1.188	11.3	205.9	509	27	1	0			1
55	NM_000713	BLVRB	Biliverdin reductase B (flavin reductase (NADPH))	1.184	8.0	137.3	120	7	0	0			
56	NM_004240	TRIP10	Thyroid hormone receptor interactor 10	1.183	8.9	154.8	314	16	2	0	1	1	
57	NM_001924	GADD45A	Growth arrest and DNA-damage-inducible, alpha	1.179	6.6	128.4	174	9	0	0			
58	NM_007182	RASSF1	Ras association RalGDS/AF-6 domain family member 1	1.178	11.2	173.3	341	21	1	0		1	
59	NM_005716	GIPC1	GIPC PDZ domain containing family, member 1	1.164	14.4	279.9	545	24	4	0		2	2
60	NM_174889	NDUFAF2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, assembly factor 2	1.161	8.3	134.9	97	6	0	0			
61	BC000209	C1orf135	Chromosome 1 open reading frame 135, mRNA	1.160	4.6	73.9	97	6	0	0			
62	NM_022077	• MANBAL	Mannosidase, beta A, lysosomal-like	1.146	4.3	93.1	129	5	1	0			1
63	NM_080677	DYNLL2	Dynein, light chain, LC8-type 2	1.145	11.2	238.5	363	14	3	0	1		2
64	NM_015846	MBD1	Methyl-CpG binding domain protein 1	1.143	8.2	127.6	419	27	0	0			
65	NM_002931	• RING1	Ring finger protein 1	1.142	10.3	226.5	395	14	4	0		4	
66	NM_013290	PSMC3IP	PSMC3 interacting protein	1.131	6.8	119.9	158	9	0	0			
67	NM_005089	ZRSR2	CCCH type zinc finger (RNA-binding motif and serine/arginine rich 2)	1.129	7.9	133.6	202	12	0	0			
68	NM_173685	NSMCE2	Non-SMC element 2, MMS21 homolog	1.112	2.4	35.9	44	3	0	0			
69	NM_014599	MAGED2	Melanoma antigen family D, 2	1.111	9.6	159.6	332	20	0	0			
70	NM_000386	BLMH	Bleomycin hydrolase	1.106	4.1	64.1	155	10	0	0			
71	NM_182705	• FAM101B	Family with sequence similarity 101, member B	1.101	6.2	108.1	421	22	2	0			2
72	NM_005949	MT1F	Metallothionein 1F	1.100	2.2	35.1	16	1	0	0			
73	NM_004843	IL27RA	Interleukin 27 receptor, alpha	1.096	10.1	180.8	589	32	1	0	1		
74	NM_005981	TSPAN31	Tetraspanin 31	1.090	5.1	90.9	159	9	0	0			
75	XM_929793	LOC646836	PREDICTED: uncharacterized protein LOC646836-like	1.090	3.9	81.4	21	1	0	0			
76	NM_020137	GRIPAP1	GRIP1 associated protein 1	1.090	10.2	182.4	553	28	3	0	1		2
77	NM_006321	• ARIH2	Ariadne homolog 2	1.084	6.1	101.0	412	25	0	0			
78	NM_002133	• HMOX1	Heme oxygenase (decycling) 1	1.083	7.5	107.7	173	12	0	0			

Rk	Accession	Gene	Definition	Enrichment (log_2)	G4 density [PQS·kb^{-1}]	Normalised ∑(G4-score)	∑(G4-score)	$N_{G4}(2)$	$N_{G4}(3)$	$N_{G4}(4)$	5'-UTR	CDS	3'-UTR
79	NM_018178	GOLPH3L	Golgi phosphoprotein 3-like	1.081	0.9	15.9	51	3	0	0			
80	NM_007277	EXOC3	Exocyst complex component 3	1.070	7.6	117.5	326	21	0	0			
81	NM_024408	• NOTCH2	Notch 2	1.068	4.0	58.1	664	46	0	0			
82	NM_004346	CASP3	Caspase 3, apoptosis-related cysteine peptidase	1.065	4.8	70.3	189	13	0	0			
83	NM_145058	RILPL2	Rab interacting lysosomal protein-like 2	1.064	9.8	189.5	272	13	1	0		1	
84	NM_006555	YKT6	YKT6 v-SNARE homolog	1.056	10.1	189.7	528	24	4	0	1		3
85	NM_006010	MANF	Mesencephalic astrocyte-derived neurotrophic factor	1.055	8.3	148.6	144	8	0	0			
86	NM_021934	• C12orf44	Chromosome 12 open reading frame 44	1.051	12.3	210.6	307	18	0	0			
87	NM_002882	• RANBP1	RAN binding protein 1	1.046	6.8	117.6	104	5	1	0	1		
88	NM_002370	MAGOH	Mago-nashi homolog, proliferation-associated	1.045	5.7	113.2	79	3	1	0			1
89	NM_003427	ZNF76	Zinc finger protein 76	1.043	9.0	133.3	356	23	1	0	1		
90	NM_181471	• RFC2	Replication factor C (activator 1) 2, 40 kDa	1.043	7.6	140.5	241	13	0	0			
91	NM_014026	• DCPS	Decapping enzyme, scavenger	1.037	8.6	171.8	259	12	1	0	1		
92	NM_015169	RRS1	RRS1 ribosome biogenesis regulator homolog	1.037	12.8	212.9	367	21	1	0		1	
93	NM_015378	• VPS13D	Vacuolar protein sorting 13 homolog D	1.035	4.0	61.6	1007	64	2	0	1		1
94	NM_138463	TLCD1	TLC domain containing 1	1.029	7.7	112.8	117	8	0	0			
95	NM_001423	EMP1	Epithelial membrane protein 1	1.028	3.9	77.7	218	9	1	1			2
96	NM_015057	• MYCBP2	MYC binding protein 2	1.026	3.9	59.4	892	59	0	0			
97	NM_024936	ZCCHC4	Zinc finger, CCHC domain containing 4	1.023	3.6	58.6	165	9	1	0		1	
98	NM_014282	HABP4	Hyaluronan binding protein 4	1.015	8.9	158.8	430	23	1	0	1		
99	NM_030645	SH3BP5L	SH3-binding domain protein 5-like	1.013	12.3	222.5	708	36	3	0	1	1	1
100	NM_182710	KAT5	Lysine acetyltransferase 5	1.012	9.4	164.5	384	21	1	0		1	
101	NM_019896	• POLE4	Polymerase (DNA-directed), epsilon 4 (p12 subunit)	1.009	5.6	102.8	73	4	0	0			
102	NM_213649	• SFXN4	Sideroflexin 4	1.007	6.4	94.3	132	9	0	0			
103	NM_002810	PSMD4	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 4	1.005	9.8	151.7	202	13	0	0			
104	NM_002812	• PSMD8	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 8	1.004	7.1	122.8	191	10	1	0			1
105	NM_024294	C6orf106	Chromosome 6 open reading frame 106	1.003	8.1	140.9	624	34	2	0	1		1
106	NM_032758	PHF5A	PHD finger protein 5A	1.002	6.3	126.7	140	6	1	0			1
										TOTAL	25	24	43

CHAPTER 7

References

1. Gu, J. and Reddy, R. (2001) **Cellular RNAs: Varied Roles.** *Encyclopedia of Life Sciences*, 1-7. Full Text
2. Schroeder, R., Barta, A. and Semrad, K. (2004) **Strategies for RNA folding and assembly.** *Nat Rev Mol Cell Biol*, **5**, 908-919. Abstract Full Text
3. Russell, R. (2008) **RNA misfolding and the action of chaperones.** *Front Biosci*, **13**, 1-20. Abstract Full Text
4. Lorsch, J.R. (2002) **RNA chaperones exist and DEAD box proteins get a life.** *Cell*, **109**, 797-800. Abstract Full Text
5. Silverman, E., Edwalds-Gilbert, G. and Lin, R.J. (2003) **DExD/H-box proteins and their partners: helping RNA helicases unwind.** *Gene*, **312**, 1-16. Abstract Full Text
6. Hilbert, M., Karow, A.R. and Klostermeier, D. (2009) **The mechanism of ATP-dependent RNA unwinding by DEAD box proteins.** *Biol Chem*, **390**, 1237-1250. Abstract Full Text
7. Linder, P. (2006) **Dead-box proteins: a family affair--active and passive players in RNP-remodeling.** *Nucleic Acids Res*, **34**, 4168-4180. Abstract Full Text
8. Jankowsky, E. and Bowers, H. (2006) **Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases.** *Nucleic Acids Res*, **34**, 4181-4188. Abstract Full Text
9. Anantharaman, V., Koonin, E.V. and Aravind, L. (2002) **Comparative genomics and evolution of proteins involved in RNA metabolism.** *Nucleic Acids Res*, **30**, 1427-1464. Abstract Full Text
10. Shiratori, A., Shibata, T., Arisawa, M., Hanaoka, F., Murakami, Y. and Eki, T. (1999) **Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related proteins in Saccharomyces cerevisiae by gene disruption and Northern analysis.** *Yeast*, **15**, 219-253. Abstract Full Text
11. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) **The sequence of the human genome.** *Science*, **291**, 1304-1351. Abstract Full Text
12. Gorbalenya, A.E. and Koonin, E.V. (1993) **Helicases: amino acid sequence comparisons and structure-function relationships.** *Curr Opin Struct Biol*, **3**, 419-429. Abstract Full Text
13. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) **Structure and mechanism of helicases and nucleic acid translocases.** *Annu Rev Biochem*, **76**, 23-50. Abstract Full Text
14. Fairman-Williams, M.E., Guenther, U.P. and Jankowsky, E. (2010) **SF1 and SF2 helicases: family matters.** *Curr Opin Struct Biol*, **20**, 313-324. Abstract Full Text
15. Durr, H., Flaus, A., Owen-Hughes, T. and Hopfner, K.P. (2006) **Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures.** *Nucleic Acids Res*, **34**, 4160-4167. Abstract Full Text
16. Tuteja, N. and Tuteja, R. (2004) **Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery.** *Eur J Biochem*, **271**, 1835-1848. Abstract Full Text
17. Tuteja, N. and Tuteja, R. (2004) **Unraveling DNA helicases. Motif, structure, mechanism and function.** *Eur J Biochem*, **271**, 1849-1863. Abstract Full Text
18. Caruthers, J.M. and McKay, D.B. (2002) **Helicase structure and mechanism.** *Curr Opin Struct Biol*, **12**, 123-133. Abstract Full Text
19. Patel, S.S. and Picha, K.M. (2000) **Structure and function of hexameric helicases.** *Annu Rev Biochem*, **69**, 651-697. Abstract Full Text
20. Enemark, E.J. and Joshua-Tor, L. (2008) **On helicases and other motor proteins.** *Curr Opin Struct Biol*, **18**, 243-257. Abstract Full Text
21. Bachrati, C.Z. and Hickson, I.D. (2008) **RecQ helicases: guardian angels of the DNA replication fork.** *Chromosoma*, **117**, 219-233. Abstract Full Text
22. Chu, W.K. and Hickson, I.D. (2009) **RecQ helicases: multifunctional genome caretakers.** *Nat Rev Cancer*, **9**, 644-654. Abstract Full Text
23. Wu, Y., Suhasini, A.N. and Brosh, R.M., Jr. (2009) **Welcome the family of FANCJ-like helicases to the block of genome stability maintenance proteins.** *Cell Mol Life Sci*, **66**, 1209-1222. Abstract Full Text
24. Kokavec, J., Podskocova, J., Zavadil, J. and Stopka, T. (2008) **Chromatin remodeling and SWI/SNF2 factors in human disease.** *Front Biosci*, **13**, 6126-6134. Abstract Full Text
25. Bleichert, F. and Baserga, S.J. (2007) **The long unwinding road of RNA helicases.** *Mol Cell*, **27**, 339-352. Abstract Full Text
26. de la Cruz, J., Kressler, D. and Linder, P. (1999) **Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families.** *Trends Biochem Sci*, **24**, 192-198. Abstract Full Text
27. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. and Yokoyama, S. (2006) **Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa.** *Cell*, **125**, 287-300. Abstract Full Text
28. Linder, P. and Jankowsky, E. (2011) **From unwinding to clamping - the DEAD box RNA helicase family.** *Nat Rev Mol Cell Biol*, **12**, 505-516. Abstract Full Text
29. Buttner, K., Nehring, S. and Hopfner, K.P. (2007) **Structural basis for DNA duplex separation by a superfamily-2 helicase.** *Nat Struct Mol Biol*, **14**, 647-652. Abstract Full Text
30. He, Y., Andersen, C.B. and Nielsen, K.H. (2011) **The function and architecture of DEAH/RHA helicases.** *BioMolecular Concepts*, **2**, 315-326. Full Text
31. Cordin, O., Banroques, J., Tanner, N.K. and Linder, P. (2006) **The DEAD-box protein family of RNA helicases.** *Gene*, **367**, 17-37. Abstract Full Text
32. Tran, H., Schilling, M., Wirbelauer, C., Hess, D. and Nagamine, Y. (2004) **Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU.** *Mol Cell*, **13**, 101-111. Abstract Full Text
33. Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., Nagamine, Y. and Akman, S.A. (2005) **The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates.** *J Biol Chem*, **280**, 38117-38120. Abstract Full Text
34. Sexton, A.N. and Collins, K. (2011) **The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation.** *Mol Cell Biol*, **31**, 736-743. Abstract Full Text
35. Harosh, I. and Deschavanne, P. (1991) **The RAD3 gene is a member of the DEAH family RNA helicase-like protein.** *Nucleic Acids Res*, **19**, 6331. Abstract Full Text
36. Company, M., Arenas, J. and Abelson, J. (1991) **Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes.** *Nature*, **349**, 487-493. Abstract Full Text
37. Tanner, N.K. and Linder, P. (2001) **DExD/H box RNA helicases: from generic motors to specific dissociation functions.** *Mol Cell*, **8**, 251-262. Abstract Full Text
38. Zhang, S. and Grosse, F. (1994) **Nuclear DNA helicase II unwinds both DNA and RNA.** *Biochemistry*, **33**, 3906-3912. Abstract
39. Giri, B., Smaldino, P.J., Thys, R.G., Creacy, S.D., Routh, E.D., Hantgan, R.R., Lattmann, S., Nagamine, Y., Akman, S.A. and Vaughn, J.P. (2011) **G4 Resolvase 1 tightly binds and unwinds unimolecular G4-DNA.** *Nucleic Acids Res*, **39**, 7161-7178. Abstract Full Text
40. Arenas, J.E. and Abelson, J.N. (1997) **Prp43: An RNA helicase-like factor involved in spliceosome disassembly.** *Proc Natl Acad Sci U S A*, **94**, 11798-11802. Abstract Full Text
41. Martin, A., Schneider, S. and Schwer, B. (2002) **Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome.** *J Biol Chem*, **277**, 17743-17750. Abstract Full Text
42. Tanaka, N., Aronova, A. and Schwer, B. (2007) **Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome.** *Genes Dev*, **21**, 2312-2325. Abstract Full Text
43. Lebaron, S., Froment, C., Fromont-Racine, M., Rain, J.C., Monsarrat, B., Caizergues-Ferrer, M. and Henry, Y. (2005) **The splicing ATPase prp43p is a component of multiple preribosomal particles.** *Mol Cell Biol*, **25**, 9269-9282. Abstract Full Text
44. Combs, D.J., Nagel, R.J., Ares, M., Jr. and Stevens, S.W. (2006) **Prp43p is a DEAH-box spliceosome disassembly factor essential for ribosome biogenesis.** *Mol Cell Biol*, **26**, 523-534. Abstract Full Text
45. Leeds, N.B., Small, E.C., Hiley, S.L., Hughes, T.R. and Staley, J.P. (2006) **The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis.** *Mol Cell Biol*, **26**, 513-522. Abstract Full Text
46. Koo, J.T., Choe, J. and Moseley, S.L. (2004) **HrpA, a DEAH-box RNA helicase, is involved in mRNA processing of a fimbrial operon in Escherichia coli.** *Mol Microbiol*, **52**, 1813-1826. Abstract Full Text
47. Schwer, B. and Gross, C.H. (1998) **Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing.** *EMBO J*, **17**, 2086-2094. Abstract Full Text

48. Wagner, J.D., Jankowsky, E., Company, M., Pyle, A.M. and Abelson, J.N. (1998) **The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes.** *EMBO J*, **17**, 2926-2937. Abstract Full Text
49. McPheeters, D.S., Schwer, B. and Muhlenkamp, P. (2000) **Interaction of the yeast DExH-box RNA helicase prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing.** *Nucleic Acids Res*, **28**, 1313-1321. Abstract Full Text
50. Schwer, B. (2008) **A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release.** *Mol Cell*, **30**, 743-754. Abstract Full Text
51. Mayas, R.M., Maita, H. and Staley, J.P. (2006) **Exon ligation is proofread by the DExD/H-box ATPase Prp22p.** *Nat Struct Mol Biol*, **13**, 482-490. Abstract Full Text
52. Graham, P.L., Schedl, T. and Kimble, J. (1993) **More mog genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of Caenorhabditis elegans.** *Dev Genet*, **14**, 471-484. Abstract Full Text
53. Puoti, A. and Kimble, J. (2000) **The hermaphrodite sperm/oocyte switch requires the Caenorhabditis elegans homologs of PRP2 and PRP22.** *Proc Natl Acad Sci U S A*, **97**, 3276-3281. Abstract Full Text
54. Colaiacovo, M.P., Stanfield, G.M., Reddy, K.C., Reinke, V., Kim, S.K. and Villeneuve, A.M. (2002) **A targeted RNAi screen for genes involved in chromosome morphogenesis and nuclear organization in the Caenorhabditis elegans germline.** *Genetics*, **162**, 113-128. Abstract
55. Walstrom, K.M., Schmidt, D., Bean, C.J. and Kelly, W.G. (2005) **RNA helicase A is important for germline transcriptional control, proliferation, and meiosis in C. elegans.** *Mech Dev*, **122**, 707-720. Abstract Full Text
56. Kuroda, M.I., Kernan, M.J., Kreber, R., Ganetzky, B. and Baker, B.S. (1991) **The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila.** *Cell*, **66**, 935-947. Abstract Full Text
57. Lee, C.G., Chang, K.A., Kuroda, M.I. and Hurwitz, J. (1997) **The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation.** *EMBO J*, **16**, 2671-2681. Abstract Full Text
58. Lee, C.G., Reichman, T.W., Baik, T. and Mathews, M.B. (2004) **MLE functions as a transcriptional regulator of the roX2 gene.** *J Biol Chem*, **279**, 47740-47745. Abstract Full Text
59. Aratani, S., Kageyama, Y., Nakamura, A., Fujita, H., Fujii, R., Nishioka, K. and Nakajima, T. (2008) **MLE activates transcription via the minimal transactivation domain in Drosophila.** *Int J Mol Med*, **21**, 469-476. Abstract
60. Morra, R., Smith, E.R., Yokoyama, R. and Lucchesi, J.C. (2008) **The MLE subunit of the Drosophila MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting.** *Mol Cell Biol*, **28**, 958-966. Abstract Full Text
61. Reenan, R.A., Hanrahan, C.J. and Ganetzky, B. (2000) **The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing.** *Neuron*, **25**, 139-149. Abstract Full Text
62. Zhang, S. and Grosse, F. (2004) **Multiple functions of nuclear DNA helicase II (RNA helicase A) in nucleic acid metabolism.** *Acta Biochim Biophys Sin (Shanghai)*, **36**, 177-183. Abstract Full Text
63. Nakajima, T., Uchida, C., Anderson, S.F., Lee, C.G., Hurwitz, J., Parvin, J.D. and Montminy, M. (1997) **RNA helicase A mediates association of CBP with RNA polymerase II.** *Cell*, **90**, 1107-1112. Abstract Full Text
64. Anderson, S.F., Schlegel, B.P., Nakajima, T., Wolpin, E.S. and Parvin, J.D. (1998) **BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A.** *Nat Genet*, **19**, 254-256. Abstract Full Text
65. Aratani, S., Fujii, R., Oishi, T., Fujita, H., Amano, T., Ohshima, T., Hagiwara, M., Fukamizu, A. and Nakajima, T. (2001) **Dual roles of RNA helicase A in CREB-dependent transcription.** *Mol Cell Biol*, **21**, 4460-4469. Abstract Full Text
66. Myohanen, S. and Baylin, S.B. (2001) **Sequence-specific DNA binding activity of RNA helicase A to the p16INK4a promoter.** *J Biol Chem*, **276**, 1634-1642. Abstract Full Text
67. Tetsuka, T., Uranishi, H., Sanda, T., Asamitsu, K., Yang, J.P., Wong-Staal, F. and Okamoto, T. (2004) **RNA helicase A interacts with nuclear factor kappaB p65 and functions as a transcriptional coactivator.** *Eur J Biochem*, **271**, 3741-3751. Abstract Full Text
68. Tang, H., Gaietta, G.M., Fischer, W.H., Ellisman, M.H. and Wong-Staal, F. (1997) **A cellular cofactor for the constitutive transport element of type D retrovirus.** *Science*, **276**, 1412-1415. Abstract Full Text
69. Li, J., Tang, H., Mullen, T.M., Westberg, C., Reddy, T.R., Rose, D.W. and Wong-Staal, F. (1999) **A role for RNA helicase A in post-transcriptional regulation of HIV type 1.** *Proc Natl Acad Sci U S A*, **96**, 709-714. Abstract Full Text
70. Roy, B.B., Hu, J., Guo, X., Russell, R.S., Guo, F., Kleiman, L. and Liang, C. (2006) **Association of RNA helicase a with human immunodeficiency virus type 1 particles.** *J Biol Chem*, **281**, 12625-12635. Abstract Full Text
71. He, Q.S., Tang, H., Zhang, J., Truong, K., Wong-Staal, F. and Zhou, D. (2008) **Comparisons of RNAi approaches for validation of human RNA helicase A as an essential factor in hepatitis C virus replication.** *J Virol Methods*, **154**, 216-219. Abstract Full Text
72. Hartman, T.R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D.R. and Boris-Lawrie, K. (2006) **RNA helicase A is necessary for translation of selected messenger RNAs.** *Nat Struct Mol Biol*, **13**, 509-516. Abstract Full Text
73. Robb, G.B. and Rana, T.M. (2007) **RNA helicase A interacts with RISC in human cells and functions in RISC loading.** *Mol Cell*, **26**, 523-537. Abstract Full Text
74. Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M. and Dreyfuss, G. (2001) **A functional interaction between the survival motor neuron complex and RNA polymerase II.** *J Cell Biol*, **152**, 75-85. Abstract Full Text
75. Pertschy, B., Schneider, C., Gnadig, M., Schafer, T., Tollervey, D. and Hurt, E. (2009) **RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1.** *J Biol Chem*, **284**, 35079-35091. Abstract Full Text
76. Imamura, O., Sugawara, M. and Furuichi, Y. (1997) **Cloning and characterization of a putative human RNA helicase gene of the DEAH-box protein family.** *Biochem Biophys Res Commun*, **240**, 335-340. Abstract Full Text
77. Gee, S., Krauss, S.W., Miller, E., Aoyagi, K., Arenas, J. and Conboy, J.G. (1997) **Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of Saccharomyces cerevisiae splicing factor Prp43.** *Proc Natl Acad Sci U S A*, **94**, 11803-11807. Abstract Full Text
78. Fouraux, M.A., Kolkman, M.J., Van der Heijden, A., De Jong, A.S., Van Venrooij, W.J. and Pruijn, G.J. (2002) **The human La (SS-B) autoantigen interacts with DDX15/hPrp43, a putative DEAH-box RNA helicase.** *RNA*, **8**, 1428-1443. Abstract Full Text
79. King, D.S. and Beggs, J.D. (1990) **Interactions of PRP2 protein with pre-mRNA splicing complexes in Saccharomyces cerevisiae.** *Nucleic Acids Res*, **18**, 6559-6564. Abstract Full Text
80. Teigelkamp, S., McGarvey, M., Plumpton, M. and Beggs, J.D. (1994) **The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA.** *EMBO J*, **13**, 888-897. Abstract
81. Kim, S.H. and Lin, R.J. (1996) **Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing.** *Mol Cell Biol*, **16**, 6810-6819. Abstract
82. Imamura, O., Saiki, K., Tani, T., Ohshima, Y., Sugawara, M. and Furuichi, Y. (1998) **Cloning and characterization of a human DEAH-box RNA helicase, a functional homolog of fission yeast Cdc28/Prp8.** *Nucleic Acids Res*, **26**, 2063-2068. Abstract Full Text
83. Pisareva, V.P., Pisarev, A.V., Komar, A.A., Hellen, C.U. and Pestova, T.V. (2008) **Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29.** *Cell*, **135**, 1237-1250. Abstract Full Text
84. Parsyan, A., Shahbazian, D., Martineau, Y., Petroulakis, E., Alain, T., Larsson, O., Mathonnet, G., Tettweiler, G., Hellen, C.U., Pestova, T.V. *et al.* (2009) **The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis.** *Proc Natl Acad Sci U S A*, **106**, 22217-22222. Abstract Full Text
85. Ye, P., Liu, S., Zhu, Y., Chen, G. and Gao, G. (2010) **DEXH-Box protein DHX30 is required for optimal function of the zinc-finger antiviral protein.** *Protein Cell*, **1**, 956-964. Abstract Full Text
86. Alli, Z., Chen, Y., Abdul Wajid, S., Al-Saud, B. and Abdelhaleem, M. (2007) **A role for DHX32 in regulating T-cell apoptosis.** *Anticancer Res*, **27**, 373-377. Abstract
87. Longman, D., Plasterk, R.H., Johnstone, I.L. and Caceres, J.F. (2007) **Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway.** *Genes Dev*, **21**, 1075-1085. Abstract Full Text

88. Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A. and Nagamine, Y. (2010) **Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU.** *Nucleic Acids Res*, **38**, 6219-6233. Abstract Full Text
89. Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S.A. and Vaughn, J.P. (2008) **G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates.** *J Biol Chem*, **283**, 34626-34634. Abstract Full Text
90. Lai, J.C., Ponti, S., Pan, D., Kohler, H., Skoda, R.C., Matthias, P. and Nagamine, Y. (2012) **The DEAH-box helicase RHAU is an essential gene and critical for mouse hematopoiesis.** *Blood*, **119**, 4291. Abstract Full Text
91. Colley, A., Beggs, J.D., Tollervey, D. and Lafontaine, D.L. (2000) **Dhr1p, a putative DEAH-box RNA helicase, is associated with the box C+D snoRNP U3.** *Mol Cell Biol*, **20**, 7238-7246. Abstract Full Text
92. Dragon, F., Gallagher, J.E., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y. *et al.* (2002) **A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis.** *Nature*, **417**, 967-970. Abstract Full Text
93. Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., Marzioch, M., Schafer, T., Kuster, B., Tschochner, H., Tollervey, D. *et al.* (2002) **90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors.** *Mol Cell*, **10**, 105-115. Abstract Full Text
94. Schwer, B. and Guthrie, C. (1991) **PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome.** *Nature*, **349**, 494-499. Abstract Full Text
95. Schwer, B. and Guthrie, C. (1992) **A dominant negative mutation in a spliceosomal ATPase affects ATP hydrolysis but not binding to the spliceosome.** *Mol Cell Biol*, **12**, 3540-3547. Abstract
96. Schwer, B. and Guthrie, C. (1992) **A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis.** *EMBO J*, **11**, 5033-5039. Abstract Full Text
97. Tseng, C.K., Liu, H.L. and Cheng, S.C. (2011) **DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps.** *RNA*, **17**, 145-154. Abstract Full Text
98. Graham, P.L. and Kimble, J. (1993) **The mog-1 gene is required for the switch from spermatogenesis to oogenesis in Caenorhabditis elegans.** *Genetics*, **133**, 919-931. Abstract
99. Zhou, Z. and Reed, R. (1998) **Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing.** *EMBO J*, **17**, 2095-2106. Abstract Full Text
100. Gillespie, D.E. and Berg, C.A. (1995) **Homeless is required for RNA localization in Drosophila oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases.** *Genes Dev*, **9**, 2495-2508. Abstract Full Text
101. Stapleton, W., Das, S. and McKee, B.D. (2001) **A role of the Drosophila homeless gene in repression of Stellate in male meiosis.** *Chromosoma*, **110**, 228-240. Abstract Full Text
102. Kennerdell, J.R., Yamaguchi, S. and Carthew, R.W. (2002) **RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E.** *Genes Dev*, **16**, 1884-1889. Abstract Full Text
103. Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V. and Zamore, P.D. (2006) **A distinct small RNA pathway silences selfish genetic elements in the germline.** *Science*, **313**, 320-324. Abstract Full Text
104. Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. and Gvozdev, V. (2006) **Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline.** *Genes Dev*, **20**, 345-354. Abstract Full Text
105. Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T. *et al.* (2009) **The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline.** *Dev Cell*, **17**, 775-787. Abstract Full Text
106. Iost, I. and Dreyfus, M. (2006) **DEAD-box RNA helicases in Escherichia coli.** *Nucleic Acids Res*, **34**, 4189-4197. Abstract Full Text
107. Kim, S.H., Smith, J., Claude, A. and Lin, R.J. (1992) **The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase.** *EMBO J*, **11**, 2319-2326. Abstract
108. Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H. and Luhrmann, R. (2009) **Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes.** *Mol Cell Biol*, **29**, 281-301. Abstract Full Text
109. Ono, Y., Ohno, M. and Shimura, Y. (1994) **Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22.** *Mol Cell Biol*, **14**, 7611-7620. Abstract
110. Ohno, M. and Shimura, Y. (1996) **A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome.** *Genes Dev*, **10**, 997-1007. Abstract Full Text
111. Gencheva, M., Kato, M., Newo, A.N. and Lin, R.J. (2010) **Contribution of DEAH-box protein DHX16 in human pre-mRNA splicing.** *Biochem J*, **429**, 25-32. Abstract Full Text
112. Puoti, A. and Kimble, J. (1999) **The Caenorhabditis elegans sex determination gene mog-1 encodes a member of the DEAH-Box protein family.** *Mol Cell Biol*, **19**, 2189-2197. Abstract
113. Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y. and Leulliot, N. (2010) **Prp43p contains a processive helicase structural architecture with a specific regulatory domain.** *EMBO J*. Abstract Full Text
114. Edwalds-Gilbert, G., Kim, D.H., Silverman, E. and Lin, R.J. (2004) **Definition of a spliceosome interaction domain in yeast Prp2 ATPase.** *RNA*, **10**, 210-220. Abstract Full Text
115. Hotz, H.R. and Schwer, B. (1998) **Mutational analysis of the yeast DEAH-box splicing factor Prp16.** *Genetics*, **149**, 807-815. Abstract
116. Wang, Y. and Guthrie, C. (1998) **PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain.** *RNA*, **4**, 1216-1229. Abstract Full Text
117. Schneider, S. and Schwer, B. (2001) **Functional domains of the yeast splicing factor Prp22p.** *J Biol Chem*, **276**, 21184-21191. Abstract Full Text
118. Tanaka, N. and Schwer, B. (2006) **Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function.** *Biochemistry*, **45**, 6510-6521. Abstract Full Text
119. Bernstein, D.A. and Keck, J.L. (2003) **Domain mapping of Escherichia coli RecQ defines the roles of conserved N- and C-terminal regions in the RecQ family.** *Nucleic Acids Res*, **31**, 2778-2785. Abstract Full Text
120. Lee, J.W., Kusumoto, R., Doherty, K.M., Lin, G.X., Zeng, W., Cheng, W.H., von Kobbe, C., Brosh, R.M., Jr., Hu, J.S. and Bohr, V.A. (2005) **Modulation of Werner syndrome protein function by a single mutation in the conserved RecQ domain.** *J Biol Chem*, **280**, 39627-39636. Abstract Full Text
121. He, Y., Andersen, G.R. and Nielsen, K.H. (2010) **Structural basis for the function of DEAH helicases.** *EMBO Rep*, **11**, 180-186. Abstract Full Text
122. Gu, M. and Rice, C.M. (2010) **Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism.** *Proc Natl Acad Sci U S A*, **107**, 521-528. Abstract Full Text
123. Assenberg, R., Mastrangelo, E., Walter, T.S., Verma, A., Milani, M., Owens, R.J., Stuart, D.I., Grimes, J.M. and Mancini, E.J. (2009) **Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication.** *J Virol*, **83**, 12895-12906. Abstract Full Text
124. Wu, J., Bera, A.K., Kuhn, R.J. and Smith, J.L. (2005) **Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing.** *J Virol*, **79**, 10268-10277. Abstract Full Text
125. Frick, D.N. (2007) **The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target.** *Curr Issues Mol Biol*, **9**, 1-20. Abstract
126. Oyama, T., Oka, H., Mayanagi, K., Shirai, T., Matoba, K., Fujikane, R., Ishino, Y. and Morikawa, K. (2009) **Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm.** *BMC Struct Biol*, **9**, 2. Abstract Full Text
127. Tanner, N.K., Cordin, O., Banroques, J., Doere, M. and Linder, P. (2003) **The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis.** *Mol Cell*, **11**, 127-138. Abstract Full Text

128. Tanner, N.K. (2003) **The newly identified Q motif of DEAD box helicases is involved in adenine recognition.** *Cell Cycle*, **2**, 18-19. Abstract Full Text
129. Luo, D., Xu, T., Watson, R.P., Scherer-Becker, D., Sampath, A., Jahnke, W., Yeong, S.S., Wang, C.H., Lim, S.P., Strongin, A. *et al.* (2008) **Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein.** *EMBO J*, **27**, 3209-3219. Abstract Full Text
130. Theissen, B., Karow, A.R., Kohler, J., Gubaev, A. and Klostermeier, D. (2008) **Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase.** *Proc Natl Acad Sci U S A*, **105**, 548-553. Abstract Full Text
131. Tanaka, N. and Schwer, B. (2005) **Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-box splicing factor Prp22.** *Biochemistry*, **44**, 9795-9803. Abstract Full Text
132. Lee, C.G. and Hurwitz, J. (1992) **A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3' to 5' direction.** *J Biol Chem*, **267**, 4398-4407. Abstract
133. Bono, F., Ebert, J., Lorentzen, E. and Conti, E. (2006) **The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA.** *Cell*, **126**, 713-725. Abstract Full Text
134. Nielsen, K.H., Chamieh, H., Andersen, C.B., Fredslund, F., Hamborg, K., Le Hir, H. and Andersen, G.R. (2009) **Mechanism of ATP turnover inhibition in the EJC.** *RNA*, **15**, 67-75. Abstract Full Text
135. Ahmadian, M.R., Stege, P., Scheffzek, K. and Wittinghofer, A. (1997) **Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras.** *Nat Struct Biol*, **4**, 686-689. Abstract Full Text
136. Elles, L.M. and Uhlenbeck, O.C. (2008) **Mutation of the arginine finger in the active site of Escherichia coli DbpA abolishes ATPase and helicase activity and confers a dominant slow growth phenotype.** *Nucleic Acids Res*, **36**, 41-50. Abstract Full Text
137. Polach, K.J. and Uhlenbeck, O.C. (2002) **Cooperative binding of ATP and RNA substrates to the DEAD/H protein DbpA.** *Biochemistry*, **41**, 3693-3702. Abstract Full Text
138. Collins, R., Karlberg, T., Lehtio, L., Schutz, P., van den Berg, S., Dahlgren, L.G., Hammarstrom, M., Weigelt, J. and Schuler, H. (2009) **The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch.** *J Biol Chem*, **284**, 10296-10300. Abstract Full Text
139. von Moeller, H., Basquin, C. and Conti, E. (2009) **The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner.** *Nat Struct Mol Biol*, **16**, 247-254. Abstract Full Text
140. Lam, A.M., Keeney, D. and Frick, D.N. (2003) **Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action.** *J Biol Chem*, **278**, 44514-44524. Abstract Full Text
141. Schneider, S., Campodonico, E. and Schwer, B. (2004) **Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8.** *J Biol Chem*, **279**, 8617-8626. Abstract Full Text
142. Lin, C. and Kim, J.L. (1999) **Structure-based mutagenesis study of hepatitis C virus NS3 helicase.** *J Virol*, **73**, 8798-8807. Abstract
143. Arcus, V. (2002) **OB-fold domains: a snapshot of the evolution of sequence, structure and function.** *Curr Opin Struct Biol*, **12**, 794-801. Abstract Full Text
144. Holm, L. and Rosenstrom, P. (2010) **Dali server: conservation mapping in 3D.** *Nucleic Acids Res*, **38 Suppl**, W545-549. Abstract Full Text
145. Ermolenko, D.N. and Makhatadze, G.I. (2002) **Bacterial cold-shock proteins.** *Cell Mol Life Sci*, **59**, 1902-1913. Abstract Full Text
146. Horn, G., Hofweber, R., Kremer, W. and Kalbitzer, H.R. (2007) **Structure and function of bacterial cold shock proteins.** *Cell Mol Life Sci*, **64**, 1457-1470. Abstract Full Text
147. Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S. (2003) **Nucleic acid recognition by OB-fold proteins.** *Annu Rev Biophys Biomol Struct*, **32**, 115-133. Abstract Full Text
148. Izzo, A., Regnard, C., Morales, V., Kremmer, E. and Becker, P.B. (2008) **Structure-function analysis of the RNA helicase maleless.** *Nucleic Acids Res*, **36**, 950-962. Abstract Full Text
149. Schneider, S., Hotz, H.R. and Schwer, B. (2002) **Characterization of dominant-negative mutants of the DEAH-box splicing factors Prp22 and Prp16.** *J Biol Chem*, **277**, 15452-15458. Abstract Full Text
150. Schwer, B. and Meszaros, T. (2000) **RNA helicase dynamics in pre-mRNA splicing.** *EMBO J*, **19**, 6582-6591. Abstract Full Text
151. Campodonico, E. and Schwer, B. (2002) **ATP-dependent remodeling of the spliceosome: intragenic suppressors of release-defective mutants of Saccharomyces cerevisiae Prp22.** *Genetics*, **160**, 407-415. Abstract
152. Plumpton, M., McGarvey, M. and Beggs, J.D. (1994) **A dominant negative mutation in the conserved RNA helicase motif 'SAT' causes splicing factor PRP2 to stall in spliceosomes.** *EMBO J*, **13**, 879-887. Abstract
153. Edwalds-Gilbert, G., Kim, D.H., Kim, S.H., Tseng, Y.H., Yu, Y. and Lin, R.J. (2000) **Dominant negative mutants of the yeast splicing factor Prp2 map to a putative cleft region in the helicase domain of DExD/H-box proteins.** *RNA*, **6**, 1106-1119. Abstract Full Text
154. Pause, A. and Sonenberg, N. (1992) **Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A.** *EMBO J*, **11**, 2643-2654. Abstract
155. Gross, C.H. and Shuman, S. (1998) **The nucleoside triphosphatase and helicase activities of vaccinia virus NPH-II are essential for virus replication.** *J Virol*, **72**, 4729-4736. Abstract
156. Karow, A.R. and Klostermeier, D. (2009) **A conformational change in the helicase core is necessary but not sufficient for RNA unwinding by the DEAD box helicase YxiN.** *Nucleic Acids Res*, **37**, 4464-4471. Abstract Full Text
157. Kos, M. and Tollervey, D. (2005) **The Putative RNA Helicase Dbp4p Is Required for Release of the U14 snoRNA from Preribosomes in Saccharomyces cerevisiae.** *Mol Cell*, **20**, 53-64. Abstract Full Text
158. Bernstein, K.A., Granneman, S., Lee, A.V., Manickam, S. and Baserga, S.J. (2006) **Comprehensive mutational analysis of yeast DEXD/H box RNA helicases involved in large ribosomal subunit biogenesis.** *Mol Cell Biol*, **26**, 1195-1208. Abstract Full Text
159. Granneman, S., Bernstein, K.A., Bleichert, F. and Baserga, S.J. (2006) **Comprehensive mutational analysis of yeast DEXD/H box RNA helicases required for small ribosomal subunit synthesis.** *Mol Cell Biol*, **26**, 1183-1194. Abstract Full Text
160. Banroques, J., Doere, M., Dreyfus, M., Linder, P. and Tanner, N.K. (2010) **Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1.** *J Mol Biol*, **396**, 949-966. Abstract Full Text
161. Banroques, J., Cordin, O., Doere, M., Linder, P. and Tanner, N.K. (2008) **A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins.** *Mol Cell Biol*, **28**, 3359-3371. Abstract Full Text
162. Gallivan, J.P. and Dougherty, D.A. (1999) **Cation-pi interactions in structural biology.** *Proc Natl Acad Sci U S A*, **96**, 9459-9464. Abstract Full Text
163. Levin, M.K., Gurjar, M. and Patel, S.S. (2005) **A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase.** *Nat Struct Mol Biol*, **12**, 429-435. Abstract Full Text
164. Appleby, T.C., Anderson, R., Fedorova, O., Pyle, A.M., Wang, R., Liu, X., Brendza, K.M. and Somoza, J.R. (2011) **Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV.** *J Mol Biol*, **405**, 1139-1153. Abstract Full Text
165. Raney, K.D., Sharma, S.D., Moustafa, I.M. and Cameron, C.E. (2010) **Hepatitis C virus non-structural protein 3 (HCV NS3): a multifunctional antiviral target.** *J Biol Chem*, **285**, 22725-22731. Abstract Full Text
166. Wang, Y., Wagner, J.D. and Guthrie, C. (1998) **The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro.** *Curr Biol*, **8**, 441-451. Abstract Full Text
167. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) **Helix formation by guanylic acid.** *Proc Natl Acad Sci U S A*, **48**, 2013-2018. Abstract Full Text
168. Ralph, R.K., Connors, W.J. and Khorana, H.G. (1962) **Secondary Structure and Aggregation in Deoxyguanosine Oligonucleotides.** *J Am Chem Soc*, **84**, 2265-2266. Full Text
169. Rich, A. (1993) **DNA comes in many forms.** *Gene*, **135**, 99-109. Abstract
170. Doktycz, M.J. (2002) **Nucleic Acids: Thermal Stability and Denaturation.** *Encyclopedia of Life Sciences*, 1-7. Full Text
171. Wells, R.D. (2007) **Non-B DNA conformations, mutagenesis and disease.** *Trends Biochem Sci*, **32**, 271-278. Abstract Full Text
172. König, S., Evans, A. and Huppert, J. (2010) **Seven essential questions on G-quadruplexes.** *Biomol Concepts*, **1**, 197-213. Full Text
173. Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) **Stability and kinetics of G-quadruplex structures.** *Nucleic Acids Res*, **36**, 5482-5515. Abstract Full Text

174. Hud, N.V. and Plavec, J. (2006) In Neidle, S. and Balasubramanian, S. (eds.), *Quadruplex Nucleic Acids*. RSC Publishing, Cambridge, UK, pp. 100-130.
175. Engelhart, A.E., Plavec, J., Persil, O. and Hud, N.V. (2008) In Glover, D. M. (ed.), *Nucleic Acid-Metal Ion Interactions*. RSC Publishing, Cambridge, UK, pp. 118-153.
176. Davis, J.T. (2004) **G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry.** *Angew Chem Int Ed Engl*, **43**, 668-698. Abstract Full Text
177. Bang, I. (1910) **Untersuchung über die Guanylsäure.** *Biochem Ztschr*, **26**, 293-311.
178. Guschlbauer, W., Chantot, J.F. and Thiele, D. (1990) **Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA.** *J Biomol Struct Dyn*, **8**, 491-511. Abstract
179. Sen, D. and Gilbert, W. (1990) **A sodium-potassium switch in the formation of four-stranded G4-DNA.** *Nature*, **344**, 410-414. Abstract Full Text
180. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., Jr. and Blackburn, E.H. (1987) **Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs.** *Cell*, **51**, 899-908. Abstract Full Text
181. Oka, Y. and Thomas, C.A., Jr. (1987) **The cohering telomeres of Oxytricha.** *Nucleic Acids Res*, **15**, 8877-8898. Abstract Full Text
182. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) **Monovalent cation-induced structure of telomeric DNA: the G-quartet model.** *Cell*, **59**, 871-880. Abstract Full Text
183. Sundquist, W.I. and Klug, A. (1989) **Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops.** *Nature*, **342**, 825-829. Abstract Full Text
184. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) **Inhibition of telomerase by G-quartet DNA structures.** *Nature*, **350**, 718-720. Abstract Full Text
185. Mergny, J.L., Gros, J., De Cian, A., Bourdoncle, A., Rosu, F., Sacca, B., Guittat, L., Amrane, S., Mills, M., Alberti, P. *et al.* (2006) In Neidle, S. and Balasubramanian, S. (eds.), *Quadruplex Nucleic Acids*. RSC Publishing, Cambridge, UK, pp. 31-80.
186. Huppert, J.L. and Balasubramanian, S. (2005) **Prevalence of quadruplexes in the human genome.** *Nucleic Acids Res*, **33**, 2908-2916. Abstract Full Text
187. Huppert, J.L. and Balasubramanian, S. (2007) **G-quadruplexes in promoters throughout the human genome.** *Nucleic Acids Res*, **35**, 406-413. Abstract Full Text
188. Todd, A.K., Johnston, M. and Neidle, S. (2005) **Highly prevalent putative quadruplex sequence motifs in human DNA.** *Nucleic Acids Res*, **33**, 2901-2907. Abstract Full Text
189. Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S. and Johnson, F.B. (2008) **Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae.** *Nucleic Acids Res*, **36**, 144-156. Abstract Full Text
190. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) **An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation.** *Nat Chem Biol*, **3**, 218-221. Abstract Full Text
191. Huppert, J.L., Bugaut, A., Kumari, S. and Balasubramanian, S. (2008) **G-quadruplexes: the beginning and end of UTRs.** *Nucleic Acids Res*, **36**, 6260-6268. Abstract Full Text
192. Eddy, J. and Maizels, N. (2008) **Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes.** *Nucleic Acids Res*, **36**, 1321-1333. Abstract Full Text
193. Neidle, S. (2010) **Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer.** *FEBS J*, **277**, 1118-1125. Abstract Full Text
194. Zhou, W., Brand, N.J. and Ying, L. (2011) **G-quadruplexes-novel mediators of gene function.** *J Cardiovasc Transl Res*, **4**, 256-270. Abstract Full Text
195. Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.M. and Lemaitre, J.M. (2012) **Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs.** *Nat Struct Mol Biol*, **19**, 837-844. Abstract Full Text
196. Chang, C.C., Chu, J.F., Kao, F.J., Chiu, Y.C., Lou, P.J., Chen, H.C. and Chang, T.C. (2006) **Verification of antiparallel G-quadruplex structure in human telomeres by using two-photon excitation fluorescence lifetime imaging microscopy of the 3,6-Bis(1-methyl-4-vinylpyridinium)carbazole diiodide molecule.** *Anal Chem*, **78**, 2810-2815. Abstract Full Text
197. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) **In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei.** *Proc Natl Acad Sci U S A*, **98**, 8572-8577. Abstract Full Text
198. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) **Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo.** *Nat Struct Mol Biol*, **12**, 847-854. Abstract Full Text
199. Grand, C.L., Han, H., Munoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) **The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo.** *Mol Cancer Ther*, **1**, 565-573. Abstract
200. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) **Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription.** *Proc Natl Acad Sci U S A*, **99**, 11593-11598. Abstract Full Text
201. Fernando, H., Rodriguez, R. and Balasubramanian, S. (2008) **Selective recognition of a DNA G-quadruplex by an engineered antibody.** *Biochemistry*, **47**, 9365-9371. Abstract Full Text
202. Fernando, H., Sewitz, S., Darot, J., Tavare, S., Huppert, J.L. and Balasubramanian, S. (2009) **Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression.** *Nucleic Acids Res*, **37**, 6716-6722. Abstract Full Text
203. Sacca, B., Lacroix, L. and Mergny, J.L. (2005) **The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides.** *Nucleic Acids Res*, **33**, 1182-1192. Abstract Full Text
204. Mergny, J.L., De Cian, A., Ghelab, A., Sacca, B. and Lacroix, L. (2005) **Kinetics of tetramolecular quadruplexes.** *Nucleic Acids Res*, **33**, 81-94. Abstract Full Text
205. Joachimi, A., Benz, A. and Hartig, J.S. (2009) **A comparison of DNA and RNA quadruplex structures and stabilities.** *Bioorg Med Chem*, **17**, 6811-6815. Abstract Full Text
206. Arora, A. and Maiti, S. (2009) **Differential biophysical behavior of human telomeric RNA and DNA quadruplex.** *J Phys Chem B*, **113**, 10515-10520. Abstract Full Text
207. Beaudoin, J.D. and Perreault, J.P. (2010) **5'-UTR G-quadruplex structures acting as translational repressors.** *Nucleic Acids Res*, **38**, 7022-7036. Abstract Full Text
208. Arora, A., Dutkiewicz, M., Scaria, V., Hariharan, M., Maiti, S. and Kurreck, J. (2008) **Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif.** *RNA*, **14**, 1290-1296. Abstract Full Text
209. Morris, M.J. and Basu, S. (2009) **An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells.** *Biochemistry*, **48**, 5313-5319. Abstract Full Text
210. Gomez, D., Guedin, A., Mergny, J.L., Salles, B., Riou, J.F., Teulade-Fichou, M.P. and Calsou, P. (2010) **A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells.** *Nucleic Acids Res*, **38**, 7187-7198. Abstract Full Text
211. Bonnal, S., Schaeffer, C., Creancier, L., Clamens, S., Moine, H., Prats, A.C. and Vagner, S. (2003) **A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons.** *J Biol Chem*, **278**, 39330-39336. Abstract Full Text
212. Morris, M.J., Negishi, Y., Pazsint, C., Schonhoft, J.D. and Basu, S. (2010) **An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES.** *J Am Chem Soc*, **132**, 17831-17839. Abstract Full Text
213. Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J.L. and Riou, J.F. (2004) **Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing.** *Nucleic Acids Res*, **32**, 371-379. Abstract Full Text

214. Kostadinov, R., Malhotra, N., Viotti, M., Shine, R., D'Antonio, L. and Bagga, P. (2006) **GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences.** *Nucleic Acids Res*, **34**, D119-124. Abstract Full Text

215. Didiot, M.C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J.L. and Moine, H. (2008) **The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer.** *Nucleic Acids Res*, **36**, 4902-4912. Abstract Full Text

216. Marcel, V., Tran, P.L., Sagne, C., Martel-Planche, G., Vaslin, L., Teulade-Fichou, M.P., Hall, J., Mergny, J.L., Hainaut, P. and Van Dyck, E. (2011) **G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms.** *Carcinogenesis*, **32**, 271-278. Abstract Full Text

217. Kikin, O., Zappala, Z., D'Antonio, L. and Bagga, P.S. (2008) **GRSDB2 and GRS_UTRdb: databases of quadruplex forming G-rich sequences in pre-mRNAs and mRNAs.** *Nucleic Acids Res*, **36**, D141-148. Abstract Full Text

218. Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M. and Gustafsson, C.M. (2010) **G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation.** *Proc Natl Acad Sci U S A*, **107**, 16072-16077. Abstract Full Text

219. Decorsiere, A., Cayrel, A., Vagner, S. and Millevoi, S. (2011) **Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage.** *Genes Dev*, **25**, 220-225. Abstract Full Text

220. Randall, A. and Griffith, J.D. (2009) **Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets.** *J Biol Chem*, **284**, 13980-13986. Abstract Full Text

221. Martadinata, H. and Phan, A.T. (2009) **Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution.** *J Am Chem Soc*, **131**, 2570-2578. Abstract Full Text

222. Xu, Y., Kaminaga, K. and Komiyama, M. (2008) **G-quadruplex formation by human telomeric repeats-containing RNA in Na+ solution.** *J Am Chem Soc*, **130**, 11179-11184. Abstract Full Text

223. Xu, Y., Suzuki, Y., Ito, K. and Komiyama, M. (2010) **Telomeric repeat-containing RNA structure in living cells.** *Proc Natl Acad Sci U S A*, **107**, 14579-14584. Abstract Full Text

224. Gros, J., Guedin, A., Mergny, J.L. and Lacroix, L. (2008) **G-Quadruplex formation interferes with P1 helix formation in the RNA component of telomerase hTERC.** *Chembiochem*, **9**, 2075-2079. Abstract Full Text

225. Wu, Y. and Brosh, R.M., Jr. (2010) **G-quadruplex nucleic acids and human disease.** *FEBS J*, **277**, 3470-3488. Abstract Full Text

226. Fry, M. and Loeb, L.A. (1999) **Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n.** *J Biol Chem*, **274**, 12797-12802. Abstract Full Text

227. Sun, H., Karow, J.K., Hickson, I.D. and Maizels, N. (1998) **The Bloom's syndrome helicase unwinds G4 DNA.** *J Biol Chem*, **273**, 27587-27592. Abstract Full Text

228. Wu, Y., Shin-ya, K. and Brosh, R.M., Jr. (2008) **FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability.** *Mol Cell Biol*, **28**, 4116-4128. Abstract Full Text

229. London, T.B., Barber, L.J., Mosedale, G., Kelly, G.P., Balasubramanian, S., Hickson, I.D., Boulton, S.J. and Hiom, K. (2008) **FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts.** *J Biol Chem*, **283**, 36132-36139. Abstract Full Text

230. Wu, Y., Sommers, J.A., Khan, I., de Winter, J.P. and Brosh, R.M., Jr. (2012) **Biochemical characterization of Warsaw breakage syndrome helicase.** *J Biol Chem*, **287**, 1007-1021. Abstract Full Text

231. van der Lelij, P., Chrzanowska, K.H., Godthelp, B.C., Rooimans, M.A., Oostra, A.B., Stumm, M., Zdzienicka, M.Z., Joenje, H. and de Winter, J.P. (2010) **Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1.** *Am J Hum Genet*, **86**, 262-266. Abstract Full Text

232. Parish, J.L., Rosa, J., Wang, X., Lahti, J.M., Doxsey, S.J. and Androphy, E.J. (2006) **The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells.** *J Cell Sci*, **119**, 4857-4865. Abstract Full Text

233. Sanders, C.M. (2010) **Human Pif1 helicase is a G-quadruplex DNA binding protein with G-quadruplex DNA unwinding activity.** *Biochem J*, **430**, 119-128. Abstract Full Text

234. Zhang, D.H., Zhou, B., Huang, Y., Xu, L.X. and Zhou, J.Q. (2006) **The human Pif1 helicase, a potential Escherichia coli RecD homologue, inhibits telomerase activity.** *Nucleic Acids Res*, **34**, 1393-1404. Abstract Full Text

235. Duxin, J.P., Dao, B., Martinsson, P., Rajala, N., Guittat, L., Campbell, J.L., Spelbrink, J.N. and Stewart, S.A. (2009) **Human Dna2 is a nuclear and mitochondrial DNA maintenance protein.** *Mol Cell Biol*, **29**, 4274-4282. Abstract Full Text

236. Zheng, L., Zhou, M., Guo, Z., Lu, H., Qian, L., Dai, H., Qiu, J., Yakubovskaya, E., Bogenhagen, D.F., Demple, B. *et al.* (2008) **Human DNA2 is a mitochondrial nuclease/helicase for efficient processing of DNA replication and repair intermediates.** *Mol Cell*, **32**, 325-336. Abstract Full Text

237. Lee, C.G., da Costa Soares, V., Newberger, C., Manova, K., Lacy, E. and Hurwitz, J. (1998) **RNA helicase A is essential for normal gastrulation.** *Proc Natl Acad Sci U S A*, **95**, 13709-13713. Abstract Full Text

238. Chakraborty, P. and Grosse, F. (2011) **Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes.** *DNA Repair (Amst)*, **10**, 654-665. Abstract Full Text

239. Cheung, I., Schertzer, M., Rose, A. and Lansdorp, P.M. (2002) **Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA.** *Nat Genet*, **31**, 405-409. Abstract Full Text

240. Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E. and Tijsterman, M. (2008) **Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective C. elegans.** *Curr Biol*, **18**, 900-905. Abstract Full Text

241. Piazza, A., Boule, J.B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M.P. and Nicolas, A. (2010) **Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae.** *Nucleic Acids Res*, **38**, 4337-4348. Abstract Full Text

242. Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M.P., Foiani, M. and Nicolas, A. (2011) **G-quadruplex-induced instability during leading-strand replication.** *EMBO J*, **30**, 4033-4046. Abstract Full Text

243. Huber, M.D., Lee, D.C. and Maizels, N. (2002) **G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition.** *Nucleic Acids Res*, **30**, 3954-3961. Abstract Full Text

244. Sun, H., Bennett, R.J. and Maizels, N. (1999) **The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs.** *Nucleic Acids Res*, **27**, 1978-1984. Abstract

245. Ribeyre, C., Lopes, J., Boule, J.B., Piazza, A., Guedin, A., Zakian, V.A., Mergny, J.L. and Nicolas, A. (2009) **The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo.** *PLoS Genet*, **5**, e1000475. Abstract Full Text

246. Masuda-Sasa, T., Polaczek, P., Peng, X.P., Chen, L. and Campbell, J.L. (2008) **Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition.** *J Biol Chem*, **283**, 24359-24373. Abstract Full Text

247. Sarkies, P., Murat, P., Phillips, L.G., Patel, K.J., Balasubramanian, S. and Sale, J.E. (2012) **FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA.** *Nucleic Acids Res*, **40**, 1485-1498. Abstract Full Text

248. Wang, W. (2007) **Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins.** *Nat Rev Genet*, **8**, 735-748. Abstract Full Text

249. Bochman, M.L., Sabouri, N. and Zakian, V.A. (2010) **Unwinding the functions of the Pif1 family helicases.** *DNA Repair (Amst)*, **9**, 237-249. Abstract Full Text

250. Ellis, N.A., Groden, J., Ye, T.Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M. and German, J. (1995) **The Bloom's syndrome gene product is homologous to RecQ helicases.** *Cell*, **83**, 655-666. Abstract

251. Yu, C.E., Oshima, J., Fu, Y.H., Wijsman, E.M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S. *et al.* (1996) **Positional cloning of the Werner's syndrome gene.** *Science*, **272**, 258-262. Abstract

252. Levran, O., Attwooll, C., Henry, R.T., Milton, K.L., Neveling, K., Rio, P., Batish, S.D., Kalb, R., Velleuer, E., Barral, S. *et al.* (2005) **The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia.** *Nat Genet*, **37**, 931-933. Abstract Full Text

253. Levitus, M., Waisfisz, Q., Godthelp, B.C., de Vries, Y., Hussain, S., Wiegant, W.W., Elghalbzouri-Maghrani, E., Steltenpool, J., Rooimans, M.A., Pals, G. *et al.* (2005) **The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J.** *Nat Genet*, **37**, 934-935. Abstract Full Text

254. Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T. and Darnell, R.B. (2001) **Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function.** *Cell*, **107**, 489-499. Abstract Full Text

255. Lu, R., Wang, H., Liang, Z., Ku, L., O'Donnell W, T., Li, W., Warren, S.T. and Feng, Y. (2004) **The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development.** *Proc Natl Acad Sci U S A*, **101**, 15201-15206. Abstract Full Text

256. Castets, M., Schaeffer, C., Bechara, E., Schenck, A., Khandjian, E.W., Luche, S., Moine, H., Rabilloud, T., Mandel, J.L. and Bardoni, B. (2005) **FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts.** *Hum Mol Genet*, **14**, 835-844. Abstract Full Text

257. Westmark, C.J. and Malter, J.S. (2007) **FMRP mediates mGluR5-dependent translation of amyloid precursor protein.** *PLoS Biol*, **5**, e52. Abstract Full Text

258. Liao, L., Park, S.K., Xu, T., Vanderklish, P. and Yates, J.R., 3rd. (2008) **Quantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in fmr1 knockout mice.** *Proc Natl Acad Sci U S A*, **105**, 15281-15286. Abstract Full Text

259. Bensaid, M., Melko, M., Bechara, E.G., Davidovic, L., Berretta, A., Catania, M.V., Gecz, J., Lalli, E. and Bardoni, B. (2009) **FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure.** *Nucleic Acids Res*, **37**, 1269-1279. Abstract Full Text

260. Millevoi, S., Moine, H. and Vagner, S. (2012) **G-quadruplexes in RNA biology.** *Wiley Interdiscip Rev RNA*, **3**, 495-507. Abstract Full Text

261. Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J. and Bardoni, B. (2011) **Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins: insights into the molecular pathology of FRAXE intellectual disability.** *Hum Mol Genet*, **20**, 1873-1885. Abstract Full Text

262. Khateb, S., Weisman-Shomer, P., Hershco, I., Loeb, L.A. and Fry, M. (2004) **Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family.** *Nucleic Acids Res*, **32**, 4145-4154. Abstract Full Text

263. Takahama, K., Kino, K., Arai, S., Kurokawa, R. and Oyoshi, T. (2011) **Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein.** *FEBS J*, **278**, 988-998. Abstract Full Text

264. Nagata, T., Kurihara, Y., Matsuda, G., Saeki, J., Kohno, T., Yanagida, Y., Ishikawa, F., Uesugi, S. and Katahira, M. (1999) **Structure and interactions with RNA of the N-terminal UUAG-specific RNA-binding domain of hnRNP D0.** *J Mol Biol*, **287**, 221-237. Abstract

265. Davidovic, L., Bechara, E., Gravel, M., Jaglin, X.H., Tremblay, S., Sik, A., Bardoni, B. and Khandjian, E.W. (2006) **The nuclear microspherule protein 58 is a novel RNA-binding protein that interacts with fragile X mental retardation protein in polyribosomal mRNPs from neurons.** *Hum Mol Genet*, **15**, 1525-1538. Abstract Full Text

266. Biffi, G., Tannahill, D. and Balasubramanian, S. (2012) **An intramolecular G-quadruplex structure is required for TERRA RNA binding to the telomeric protein TRF2.** *J Am Chem Soc*, **134**, 11974-11976. Abstract Full Text

267. Bashkirov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M. and Heyer, W.D. (1997) **A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates.** *J Cell Biol*, **136**, 761-773. Abstract Full Text

268. Chalupnikova, K., Lattmann, S., Selak, N., Iwamoto, F., Fujiki, Y. and Nagamine, Y. (2008) **Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain.** *J Biol Chem*, **283**, 35186-35198. Abstract Full Text

269. Iwamoto, F., Stadler, M., Chalupnikova, K., Oakeley, E. and Nagamine, Y. (2008) **Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU.** *Exp Cell Res*, **314**, 1378-1391. Abstract Full Text

270. Askjaer, P., Rosendahl, R. and Kjems, J. (2000) **Nuclear export of the DEAD box An3 protein by CRM1 is coupled to An3 helicase activity.** *J Biol Chem*, **275**, 11561-11568. Abstract Full Text

271. Anderson, P. and Kedersha, N. (2006) **RNA granules.** *J Cell Biol*, **172**, 803-808. Abstract Full Text

272. Anderson, P. and Kedersha, N. (2008) **Stress granules: the Tao of RNA triage.** *Trends Biochem Sci*, **33**, 141-150. Abstract Full Text

273. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd *et al.* (2009) **BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources.** *Genome Biol*, **10**, R130. Abstract Full Text

274. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. *et al.* (2004) **A gene atlas of the mouse and human protein-encoding transcriptomes.** *Proc Natl Acad Sci U S A*, **101**, 6062-6067. Abstract Full Text

275. Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A. *et al.* (2002) **Large-scale analysis of the human and mouse transcriptomes.** *Proc Natl Acad Sci U S A*, **99**, 4465-4470. Abstract Full Text

276. Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T. *et al.* (2007) **DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs.** *Nat Cell Biol*, **9**, 604-611. Abstract Full Text

277. Fukunaga, A., Tanaka, A. and Oishi, K. (1975) **Maleless, a recessive autosomal mutant of Drosophila melanogaster that specifically kills male zygotes.** *Genetics*, **81**, 135-141. Abstract

278. Gonzalez-Reyes, A., Elliott, H. and St Johnston, D. (1997) **Oocyte determination and the origin of polarity in Drosophila: the role of the spindle genes.** *Development*, **124**, 4927-4937. Abstract

279. English, M.A., Lei, L., Blake, T., Wincovitch, S.M., Sr., Sood, R., Azuma, M., Hickstein, D. and Paul Liu, P. (2012) **Incomplete splicing, cell division defects and hematopoietic blockage in dhx8 mutant zebrafish.** *Dev Dyn*, **241**, 879-889. Abstract Full Text

280. Hoffbrand, A.V., Moss, P.A.H. and Pettit, J.E. (2006), *Essential Haematology*. 5th ed. Blackwell Publishing Ltd, Oxford, UK, pp. 123-128.

281. Kim, T., Pazhoor, S., Bao, M., Zhang, Z., Hanabuchi, S., Facchinetti, V., Bover, L., Plumas, J., Chaperot, L., Qin, J. *et al.* (2010) **Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells.** *Proc Natl Acad Sci U S A*. Abstract Full Text

282. Hokeness-Antonelli, K.L., Crane, M.J., Dragoi, A.M., Chu, W.M. and Salazar-Mather, T.P. (2007) **IFN-alphabeta-mediated inflammatory responses and antiviral defense in liver is TLR9-independent but MyD88-dependent during murine cytomegalovirus infection.** *J Immunol*, **179**, 6176-6183. Abstract

283. Jung, A., Kato, H., Kumagai, Y., Kumar, H., Kawai, T., Takeuchi, O. and Akira, S. (2008) **Lymphocytoid choriomeningitis virus activates plasmacytoid dendritic cells and induces a cytotoxic T-cell response via MyD88.** *J Virol*, **82**, 196-206. Abstract Full Text

284. Hochrein, H., Schlatter, B., O'Keeffe, M., Wagner, C., Schmitz, F., Schiemann, M., Bauer, S., Suter, M. and Wagner, H. (2004) **Herpes simplex virus type-1 induces IFN-alpha production via Toll-like receptor 9-dependent and -independent pathways.** *Proc Natl Acad Sci U S A*, **101**, 11416-11421. Abstract Full Text

285. Hopfner, K.P., Cui, S., Kirchhofer, A. and Pippig, D. (2010) In Jankowsky, E. (ed.), *RNA Helicases*. The Royal Society of Chemistry, Cambridge, UK, pp. 121-148.

286. Loo, Y.M. and Gale, M., Jr. (2011) **Immune signaling by RIG-I-like receptors.** *Immunity*, **34**, 680-692. Abstract Full Text

287. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) **Site-directed mutagenesis by overlap extension using the polymerase chain reaction.** *Gene*, **77**, 51-59. Abstract Full Text

288. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. (1989) **Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension.** *Gene*, **77**, 61-68. Abstract Full Text

289. Zheng, L., Baumann, U. and Reymond, J.L. (2004) **An efficient one-step site-directed and site-saturation mutagenesis protocol.** *Nucleic Acids Res*, **32**, e115. Abstract Full Text
290. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol*, **5**, R80. Abstract Full Text
291. Carvalho, B.S. and Irizarry, R.A. (2010) **A framework for oligonucleotide microarray preprocessing.** *Bioinformatics*, **26**, 2363-2367. Abstract Full Text
292. Smyth, G.K. (2004) **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol*, **3**, Article3. Abstract Full Text
293. Kikin, O., D'Antonio, L. and Bagga, P.S. (2006) **QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.** *Nucleic Acids Res*, **34**, W676-682. Abstract Full Text
294. Altschul, S.F. and Erickson, B.W. (1985) **Significance of nucleotide sequence alignments: a method for random sequence permutation that preserves dinucleotide and codon usage.** *Mol Biol Evol*, **2**, 526-538. Abstract
295. Livak, K.J. and Schmittgen, T.D. (2001) **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods*, **25**, 402-408. Abstract Full Text
296. Herbert, B.S., Hochreiter, A.E., Wright, W.E. and Shay, J.W. (2006) **Nonradioactive detection of telomerase activity using the telomeric repeat amplification protocol.** *Nat Protoc*, **1**, 1583-1590. Abstract Full Text
297. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H. and Karplus, M. (1995) **Evaluation of comparative protein modeling by MODELLER.** *Proteins*, **23**, 318-326. Abstract Full Text
298. Eswar, N., Eramian, D., Webb, B., Shen, M.Y. and Sali, A. (2008) **Protein structure modeling with MODELLER.** *Methods Mol Biol*, **426**, 145-159. Abstract Full Text
299. Benkert, P., Biasini, M. and Schwede, T. (2011) **Toward the estimation of the absolute quality of individual protein structure models.** *Bioinformatics*, **27**, 343-350. Abstract Full Text
300. McGuffin, L.J. and Roche, D.B. (2010) **Rapid model quality assessment for protein structure predictions using the comparison of multiple models without structural alignments.** *Bioinformatics*, **26**, 182-188. Abstract Full Text
301. Schrodinger, LLC. (2010) **The PyMOL Molecular Graphics System, Version 1.5.**
302. Huber, M.D., Duquette, M.L., Shiels, J.C. and Maizels, N. (2006) **A conserved G4 DNA binding domain in RecQ family helicases.** *J Mol Biol*, **358**, 1071-1080. Abstract Full Text
303. Baran, N., Pucshansky, L., Marco, Y., Benjamin, S. and Manor, H. (1997) **The SV40 large T-antigen helicase can unwind four stranded DNA structures linked by G-quartets.** *Nucleic Acids Res*, **25**, 297-303. Abstract Full Text
304. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) **Crystal structure of parallel quadruplexes from human telomeric DNA.** *Nature*, **417**, 876-880. Abstract Full Text
305. Phan, A.T. and Patel, D.J. (2003) **Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics.** *J Am Chem Soc*, **125**, 15021-15027. Abstract Full Text
306. Gray, D.M., Wen, J.D., Gray, C.W., Repges, R., Repges, C., Raabe, G. and Fleischhauer, J. (2008) **Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities.** *Chirality*, **20**, 431-440. Abstract Full Text
307. Kypr, J., Kejnovska, I., Renciuk, D. and Vorlickova, M. (2009) **Circular dichroism and conformational polymorphism of DNA.** *Nucleic Acids Res*, **37**, 1713-1725. Abstract Full Text
308. Paramasivan, S., Rujan, I. and Bolton, P.H. (2007) **Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding.** *Methods*, **43**, 324-331. Abstract Full Text
309. Tang, C.F. and Shafer, R.H. (2006) **Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes.** *J Am Chem Soc*, **128**, 5966-5973. Abstract Full Text
310. Pilch, D.S., Plum, G.E. and Breslauer, K.J. (1995) **The thermodynamics of DNA structures that contain lesions or guanine tetrads.** *Curr Opin Struct Biol*, **5**, 334-342. Abstract Full Text
311. Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C. and Caron, P.R. (1998) **Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.** *Structure*, **6**, 89-100. Abstract Full Text
312. Del Campo, M. and Lambowitz, A.M. (2009) **Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA.** *Mol Cell*, **35**, 598-609. Abstract Full Text
313. Katoh, K. and Toh, H. (2008) **Recent developments in the MAFFT multiple sequence alignment program.** *Brief Bioinform*, **9**, 286-298. Abstract Full Text
314. Cole, C., Barber, J.D. and Barton, G.J. (2008) **The Jpred 3 secondary structure prediction server.** *Nucleic Acids Res*, **36**, W197-201. Abstract Full Text
315. Aasland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford, M.T., Cesareni, G., Gimona, M., Hurley, J.H., Jarchau, T., Lehto, V.P. *et al.* (2002) **Normalization of nomenclature for peptide motifs as ligands of modular protein domains.** *FEBS Lett*, **513**, 141-144. Abstract Full Text
316. Qin, Y. and Hurley, L.H. (2008) **Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions.** *Biochimie*, **90**, 1149-1171. Abstract Full Text
317. Wieland, M. and Hartig, J.S. (2007) **RNA quadruplex-based modulation of gene expression.** *Chem Biol*, **14**, 757-763. Abstract Full Text
318. Kang, Y.H., Lee, C.H. and Seo, Y.S. (2010) **Dna2 on the road to Okazaki fragment processing and genome stability in eukaryotes.** *Crit Rev Biochem Mol Biol*, **45**, 71-96. Abstract Full Text
319. Singleton, M.R. and Wigley, D.B. (2002) **Modularity and specialization in superfamily 1 and 2 helicases.** *J Bacteriol*, **184**, 1819-1826. Abstract Full Text
320. Karginov, F.V., Caruthers, J.M., Hu, Y., McKay, D.B. and Uhlenbeck, O.C. (2005) **YxiN is a modular protein combining a DEx(D/H) core and a specific RNA-binding domain.** *J Biol Chem*, **280**, 35499-35505. Abstract Full Text
321. Grohman, J.K., Del Campo, M., Bhaskaran, H., Tijerina, P., Lambowitz, A.M. and Russell, R. (2007) **Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA.** *Biochemistry*, **46**, 3013-3022. Abstract Full Text
322. Mohr, G., Del Campo, M., Mohr, S., Yang, Q., Jia, H., Jankowsky, E. and Lambowitz, A.M. (2008) **Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro.** *J Mol Biol*, **375**, 1344-1364. Abstract Full Text
323. Diges, C.M. and Uhlenbeck, O.C. (2001) **Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA.** *EMBO J*, **20**, 5503-5512. Abstract Full Text
324. Tsu, C.A., Kossen, K. and Uhlenbeck, O.C. (2001) **The Escherichia coli DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA.** *RNA*, **7**, 702-709. Abstract Full Text
325. Kossen, K., Karginov, F.V. and Uhlenbeck, O.C. (2002) **The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA.** *J Mol Biol*, **324**, 625-636. Abstract Full Text
326. Gibson, T.J. and Thompson, J.D. (1994) **Detection of dsRNA-binding domains in RNA helicase A and Drosophila maleless: implications for monomeric RNA helicases.** *Nucleic Acids Res*, **22**, 2552-2556. Abstract Full Text
327. Ji, X., Sun, H., Zhou, H., Xiang, J., Tang, Y. and Zhao, C. (2011) **Research progress of RNA quadruplex.** *Nucleic Acid Ther*, **21**, 185-200. Abstract Full Text
328. Fry, M. and Loeb, L.A. (1994) **The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure.** *Proc Natl Acad Sci U S A*, **91**, 4950-4954. Abstract Full Text
329. Usdin, K. and Woodford, K.J. (1995) **CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro.** *Nucleic Acids Res*, **23**, 4202-4209. Abstract Full Text
330. Collins, K. (2008) **Physiological assembly and activity of human telomerase complexes.** *Mech Ageing Dev*, **129**, 91-98. Abstract Full Text

331. Mitchell, J.R., Wood, E. and Collins, K. (1999) **A telomerase component is defective in the human disease dyskeratosis congenita.** *Nature*, **402**, 551-555. Abstract Full Text
332. Bohnsack, M.T., Martin, R., Granneman, S., Ruprecht, M., Schleiff, E. and Tollervey, D. (2009) **Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis.** *Mol Cell*, **36**, 583-592. Abstract Full Text
333. Savage, S.A. and Alter, B.P. (2009) **Dyskeratosis congenita.** *Hematol Oncol Clin North Am*, **23**, 215-231. Abstract Full Text
334. Canela, A., Vera, E., Klatt, P. and Blasco, M.A. (2007) **High-throughput telomere length quantification by FISH and its application to human population studies.** *Proc Natl Acad Sci U S A*, **104**, 5300-5305. Abstract Full Text
335. Halder, K., Wieland, M. and Hartig, J.S. (2009) **Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes.** *Nucleic Acids Res*, **37**, 6811-6817. Abstract Full Text
336. Chen, J.L., Blasco, M.A. and Greider, C.W. (2000) **Secondary structure of vertebrate telomerase RNA.** *Cell*, **100**, 503-514. Abstract Full Text
337. Chen, J.L. and Greider, C.W. (2003) **Template boundary definition in mammalian telomerase.** *Genes Dev*, **17**, 2747-2752. Abstract Full Text
338. Parkinson, E.K., Fitchett, C. and Cereser, B. (2008) **Dissecting the non-canonical functions of telomerase.** *Cytogenet Genome Res*, **122**, 273-280. Abstract Full Text
339. Krieger, E., Nabuurs, S.B. and Vriend, G. (2003) **Homology modeling.** *Methods Biochem Anal*, **44**, 509-523. Abstract
340. Xiang, Z. (2006) **Advances in homology protein structure modeling.** *Curr Protein Pept Sci*, **7**, 217-227. Abstract
341. Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S.Y., Ghaffari, A.A., Qin, J., Cheng, G. and Liu, Y.J. (2011) **DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells.** *Immunity*, **34**, 866-878. Abstract Full Text
342. Kim, H.N., Lee, J.H., Bae, S.C., Ryoo, H.M., Kim, H.H., Ha, H. and Lee, Z.H. (2011) **Histone deacetylase inhibitor MS-275 stimulates bone formation in part by enhancing Dhx36-mediated TNAP transcription.** *J Bone Miner Res*, **26**, 2161-2173. Abstract Full Text
343. Huang, W., Smaldino, P.J., Zhang, Q., Miller, L.D., Cao, P., Stadelman, K., Wan, M., Giri, B., Lei, M., Nagamine, Y. *et al.* (2012) **Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1.** *Nucleic Acids Res*, **40**, 1033-1049. Abstract Full Text
344. Subramanian, M., Rage, F., Tabet, R., Flatter, E., Mandel, J.L. and Moine, H. (2011) **G-quadruplex RNA structure as a signal for neurite mRNA targeting.** *EMBO Rep*, **12**, 697-704. Abstract Full Text
345. Hartmann, G., Weiner, G.J. and Krieg, A.M. (1999) **CpG DNA: a potent signal for growth, activation, and maturation of human dendritic cells.** *Proc Natl Acad Sci U S A*, **96**, 9305-9310. Abstract
346. Krug, A., Rothenfusser, S., Hornung, V., Jahrsdorfer, B., Blackwell, S., Ballas, Z.K., Endres, S., Krieg, A.M. and Hartmann, G. (2001) **Identification of CpG oligonucleotide sequences with high induction of IFN-alpha/beta in plasmacytoid dendritic cells.** *Eur J Immunol*, **31**, 2154-2163. Abstract Full Text
347. Kerkmann, M., Costa, L.T., Richter, C., Rothenfusser, S., Battiany, J., Hornung, V., Johnson, J., Englert, S., Ketterer, T., Heckl, W. *et al.* (2005) **Spontaneous formation of nucleic acid-based nanoparticles is responsible for high interferon-alpha induction by CpG-A in plasmacytoid dendritic cells.** *J Biol Chem*, **280**, 8086-8093. Abstract Full Text
348. Fry, M. (2007) **Tetraplex DNA and its interacting proteins.** *Front Biosci*, **12**, 4336-4351. Abstract Full Text
349. Sissi, C., Gatto, B. and Palumbo, M. (2011) **The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective.** *Biochimie*, **93**, 1219-1230. Abstract Full Text
350. Padmanabhan, K. and Tulinsky, A. (1996) **An ambiguous structure of a DNA 15-mer thrombin complex.** *Acta Crystallogr D Biol Crystallogr*, **52**, 272-282. Abstract Full Text
351. Russo Krauss, I., Merlino, A., Giancola, C., Randazzo, A., Mazzarella, L. and Sica, F. (2011) **Thrombin-aptamer recognition: a revealed ambiguity.** *Nucleic Acids Res*, **39**, 7858-7867. Abstract Full Text
352. Horvath, M.P. and Schultz, S.C. (2001) **DNA G-quartets in a 1.86 A resolution structure of an Oxytricha nova telomeric protein-DNA complex.** *J Mol Biol*, **310**, 367-377. Abstract Full Text
353. Phan, A.T., Kuryavyi, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D. *et al.* (2011) **Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction.** *Nat Struct Mol Biol*, **18**, 796-804. Abstract Full Text
354. Wen, J.D. and Gray, D.M. (2002) **The Ff gene 5 single-stranded DNA-binding protein binds to the transiently folded form of an intramolecular G-quadruplex.** *Biochemistry*, **41**, 11438-11448. Abstract
355. Gonzalez, V. and Hurley, L.H. (2010) **The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity.** *Biochemistry*, **49**, 9706-9714. Abstract Full Text
356. Mohaghegh, P., Karow, J.K., Brosh, R.M., Jr., Bohr, V.A. and Hickson, I.D. (2001) **The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases.** *Nucleic Acids Res*, **29**, 2843-2849. Abstract
357. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) **Specific association of human telomerase activity with immortal cells and cancer.** *Science*, **266**, 2011-2015. Abstract
358. Shay, J.W. and Wright, W.E. (2006) **Telomerase therapeutics for cancer: challenges and new directions.** *Nat Rev Drug Discov*, **5**, 577-584. Abstract Full Text
359. De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.F. and Mergny, J.L. (2008) **Targeting telomeres and telomerase.** *Biochimie*, **90**, 131-155. Abstract Full Text
360. Monchaud, D. and Teulade-Fichou, M.P. (2008) **A hitchhiker's guide to G-quadruplex ligands.** *Org Biomol Chem*, **6**, 627-636. Abstract Full Text
361. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) **Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?** *Nat Rev Drug Discov*, **10**, 261-275. Abstract Full Text
362. Lipps, H.J. and Rhodes, D. (2009) **G-quadruplex structures: in vivo evidence and function.** *Trends Cell Biol*, **19**, 414-422. Abstract Full Text
363. Lacroix, L., Seosse, A. and Mergny, J.L. (2011) **Fluorescence-based duplex-quadruplex competition test to screen for telomerase RNA quadruplex ligands.** *Nucleic Acids Res*, **39**, e21. Abstract Full Text
364. Bates, P.J., Laber, D.A., Miller, D.M., Thomas, S.D. and Trent, J.O. (2009) **Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer.** *Exp Mol Pathol*, **86**, 151-164. Abstract Full Text
365. Reyes-Reyes, E.M., Teng, Y. and Bates, P.J. (2010) **A new paradigm for aptamer therapeutic AS1411 action: uptake by macropinocytosis and its stimulation by a nucleolin-dependent mechanism.** *Cancer Res*, **70**, 8617-8629. Abstract Full Text
366. Krieg, A.M. (2008) **Toll-like receptor 9 (TLR9) agonists in the treatment of cancer.** *Oncogene*, **27**, 161-167. Abstract Full Text
367. Katoh, K. and Toh, H. (2008) **Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework.** *BMC Bioinformatics*, **9**, 212. Abstract Full Text
368. Gille, C. and Frommel, C. (2001) **STRAP: editor for STRuctural Alignments of Proteins.** *Bioinformatics*, **17**, 377-378. Abstract Full Text
369. Gille, C., Lorenzen, S., Michalsky, E. and Frommel, C. (2003) **KISS for STRAP: user extensions for a protein alignment editor.** *Bioinformatics*, **19**, 2489-2491. Abstract Full Text
370. Chintapalli, V.R., Wang, J. and Dow, J.A. (2007) **Using FlyAtlas to identify better Drosophila melanogaster models of human disease.** *Nat Genet*, **39**, 715-720. Abstract Full Text
371. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) **WebLogo: a sequence logo generator.** *Genome Res*, **14**, 1188-1190. Abstract Full Text

CHAPTER 8

Publications

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 50, pp. 35186–35198, December 12, 2008

Recruitment of the RNA Helicase RHAU to Stress Granules via a Unique RNA-binding Domain*[S]

Received for publication, June 25, 2008, and in revised form, September 16, 2008 Published, JBC Papers in Press, October 14, 2008, DOI 10.1074/jbc.M804857200

Kateřina Chalupníková[‡], Simon Lattmann[‡], Nives Selak[‡], Fumiko Iwamoto[‡1], Yukio Fujiki[§], and Yoshikuni Nagamine[‡2]

From the [‡]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, 4058 Basel, Switzerland and the [§]Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-8581, Japan

In response to environmental stress, the translation machinery of cells is reprogrammed. The majority of actively translated mRNAs are released from polysomes and driven to specific cytoplasmic foci called stress granules (SGs) where dynamic changes in protein-RNA interaction determine the subsequent fate of mRNAs. Here we show that the DEAH box RNA helicase RHAU is a novel SG-associated protein. Although RHAU protein was originally identified as an AU-rich element-associated protein involved in urokinase-type plasminogen activator mRNA decay, it was not clear whether RHAU could directly interact with RNA. We have demonstrated that RHAU physically interacts with RNA *in vitro* and *in vivo* through a newly identified N-terminal RNA-binding domain, which was found to be both essential and sufficient for RHAU localization in SGs. We have also shown that the ATPase activity of RHAU plays a role in the RNA interaction and in the regulation of protein retention in SGs. Thus, our results show that RHAU is the fourth RNA helicase detected in SGs, after rck/p54, DDX3, and eIF4A, and that its association with SGs is dynamic and mediated by an RHAU-specific RNA-binding domain.

Posttranscriptional regulation of gene expression is important and highly regulated in response to developmental, environmental, and metabolic signals (1). During stress conditions such as heat shock, oxidative stress, ischemia, or viral infection, mRNA translation is reprogrammed and allows the selective synthesis of stress response and repair proteins (2). Under these conditions, the translation of housekeeping genes is arrested, and untranslated mRNAs accumulate in cytoplasmic foci known as stress granules (SGs) (3). SG[3] formation is triggered by translation initiation arrest involving eIF2α phosphorylation or by inhibition of eIF4A helicase function. Phosphorylation of eIF2α can be mediated by several stress-induced protein kinases (for a review, see Ref. 4), suggesting that the formation of SGs is highly regulated and that eIF2α plays a pivotal role in sensing stress. Upon translational arrest, polysome-free 48 S preinitiation complexes containing initiation factors, small ribosomal subunits, and poly(A)-binding protein 1 (PABP-1) aggregate into SGs (3, 5). Several known SG-associated mRNA-binding proteins have been identified and shown to induce or inhibit SG aggregation when overexpressed. It is presumed that the overexpression of mRNA-binding proteins, which are able to oligomerize, disturb the equilibrium of mRNA distribution between polysomes and polysome-free ribonucleoprotein (RNP) complexes and thus induce SG formation by their aggregation (6). Nevertheless some RNA-binding proteins do not induce SG formation upon overexpression, *e.g.* a zipcode-binding protein 1 (ZBP1), heterogeneous nuclear RNP A1, or PABP-1 (7–9). Under normal conditions, most RNA-binding proteins are involved in various aspects of mRNA metabolism, such as translation (TIA-1, TIA-1-related protein, PCBP2, Pumilio 2, and cytoplasmic polyadenylation element-binding protein), degradation (G3BP, TTP, Brf1, rck/p54, KH domain RNA binding protein [KSRP], and PMR1), stability (HuR), and specific intracellular localization (ZBP1, Staufen, Smaug, Caprin-1, and FMRP) (for a review, see Ref. 10). The differing flux of these SG-associated proteins and poly(A)$^+$ mRNAs revealed by fluorescent recovery after photobleaching (FRAP) analysis suggests that SGs are dynamic foci (6, 9, 11–14). They are also considered to be sites at which RNPs undergo structural and compositional remodeling and may be temporally stored, returned to polysomes for translation, or packaged for degradation (6).

Immediately after transcription, RNA forms with RNA-binding proteins RNPs that are dynamic and take part in RNA metabolism. RNP remodeling, which is essential for the cellular localization, processing, function, and fate of RNA (15), is mainly regulated by a large family of proteins called RNA helicases. These enzymes use energy released by ATP hydrolysis to unwind secondary structures of RNA or displace proteins from RNA (16). The majority of RNA helicases are assigned to super-

* This work was supported in part by CREST grant (to Y. F.) from The Science and Technology Agency of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1 and 2.

[1] Present address: Takara Bio Inc., Seta 3-4-1, Otsu, Shiga 520-2134, Japan.

[2] To whom correspondence should be addressed. Tel.: 41-61-697-6669; Fax: 41-61-697-3976; E-mail: Yoshikuni.nagamine@fmi.ch.

[3] The abbreviations used are: SG, stress granule; RNP, ribonucleoprotein; FRAP, fluorescence recovery after photobleaching; CCCP, carbonyl cyanide *m*-chlorophenylhydrazone; CLIP, cross-linking immunoprecipitation; ARE, AU-rich element; RSM, RHAU-specific motif; uPA, urokinase-type plasminogen activator; PABP-1, poly(A)-binding protein 1; ZBP1, zipcode-binding protein 1; aa, amino acids; TTP, tristetraprolin; HuR, ELAV-like protein 1 (Hu-antigen R); FMRP, Fragile X mental retardation; PBS, phosphate-buffered saline; PIPES, 1,4-piperazinediethanesulfonic acid; DAPI, 4′,6-diamidino-2-phenylindole; bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; GST, glutathione *S*-transferase; RISP, RNA-Interaction Site Prediction; ROI, regions of interest; EGFP, enhanced green fluorescent protein; Nter, N terminus; HCR, helicase core region; Cter, C terminus; WT, full-length RHAU (wild type); β-gal, β-galactosidase.

family 2 (SF2), which is divided into three subfamilies named after the sequence in the helicase motif II: DEAD, DEAH, and DE*X*H (17). Several DE*X*(H/D) proteins have been shown to unwind double-stranded RNA in an ATP-dependent manner *in vitro* (16, 18), but most are involved in the ATP-dependent remodeling of RNPs (16). Although RNA helicases contain a highly conserved helicase core region, they are involved in all of the RNA processes ranging from transcription, pre-mRNA splicing, ribosome biogenesis, RNA export, and translation initiation to RNA decay (for reviews, see Refs. 19–21). The specific functions of individual enzymes are attributed to the less conserved N/C termini, which are responsible for substrate specificity, subcellular localization, and cofactor requirements (22–25).

RHAU is a DEAH box helicase (DHX36) originally identified as an RNA helicase associated with AU-rich element (ARE) of urokinase-type plasminogen activator (uPA) mRNA together with NF90 and HuR (26). RHAU is a nucleocytoplasmic shuttling protein found predominantly in the nucleus and to a lesser extent in the cytoplasm. As with other helicases (27–29), ATPase activity is necessary for RHAU function in the decay of uPA mRNA and for its nuclear localization (26, 30). Despite the role of RHAU as a factor destabilizing uPA mRNA, global analysis of gene expression in RHAU knockdown cells revealed that changes in steady-state levels of mRNAs were only partially influenced by mRNA decay regulation (30). Indeed the nuclear localization of RHAU and its guanine quadruplex (G4) DNA/RNA-resolving activity (31) may reflect that RHAU regulates gene expression at various steps other than mRNA decay.

Given that SGs are sites at which dynamic changes in protein-RNA interaction determine the fate of mRNAs during and after stress, it is surprising that the role of SG-associated RNA helicases with an essential function in remodeling protein-RNA interactions is largely unknown. Here we show that RHAU is a novel component of SGs and that its recruitment to SGs is mediated by an RNA interaction. Although identified as an ARE-associated protein involved in ARE-mediated mRNA decay of uPA, it was not immediately clear whether RHAU directly binds RNA. Here we show that RHAU physically interacts with RNA *in vitro* and *in vivo* via a unique N-terminal RNA-binding domain composed of a G-rich region and an RHAU-specific motif (RSM) that is highly conserved between RHAU orthologs. The same RNA-binding domain is necessary and sufficient for RHAU recruitment to SGs. Finally we show that the ATPase activity of RHAU is involved in the dynamic regulation of RHAU shuttling into and out of SGs.

EXPERIMENTAL PROCEDURES

Plasmid Constructs—The plasmids pTER-shRHAU and pTER-shLuc were described previously (30). Plasmids EGFP-RHAU, EGFP-Nter, EGFP-HCR, and EGFP-Cter were based on pEGFP-C1 (Clontech) and EGFP-E335A (also termed DAIH in this study), which was derived from EGFP-RHAU by point mutation in motif II as described previously (30). Plasmids RHAU-EGFP, (50–1008)-EGFP, (105–1008)-EGFP, (1–130)-EGFP, and (1–105)-EGFP were based on pEGFP-N1 (Clontech) by inserting corresponding fragments into the BglII (XhoI for RHAU full length) and AgeI sites of the vector. Plasmids EGFP-(50–1008) and EGFP-(105–1008) were prepared by inserting the corresponding fragments between BglII and BamHI sites of pEGFP-C1. RHAU-FLAG, (50–1008)-FLAG, (105–1008)-FLAG, (1–105)-FLAG, and (1–130)-FLAG were prepared by inserting corresponding RHAU fragments into the BglII (NheI for RHAU full length) and AgeI sites of pIRES.EGFP-N1-FLAG. The vector was prepared by insertion of IRES.EGFP between AgeI and NotI of pEGFP-N1 (Clontech). The IRES.EGFP insert was designed by PCR using the pIRES.ECMV-EGFP vector, which was kindly provided by D. Schmitz Rohmer and B. A. Hemmings, as a template. The PCR product of the IRES.EGFP insert contained FLAG sequence with the AgeI site on the 5′ end and the NotI site on the 3′ end. Plasmids β-gal-(1–52)-EGFP, β-gal-(1–130)-EGFP, and β-gal-(52–200)-EGFP were prepared by inserting corresponding RHAU fragments into EcoRI and SalI sites of pβ-gal-EGFP.[4] The GST-RHAU vector was designed as described previously (26). GST-Nter was prepared by inserting the fragment (1–200 aa) between BamHI and EcoRI sites of pGEX-2T (Amersham Biosciences). The oligonucleotides used in this work and detailed descriptions of the plasmid construction are available upon request.

Antibodies—Mouse monoclonal anti-RHAU antibody (12F33) was generated against a peptide corresponding to the C terminus of RHAU, 991–1007 aa, as described previously (31). Commercially obtained antibodies were: mouse anti-green fluorescent protein (B-2, sc-9996), goat anti-TIA-1 (sc-1751), goat anti-eIF3b (N-20, sc-16377), mouse anti-HuR (3A2, sc-5261), and rabbit anti-eIF2α (FL-315, sc-11386) (Santa Cruz Biotechnology, Santa Cruz, CA); mouse anti-actin (pan Ab-5, Thermo Fisher Scientific, Fremont, CA); rabbit monoclonal anti-eIF2α-P (Ser-51, 119A11, Cell Signaling Technology, Danvers, MA); and mouse anti-FLAG M2 (Sigma-Aldrich). The mouse antibodies were all monoclonal.

Cell Culture, Transfection, and Stress Treatments—HeLa cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at 37 °C in the presence of 5% CO_2. T-REx-HeLa cells stably transfected with pTER-shRHAU were maintained as described previously (30). To induce short hairpin RNA expression and consequent depletion of endogenous RHAU, cells were treated with 1 μg/ml doxycycline (Sigma-Aldrich) for 7 days as described in supplemental Fig. 1. For immunofluorescence analysis, transient transfection of DNA plasmids using FuGENE 6 (Roche Applied Science) was performed according to the manual provided using 1 μg of plasmid DNA and 3 μl of FuGENE 6/35-mm dish. For immunoprecipitation analysis, cells were transfected by Lipofectamine 2000 (Invitrogen) in Opti-MEM I medium (Invitrogen). Briefly HeLa cells were seeded at 0.8×10^6 cells/35-mm dish and 24 h later were transfected with 4 μg of plasmid DNA and 10 μl of Lipofectamine 2000. The cells were used 24 h later for the following experiments. RNA interference silencing was induced by transient transfection of small interfering RNAs with INTERFERin (Polyplus Transfection, New York, NY) following the manual instructions. Small interfering RNA was added to give a final concentration of 2.5 nM in 2 ml of

[4] F. Iwamoto and Y. Fujiki, unpublished data.

medium and 8 μl of INTERFERin for transfection of 40% confluent cells in each 35-mm dish. To test SG formation, 0.5 mM sodium arsenite (Sigma-Aldrich) or 1 μM hippuristanol (kindly provided by J. Tanaka (32)) was added in conditioned medium for 45 or 30 min, respectively. To induce SGs with the ionophore carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), cells were washed with 1× PBS^- (PBS without Ca^{2+} and Mg^{2+}) and cultured in glucose- and pyruvate-free Dulbecco's modified Eagle's medium containing 1 μM CCCP. Heat shock was performed at 44 °C in a 5% CO_2 incubator for 45 min.

Immunocytochemistry and Image Processing—At 24 h after transfection by FuGENE 6, HeLa cells were replated in 12-well dishes with coverslips coated with 0.2% gelatin. The next day, cells were treated with the indicated stimuli, fixed with 3.8% paraformaldehyde in 1× PBS^- for 10 min, permeabilized with 0.2% Triton X-100 in PHEM buffer (25 mM HEPES, 10 mM EGTA, 60 mM PIPES, 2 mM $MgCl_2$, pH 6.9) for 30 min and blocked with 5% horse serum in PHEM buffer for 1 h. All steps were performed at room temperature. Samples were incubated with primary antibodies (goat anti-TIA-1 (1:200), mouse anti-HuR (1:200), and goat anti-eIF3b (1:200)) diluted in the blocking buffer overnight at 4 °C. Coverslips with fixed cells were washed three times with 0.2% Triton X-100 in PHEM buffer and incubated in the dark with the secondary antibody and 500 ng/ml DAPI (Santa Cruz Biotechnology) to identify the nuclei for 40 min at room temperature. Cy2-, Cy3-, or Cy5-conjugated donkey secondary antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA) were used at dilutions of 1:200, 1:2,000, or 1:200 with 2.5% horse serum in PHEM buffer, respectively. The cells were mounted in FluoroMount reagent (SouthernBiotech, Birmingham, AL). Fluorescent images were captured with a confocal microscope (LSM 510 META, Carl Zeiss, Jena, Germany) as described previously (30) except that a Plan-NeoFluar 100×/1.3 oil differential interference contrast objective was used. The data obtained were processed using standard image software (Bitplane Imaris 5.7.1, Adobe Photoshop, and Adobe Illustrator). To quantify association of RHAU with SGs, at least 100 transfected cells were analyzed under the wide spectrum microscope (Axioplan 2, Carl Zeiss) and scored as positive when the green fluorescent protein signal was enriched and co-localized with TIA-1 in SGs. Three ($n = 3$) independent transfections were analyzed to calculate mean percentages and ±S.E.

Protein Extraction and Western Blotting—To prepare total cell lysates, cells were lysed with Nonidet P-40 buffer (50 mM Tris-HCl, pH 7.5, 120 mM NaCl, 1% Nonidet P-40, 1 mM EDTA, 5 mM Na_3VO_4, 5 mM NaF, 0.5 μg/ml aprotinin, 1 μg/ml leupeptin) on ice for 10 min and centrifuged at 20,000 × *g* for 15 min at 4 °C to remove cell debris. Typically 30 μg of the total cell lysate was loaded for Western blotting. The protein bands were visualized with the direct infrared fluorescence method or the chemiluminescence method as described previously (30).

Cross-linking Immunoprecipitation (CLIP)—RHAU and RNA interaction was determined by the previously reported CLIP method with slight modifications (33). HeLa cells (0.8×10^6/35-mm dish) were rinsed twice with ice-cold PBS and then UV irradiated (400 mJ/cm^2) to induce cross-linking between protein and RNA. Cells were then lysed with 200 μl of RIPA buffer (1% Nonidet P-40, 1% deoxycholate, 0.1% SDS, 50 mM NaCl, 10 mM sodium phosphate, pH 7.2, 2 mM EDTA, 50 mM NaF, 0.2 mM sodium vanadate, 100 units/ml aprotinin)/well in a 6-well dish and shaken for 15 min at 4 °C. Pooled lysates from 6 wells were treated with 30 μl of RQ1 RNase-free DNase (1 unit/μl; Promega, Madison, WI) and with 31 units of RNase A (31 units/μl; USB Corp.) as described in the CLIP protocol (33). Treated samples were centrifuged at 20,000 × *g* for 20 min at 4 °C. The supernatants (600 μg) were incubated with 4.5 μg of a mouse anti-RHAU monoclonal antibody (12F33) or 10 μl (bed volume) of anti-FLAG M2 affinity gel (A2220, Sigma-Aldrich) with rotation for 2 h at 4 °C. Beads were washed twice with RIPA buffer, twice with high salt washing buffer (5× PBS, 0.1% SDS, 0.5% deoxycholate, 0.5% Nonidet P-40) and twice with 1× PNK buffer (50 mM Tris-HCl, pH 7.4, 10 mM $MgCl_2$, 0.5% Nonidet P-40). The associated nucleic acids were radiolabeled with [γ-^{32}P]ATP using T4 polynucleotide kinase (Roche Applied Science) as described in the CLIP protocol (33), and RHAU-RNA complexes were resolved in a NuPAGE 4–12% bis-Tris gel (Invitrogen). Immunoprecipitated RHAU was detected by Coomassie Blue staining and in-gel Western blotting according to the manual of Odyssey In-Gel Western detection (LI-COR Biosciences). Half of the samples were transferred to a polyvinylidene difluoride membrane to facilitate better protein detection by Western blotting and to remove free RNA. The proteins were detected by the Odyssey infrared imager as described above. Radiolabeled RNA was detected by a phosphorimaging system, Typhoon 9400 (GE Healthcare), and analyzed using the ImageQuant TL program. To test whether RHAU associates with RNA, bound nucleic acids were isolated and radiolabeled according to the CLIP protocol (33). Nucleic acids were mixed with increasing amounts of RNase A (0.015, 0.15, 1.5, and 15 units) in H_2O and 1 unit of RQ1 DNase in 1× RQ1 DNase reaction buffer. Reactions were incubated for 30 min at 37 °C and resolved by denaturing 8% PAGE in 1× Tris borate-EDTA buffer.

Protein Purification—Escherichia coli BL21 (DE3) transformed with glutathione *S*-transferase (GST) or GST-Nter proteins were induced by 1 mM isopropyl 1-thio-β-D-galactopyranoside for 12 h at 25 °C and purified by affinity chromatography with glutathione-Sepharose 4B (Amersham Biosciences) according to the manufacturer's instructions. GST-RHAU protein was expressed in Sf9 cells according to the supplier's instructions (BD Biosciences Pharmingen) and purified as above. The purity of recombinant proteins was analyzed by 10% SDS-PAGE and Coomassie Blue staining as reported in supplemental Fig. 2.

Double Filter RNA Binding Assay—5 μg of total RNA isolated from HeLa cells was alkali-treated with 0.1 M NaOH on ice for 10 min and then EtOH-precipitated. Redissolved RNA or poly(rU) (P9528, Sigma-Aldrich) was 5′-end-labeled using [γ-^{32}P]ATP (Hartmann Analytic GmbH, Braunschweig, Germany) and T4 polynucleotide kinase (Roche Applied Science) at 37 °C for 30 min and passed through a G-50 column (GE Healthcare) to remove free nucleotides. Reaction mixtures (50 μl) containing varying amounts of recombinant proteins (0–150 nM as specified in text), radiolabeled RNA (poly(rU), 10,000 cpm; total RNA, 5,000 cpm) and 2 units of RNase inhibitor (RNAguard, Roche Applied Science) in the binding buffer

(50 mM Tris-HCl, pH 8.0, 1 mM dithiothreitol, 5 mM NaCl) were incubated for 30 min at 37 °C. The double filter RNA binding assay was performed with a slot-blot apparatus using a 0.45-μm nitrocellulose (Protran, Whatman) and nylon membranes (positively charged; Roche Diagnostics) that was presoaked in different buffers as described previously (34). Loaded samples were washed three times with 200 μl of the binding buffer. Retained RNA was detected with a phosphorimaging system, Typhoon 9400, and analyzed using the ImageQuant TL program. The nitrocellulose membrane retains only RNA-protein complexes, and free RNAs are captured on the nylon membrane. The ratio of RNA that was bound to GST-RHAU or GST-Nter was calculate using the following formula: bound RNA (%) = $100((\text{signal}_{\text{nitrocellulose}})/(\text{signal}_{\text{nitrocellulose}} + \text{signal}_{\text{nylon}}))$.

Bioinformatics—The program RNABindR (35) was used to predict RNA binding potential in amino acid sequences. Programs RISP (RNA-Interaction Site Prediction) and BindN+ were used to confirm the reliability of the RNABindR program: RISP (36) runs with 72.2% RNA binding prediction accuracy, and BindN+ (37) runs with 68% RNA binding prediction accuracy. For multiple sequence alignments of N termini, RHAU orthologs were identified by a BLASTP (version 2.2.18+) search of non-redundant protein entries in the NCBI data base using the entire sequence of RHAU as a query. Multiple sequence alignment was carried out with ProbCons (version 1.12) (38). Similarity of groups was generated using GeneDoc (version 2.7) with the BLOSUM62 scoring matrix.

Fluorescence Recovery after Photobleaching—HeLa cells (1.5×10^5/35-mm dish) were plated and transfected by FuGENE 6 on glass-bottomed dishes (Micro-Dish 35 mm, Fisher Scientific). FRAP experiments were carried out with a confocal microscope (LSM 510, Carl Zeiss) using a 63×/1.4 numerical aperture oil differential interference contrast objective. Bleaching was performed using the 488-nm lines from a 40-milliwatt argon laser operating at 75% laser power. Bleaching of circular regions of interest (ROI) was done with three scan iterations lasting a total of 1.5 s. Fluorescence recovery was monitored at low laser intensity (~2% of the 40-milliwatt laser) at 0.5-s intervals. Two additional ROI were monitored in parallel to detect fluorescence fluctuations that were independent of bleaching (ROI of non-bleached SG) and to extract a background (ROI in the cytoplasm). Raw data of one protein were obtained from ≥10 different FRAP analyses from three independent transfections. Fluorescence recovery was normalized with the function $((I_t - B_t)/(C_t - B_t))/I_0$ where I_t is the mean intensity in ROI of bleached SGs at time point t, B_t is the mean intensity in ROI of the background at time point t, C_t is the mean intensity in ROI of control SGs at time point t, and I_0 the average intensity in ROI of bleached SGs during prebleach. The average fluorescence intensities ±S.E. from at least ≥10 independent FRAP analyses were extrapolated to the graph as recovery curves.

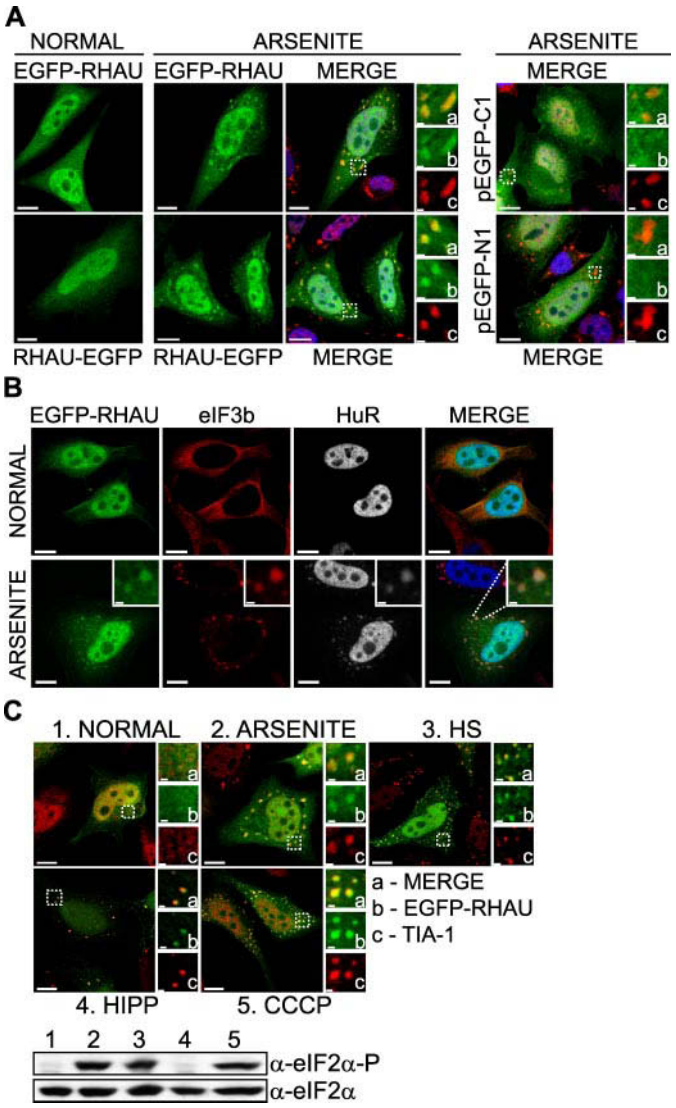


FIGURE 1. **RHAU helicase accumulates in SGs in response to all tested stress stimuli.** *A*, immunofluorescence analysis of RHAU localization in arsenite-induced SGs. HeLa cells, transiently transfected with EGFP-RHAU or RHAU-EGFP for 48 h, were treated with 0.5 mM sodium arsenite for 45 min or cultured in normal conditions. After fixation, cells were stained with DAPI (*blue*) to visualize nuclei and with anti-TIA-1 antibody (*red*) to detect SGs. *Left*, distribution of EGFP-RHAU or RHAU-EGFP (*green*) under normal and arsenite-treated conditions. Panels denoted as *MERGE* show the digital combination of EGFP, TIA-1, and DAPI images under arsenite-treated conditions; *enlargements* of *boxed regions* are shown in *small panels* indicating merge (*a*), EGFP (*b*), and TIA-1 (*c*). *Right*, merges of DAPI and TIA-1 with empty vectors pEGFP-C1 or pEGFP-N1; *enlargements* of *boxed regions* are shown in *small panels* as above. *B*, co-localization of RHAU with eIF3b and HuR in arsenite-induced SGs. HeLa cells were transfected with EGFP-RHAU and treated with or without arsenite as above. The *extreme left panels* show EGFP-RHAU (*green*), and the *two middle panels* are immunofluorescent images for eIF3b (*red*) and HuR (*white*; *blue* in merge), two SG markers. The *extreme right panels* are their merges. *C*, effects of different stress stimuli on eIF2α phosphorylation and RHAU localization in SGs. HeLa cells transiently transfected with EGFP-RHAU were cultured without (*1*) or with the following stress inducers: 0.5 mM arsenite for 45 min (*2*), heat shock (*HS*) at 42 °C for 45 min (*3*), 1 μM hippuristanol (*HIPP*) for 30 min (*4*), and 1 μM CCCP for 90 min (*5*). The images show merges of EGFP-RHAU (*green*) and TIA-1 (*red*). *Enlargements* of *boxed regions* are shown in *small panels* indicating merge (*a*), EGFP (*b*), and TIA-1 (*c*). The total eIF2α and phosphorylated eIF2α (*eIF2α-P*) were detected by Western blotting. *Bar*, 10 μm (1 μm in *enlargements*).

RESULTS

RHAU Protein Associates with SGs in Response to Arsenite-induced Stress—Various research groups have shown that different ARE-binding proteins, such as TTP, Brf1, HuR, TIA-1, or TIA-1-related protein, associate with SGs in cells exposed to environmental stress (8, 39, 40). To determine whether RHAU is also recruited to SGs, HeLa cells were transfected with EGFP-

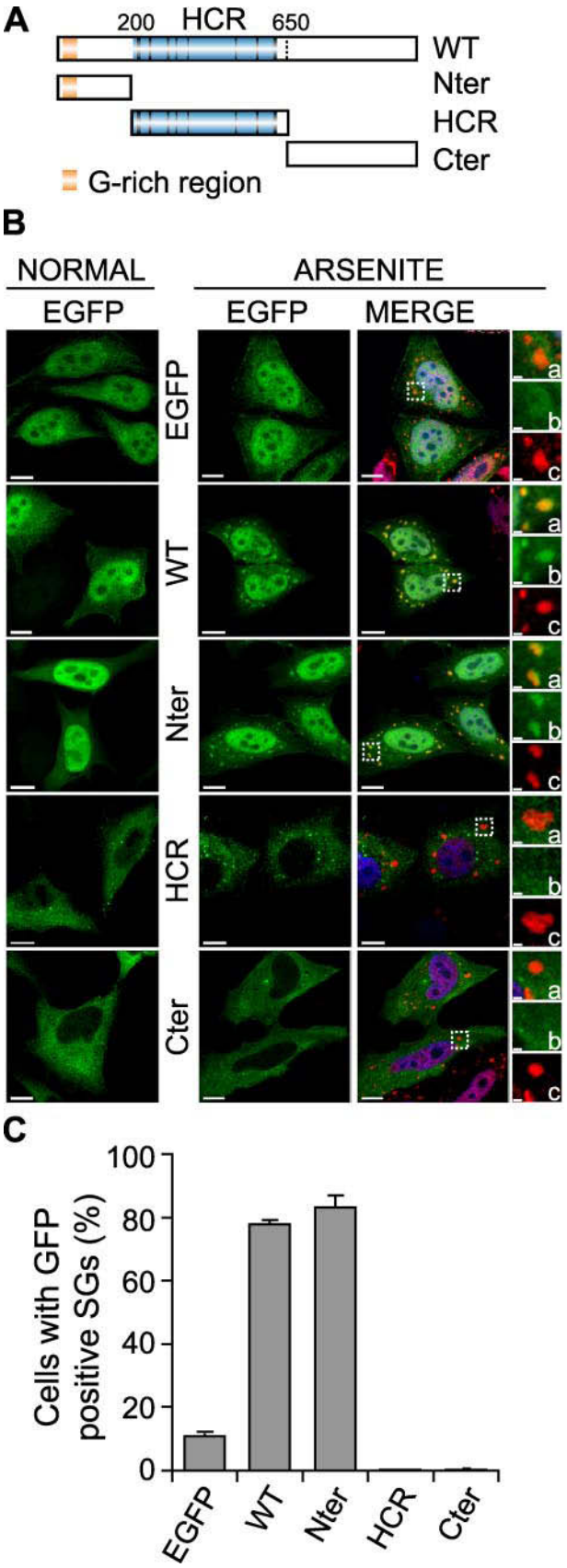


FIGURE 2. **The N-terminal domain of RHAU localizes in SGs.** *A*, scheme of wild-type RHAU and its fragments. Nter, 1–200 aa; HCR, 200–650 aa; Cter, 200–1008 aa. *B*, intracellular localization of different parts of RHAU. HeLa cells were transfected with an empty pEGFP-C1 vector or plasmids expressing EGFP-fused RHAU (*WT*) or fragments of RHAU, Nter, HCR, and Cter. After 48 h, cells were cultured in normal conditions or treated with 0.5 mM arsenite for 45 min, fixed, and stained for TIA-1 (*red*). Nuclei were visualized with DAPI (*blue*). *Small panels* show *enlargements* of *boxed regions*: merge (*a*), EGFP-fused fragments of RHAU (*b*) and TIA-1 (*c*). *C*, data obtained by quantitative immunofluorescence analysis showing the percentage of transfected cells in which EGFP signals were detected in SGs. Values ± S.E.M. (standard errors of means) were derived from three independent experiments. *Bar*, 10 μm (1 μm in *enlargements*). *GFP*, green fluorescent protein.

tagged RHAU and treated with arsenite, a typical inducer of SGs. As shown in Fig. 1*A*, the arsenite treatment induced, in addition to diffuse nuclear and cytoplasmic localizations, the accumulation of RHAU in distinct cytoplasmic foci that co-localized with TIA-1 protein, an SG marker. Because all available anti-RHAU antibodies, both commercial and our own preparation, were not suitable to visualize endogenous RHAU by immunofluorescence analysis, we examined the intracellular localization of RHAU by transiently transfecting cells. We used either EGFP-RHAU or RHAU-EGFP to eliminate the unwanted effects of EGFP on RHAU structure and subsequent subcellular localization. Indeed both EGFP-RHAU and RHAU-EGFP showed similar localization patterns in normal and stress conditions (Fig. 1*A*). EGFP itself did not accumulate in SGs (Fig. 1*A*). The localization of RHAU in arsenite-induced cytoplasmic foci was confirmed using other SG markers such as HuR and eIF3b (p116) (Fig. 1*B*), indicating that RHAU is a novel SG component.

Although the formation of SGs was not abolished or reduced in T-Rex-HeLa cells after RNA interference-mediated down-regulation of endogenous RHAU (supplemental Fig. 1), we asked whether the recruitment of RHAU to SGs is restricted to a specific type of stress. To address this question, we treated EGFP-RHAU transfected HeLa cells with different SG-inducing stimuli, such as arsenite, CCCP (a mitochondrial inhibitor), heat shock, and hippuristanol (an inhibitor of eIF4A RNA binding activity known to induce SG aggregation independently of eIF2α phosphorylation) (32). As shown in Fig. 1*C*, EGFP-RHAU co-localized with TIA-1 in SGs in response to all stimuli tested. The phosphorylation of eIF2α was detected after treatment with arsenite, heat shock, and CCCP but, as expected, not with hippuristanol (Fig. 1*C*). These results indicate that the formation of SGs, rather than phosphorylation of eIF2α, is required for RHAU protein recruitment to SGs.

The N-terminal Domain Recruits RHAU to SGs—RHAU protein consists of three main domains: the unique N terminus (Nter), the evolutionarily conserved helicase core region (HCR), and the C terminus (Cter) partially conserved among DEAH subfamily members. To identify domains essential for localization of RHAU in SGs, we generated deletion mutants of RHAU: EGFP-Nter (1–200 aa), EGFP-HCR (201–650 aa), and EGFP-Cter (651–1008 aa) (Fig. 2*A*). As shown in Fig. 2*B*, the full-length RHAU (WT) as well as the N terminus alone accumulated in arsenite-induced SGs. Unlike EGFP-WT and EGFP-Nter, EGFP-HCR and EGFP-Cter were excluded from the nucleus and diffused in the cytoplasm as described previously (30). After arsenite treatment, they showed low level accumulation in small foci other than SGs (Fig. 2*B*). A similar proportion of cells positive for EGFP signals in arsenite-

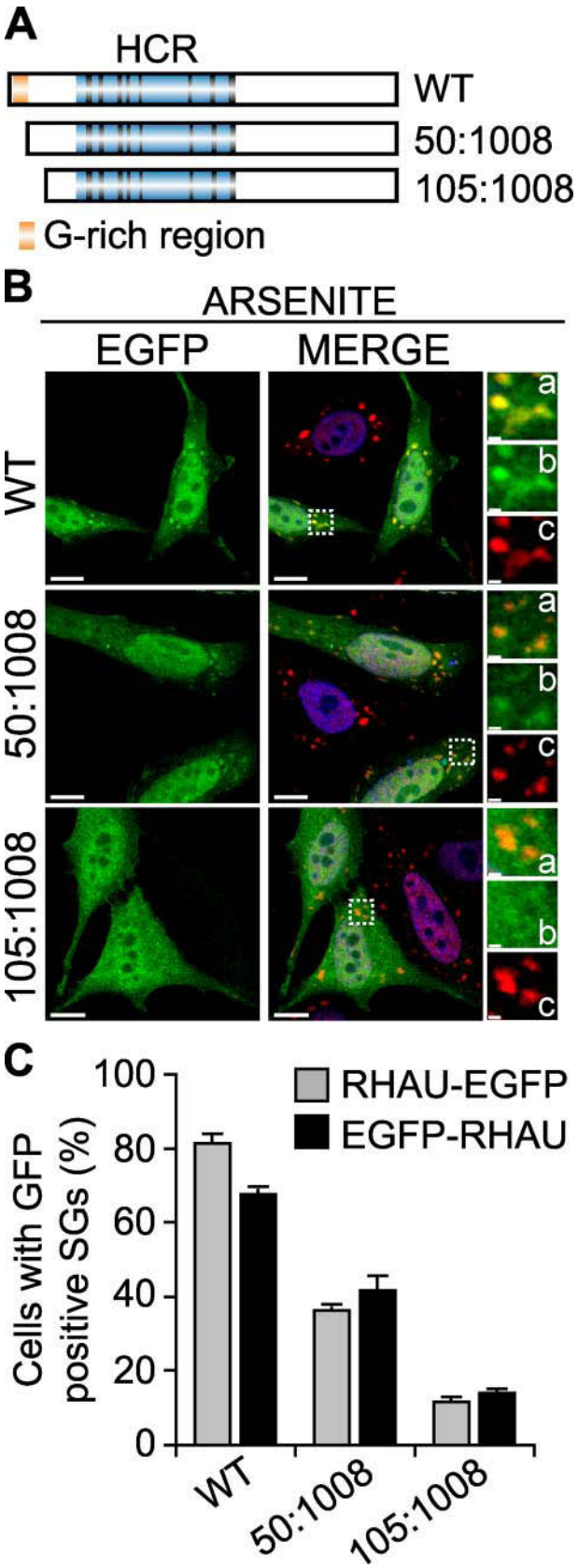


FIGURE 3. **The first N-terminal 105 amino acids of RHAU are necessary for RHAU localization in SG.** *A*, scheme of wild-type RHAU and its deletion mutants. *B*, intracellular localization of RHAU N-terminal deletion mutants. HeLa cells were transfected with plasmids expressing EGFP-fused RHAU (*WT*) or N-terminal deletion mutants of RHAU, 50–1008 and 105–1008. After 48 h, cells were treated with 0.5 mM arsenite for 45 min, fixed, and stained for TIA-1 (*red*). Nuclei were visualized with DAPI (*blue*). Small panels show enlargements of boxed regions: merge (*a*), EGFP-fused fragments of RHAU (*b*), and TIA-1 (*c*). *C*, quantitative immunofluorescence analysis showing the percentage of transfected cells in which EGFP signals were detected in SGs. Values ± S.E.M. (standard errors of means) were derived from three independent experiments. *Bar*, 10 μm (1 μm in *enlargements a–c*). *GFP*, green fluorescent protein.

induced SGs was observed with EGFP-WT (78%)- and EGFP-Nter (83%)-transfected cells, but no positive EGFP signals in arsenite-induced SGs were detected after transfection with EGFP-HCR and EGFP-Cter (Fig. 2*C*). Taken together, these results indicate that the N terminus of RHAU is essential for stress-induced accumulation of RHAU in SGs.

Because only the first 200 amino acids of RHAU were necessary for SG localization, we further studied functional domains within the N terminus. For this we generated two new N-terminally truncated forms of EGFP-fused RHAU by stepwise deletion as depicted in Fig. 3*A* where the first deletion mutant (50–1008 aa) lacked a G-rich region. These truncated mutants and WT were transfected into HeLa cells. The association of RHAU and its mutants with arsenite-induced SGs was assessed by immunofluorescence microscopy (Fig. 3*B*) and quantified (Fig. 3*C*). Deletion of the G-rich region significantly decreased RHAU association with SGs from 81 to 36% of transfected cells. The truncation of amino acids 1–105 further reduced RHAU accumulation in SGs to 12%. As shown in a graph (Fig. 3*C*), EGFP fusion at the N or C terminus with the WT or the mutants did not affect their localization efficiency in SGs. Taken together, these data suggest that the first 105 amino acids of the N-terminal domain are essential for the recruitment of RHAU protein to SGs and that an additional site, other than the G-rich region, plays a role in this recruitment process.

RHAU Binds to RNA via the N-terminal Domain—The recruitment of proteins to SGs was reported to be mediated by protein-protein or RNA-protein interactions (10). Because RHAU is an RNA helicase, we first examined whether RHAU binds to RNA and, if so, whether RNA binding activity is required for its accumulation in SGs. To test this, we used adapted CLIP (33), utilizing irreversible UV cross-linking between protein and RNA to purify protein-RNA complexes. The tightly associated RNA molecules migrating together with immunoprecipitated protein were radiolabeled with [γ-^{32}P]ATP and detected with a phosphorimaging system. Endogenous RHAU was immunoprecipitated with monoclonal anti-RHAU antibody. Phosphorimaging analysis of radiolabeled nucleic acids showed a prominent signal at 120 kDa corresponding to RHAU (Fig. 4*A*). The signal was absent when RHAU was depleted by RNA interference before immunoprecipitation. This indicated that under normal conditions the strong signal at 120 kDa was derived from RHAU-associated nucleic acids. UV irradiation not only cross-links proteins and RNA but also proteins and DNA. As the CLIP method includes a stringent DNase treatment prior to immunoprecipitation, it is likely that the detected signals involve RNA. Nevertheless to test this further, we purified co-immunoprecipitated nucleic acids, radiolabeled them, and examined their sensitivity to an increasing concentration of RNase (*lanes 2–5*) and 1 unit

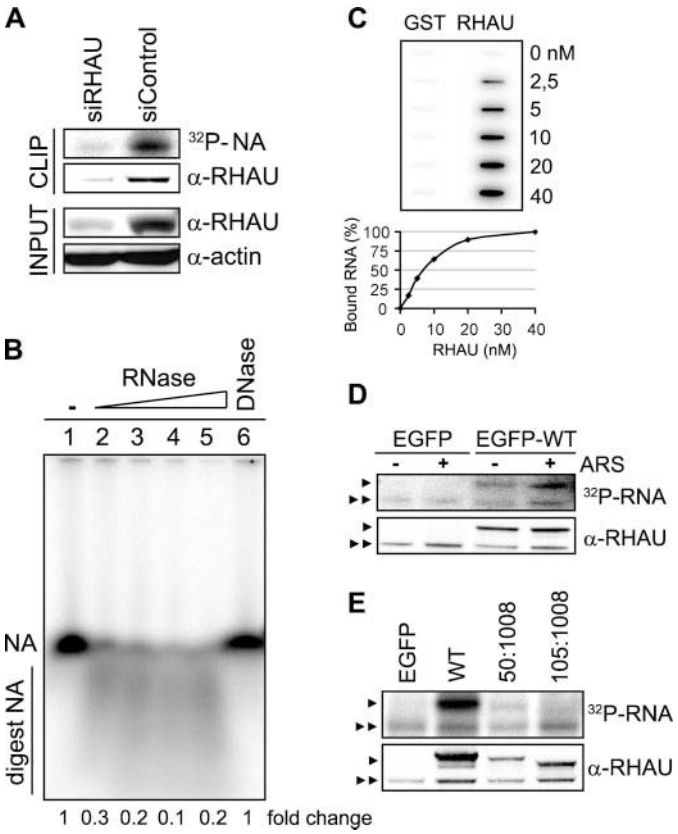


FIGURE 4. **RHAU associates with RNA through its N terminus.** *A*, CLIP method to identify interactions of RHAU with nucleic acids (*NA*). HeLa cells were transiently transfected with small interfering RNAs against RHAU (*siRHAU*) or luciferase (*siControl*). Cells were UV irradiated 72 h later and harvested. Immunoprecipitation was performed with a RHAU antibody. Bound nucleic acids were radiolabeled and detected as a band of ~120 kDa corresponding to RHAU (*lane siControl*). However, this band was strongly reduced when endogenous RHAU was depleted (*lane siRHAU*). Western blot (*INPUT*) shows efficient RHAU depletion by small interfering RNAs against RHAU. Actin was used as a loading control. *B*, identification of associated nucleic acids as RNA. After immunoprecipitation, the nucleic acids bound to RHAU were isolated, radiolabeled, and treated with increasing concentrations (0.015, 0.15, 1.5, and 15 units) of RNase A and 1 unit of RQ1 DNase. *Lane 1*, isolated and non-treated nucleic acids (*NA*); *lanes 2–5*, isolated nucleic acids treated with increasing concentrations of RNase; *lane 6*, nucleic acids treated with DNase. -Fold changes in intensity ratio (*NA/digest NA*) are depicted relative to the intensity of non-treated nucleic acids in *lane 1*, set as 1. Note that nucleic acids are RNase-sensitive but DNase-insensitive. *C*, RHAU binds to RNA *in vitro*. 0, 2.5, 5, 10, 20, and 40 nM GST or GST-RHAU (*RHAU*) was incubated with 5,000 cpm 5′-end-labeled RNA for 30 min at 37 °C and filtered through nitrocellulose and nylon membranes. The percentage of bound RNA was plotted as a function of increasing concentration of GST-RHAU. Note that dose-dependent RNA binding was seen only with GST-RHAU. *D*, effect of stress on RHAU association with RNA *in vivo*. HeLa cells were transfected with EGFP alone or EGFP-WT (a full-length RHAU). Half were treated 24 h later with arsenite (0.5 mM for 45 min) followed by UV irradiation and radiolabeling of RHAU-associated nucleic acids. RHAU and associated RNA were analyzed by Western blotting and a phosphorimaging system to detect levels of protein expression and the amount of radiolabeled RNA, respectively. Note that the arsenite treatment (+) did not abolish or dramatically change the RNA binding activity of RHAU. *ARS*, arsenite. *E*, comparison of RNA binding activities of wild-type RHAU and its N-terminal deletion mutants. HeLa cells were transfected with EGFP, EGFP-WT, EGFP-(50–1008), or EGFP-(105–1008). Protein and associated RNA were analyzed as in *A*. Radioactivity of bound RNA was normalized to the expression levels of corresponding proteins. In sharp contrast to constant RNA binding of endogenous RHAU in each lane, N-terminal deletion mutants showed stepwise reduction of RNA interaction compared with WT. ▶, overexpressed EGFP-RHAU or its N-terminal deletion mutants; ▶▶, endogenous RHAU.

of DNase (*lane 6*). RHAU-associated nucleic acids were digested by RNase but not by DNase (Fig. 4*B*) indicating that RHAU interacts with RNA *in vivo*.

To confirm our *in vivo* observation, the binding of RHAU to RNA was examined *in vitro* using purified, recombinant GST and GST-RHAU by double filter RNA binding assays (supplemental Fig. 2A and Fig. 4*C*). As we do not know yet the *in vivo* RNA targets of RHAU, we incubated radiolabeled, total RNA (5,000 cpm) from HeLa extract with increasing amounts of GST or GST-RHAU (0–40 nM) and filtered the mixture through nitrocellulose and nylon membranes. As shown in Fig. 4*C*, unlike GST itself, GST-RHAU was retained with associated RNA on nitrocellulose, and moreover it displayed dose-dependent RNA binding, suggesting that RHAU can directly associate with RNA. We express the relative binding affinity (K_D) as the concentration of GST-RHAU at which 50% of RNA was bound. The apparent K_D of GST-RHAU was 7 nM (Fig. 4*C*).

Next we examined the influence of arsenite stress on the RHAU-RNA interaction by treating the EGFP- or EGFP-WT-transfected HeLa cells with arsenite before UV irradiation and comparing the RNA binding activity of RHAU in normal and stress conditions. After immunoprecipitation by a monoclonal anti-RHAU antibody, two radiolabeled RNA bands per the lane were detected; the lower and upper bands correspond to endogenous RHAU and transfected EGFP-WT, respectively (Fig. 4*D*). The association of endogenous RHAU with RNA was used as a loading control. The intensity of the RNA signals normalized over RHAU protein levels was similar in control and arsenite-treated conditions, suggesting that the amount of RHAU-associated RNA was not influenced by stress. The RNA band migrating in the region corresponding to EGFP-WT was not present in extracts of EGFP-transfected cells, verifying that the RNA signals were derived from RHAU-associated RNA.

Given the apparent significance of the first 105 amino acids for RHAU localization in SGs, we examined whether the N terminus was also required for RHAU-RNA interaction. After cells were transiently transfected with EGFP-WT and its N-terminal deletion mutants, the extent of RNA binding to RHAU and mutants was examined as described above and normalized over RHAU protein levels. Although the CLIP method is not a suitable quantitative assay, in six independent assays EGFP-(50–1008) and EGFP-(105–1008) showed a reproducible decrease in RNA binding activity of ~50 and 20%, respectively, compared with EGFP-WT (Fig. 4*E*). A stepwise decrease in the RNA binding activity of the two deletion mutants correlated with the decreased RHAU localization in arsenite-induced SGs (see Fig. 3*C*). The small amount of RNA remaining associated with EGFP-(105–1008) may reflect residual RNA binding activity derived from the helicase core domain of RHAU. Thus, the majority of RNA associated with RHAU is bound to its N terminus.

Bioinformatics Analysis of the N Terminus Revealed a Putative RNA-binding Domain—Knowing that RHAU interacts with RNA via the first 105 amino acids, we submitted the RHAU protein sequence to RNABindR, a computational program that predicts RNA binding potential of amino acid residues (35). This program is focused on sequence-based predictions and runs on a naïve Bayes classifier. According to this program, residues from 3 to 75 were highly positive for the RNA interaction (Fig. 5*A*). A similar prediction was obtained by two other programs based on different algo-

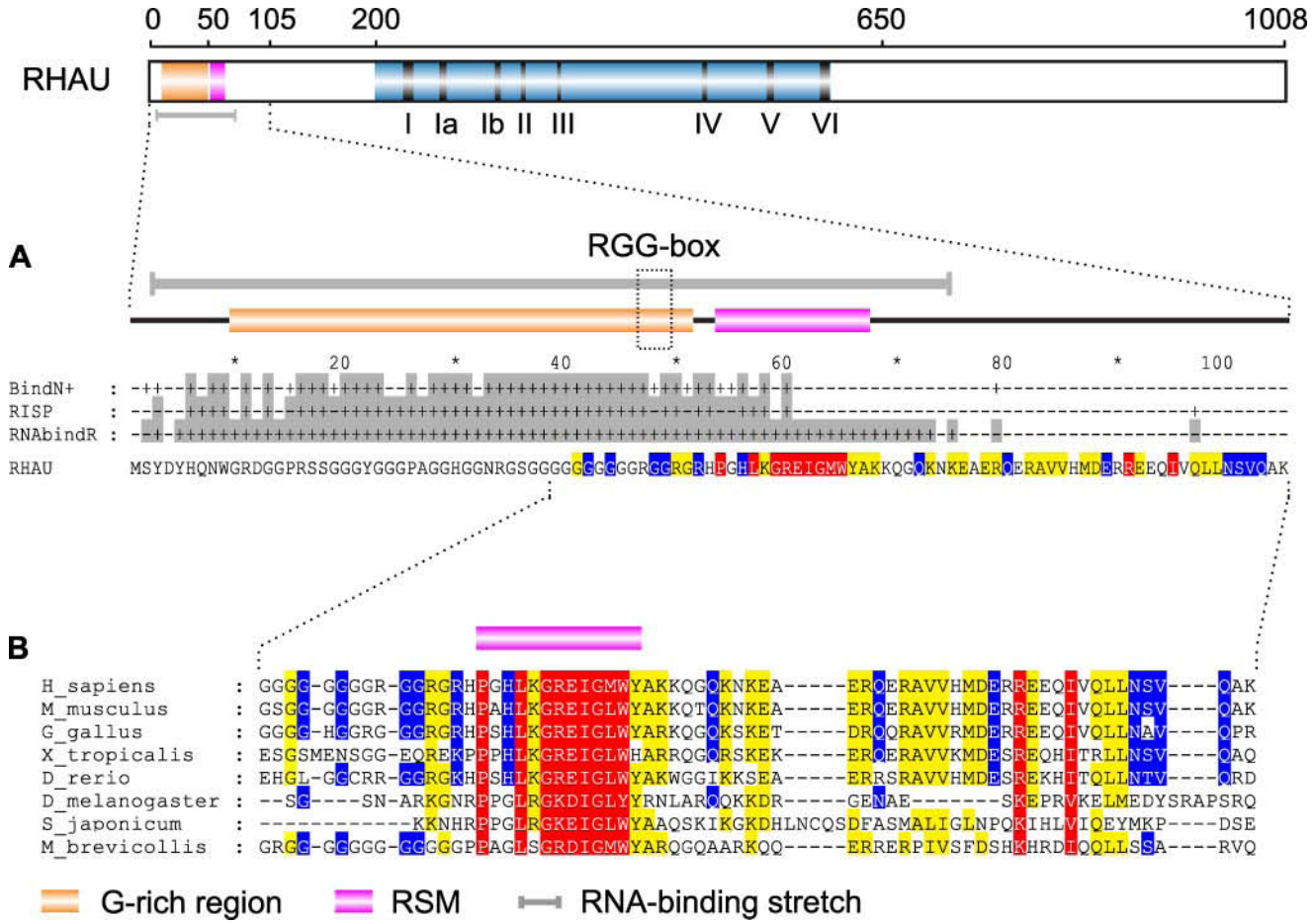


FIGURE 5. **Bioinformatics analysis of the RHAU N-terminal domain.** *A*, computational analysis for RNA-binding residues within the RHAU N terminus. The human RHAU amino acid sequence was subjected to three different programs, BindN+, RISP, and RNABindR, available on line. Residues predicted to be positive for RNA binding are indicated by "+" and depicted *above* the RHAU amino acid sequence. The *gray shading* corresponds to RNA binding consensus per one amino acid in all three programs. Note that the predicted RNA-binding domain is a long stretch from 3 to 75 aa (*gray bar*) that overlaps with the G-rich region containing an RGG box motif and with the highly conserved RSM. *B*, conservation of the RSM through evolution. Orthologs of RHAU used for the multiple sequence alignment were identified as described under "Experimental Procedures." Similarity of groups was estimated using GeneDoc (version 2.7) with the BLOSUM62 scoring matrix. *Red*, 90–100% similarity; *yellow*, 70–90% similarity; and *blue*, 50–70% similarity. Species and NCBI and Ensembl (in case of *X. tropicallis*) accession numbers are as follows: human (*Homo sapiens*, NP_065916), mouse (*Mus musculus*, NP_082412), chicken (*Gallus gallus*, XP_422834), frog (*Xenopus tropicalis*, ENSXETP00000016953), zebrafish (*Danio rerio*, CAM56669), fruit fly (*Drosophila melanogaster*, NP_610056), trematoda (*Schistosoma japonicum*, AAW25528), and choanoflagellate (*Monosiga brevicollis*, EDQ87802).

rithms, BindN+ and RISP (36, 37) (Fig. 5*A*). Moreover sequence alignment within the first 105 amino acids of RHAU proteins from choanoflagellate to humans revealed a novel highly conserved domain (54–66 aa) termed RSM next to the already known G-rich region (10–50 aa) with one RGG box motif (47–49 aa) (Fig. 5*B*). Although RGG box motifs have been reported to be involved in RNA binding (41, 42), RSM has no similarity to any known functional sequence elements. Thus, RHAU likely harbors an atypical conserved RNA-binding domain in the N terminus.

The N-terminal RNA-binding Domain Is Essential and Sufficient for RNA Interaction and Localization of RHAU in SGs—To further test whether these 105 N-terminal residues are sufficient for RNA binding and subsequent RHAU recruitment to SGs, we designed FLAG-tagged truncated forms of the RHAU protein: WT-FLAG, (50–1008)-FLAG, (105–1008)-FLAG, (1:105)-FLAG and (1:130)-FLAG. For the RNA binding experiment, CLIP analyses were performed using a monoclonal mouse anti-FLAG M2 antibody coupled to agarose beads. Western blotting confirmed the successful expression and immunoprecipitation of all fragments with (1–105)-FLAG and (1–130)-FLAG at higher protein levels (Fig. 6*A*, *left panel*). To analyze the RNA binding activities of these proteins, the measured radiolabeled RNA levels were normalized over protein levels. As shown in Fig. 6*A* (*right panel*), compared with WT-FLAG (*lane 2*), the amounts of RNA associated with (50–1008)-FLAG (*lane 3*) and (105–1008)-FLAG (*lane 4*) were stepwise reduced to ~70 and 30%, respectively. The high levels of RNA binding (before normalization) with (1–105)-FLAG and (1–130)-FLAG (*lanes 5* and *6*) may partly reflect their expression levels (Fig. 6*A*, *right panel*); however, they still showed ~10-fold higher RNA binding activity than WT-FLAG after normalization. Thus, these results demonstrate that the first 105 amino acids are essential and sufficient for RHAU to bind RNA.

Because the N-terminal fragment of RHAU (1–105)-FLAG has been shown to bind RNA *in vivo*, we again performed a double filter RNA binding assay using the recombinant GST-Nter fragment (1–200 aa). Purified, recombinant GST and GST-Nter (supplemental Fig. 2*B*) were incubated with 5′-end-

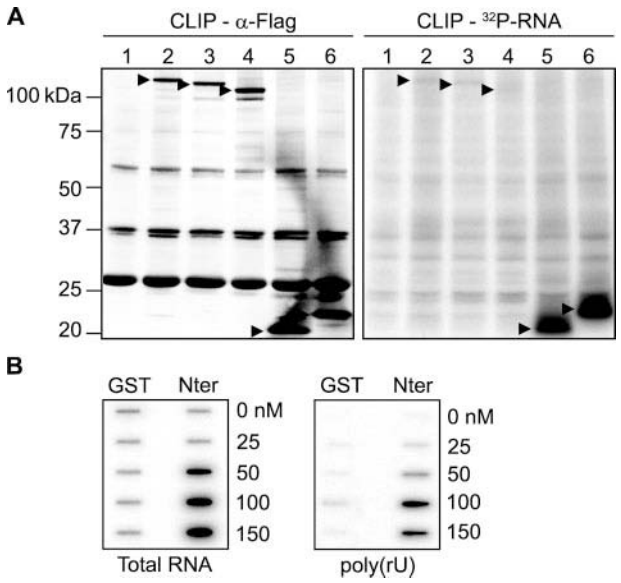


FIGURE 6. **The N-terminal RNA-binding domain is essential and sufficient for RNA interaction *in vivo* and *in vitro*.** *A*, CLIP method to show RNA binding activity of the N-terminal 105 amino acid residues of RHAU. HeLa cells were transfected with an empty vector (*lane 1*), vectors expressing FLAG-tagged RHAU wild type (*lane 2*), or its fragments, (50–1008) (*lane 3*), (105–1008) (*lane 4*), (1–105) (*lane 5*), and (1–130) (*lane 6*). Cells were UV irradiated 24 h later, harvested, and immunoprecipitated by FLAG antibody. Bound RNA was labeled with [γ-^{32}P]ATP and detected by a phosphorimaging system as described under "Experimental Procedures." *Left panel*, Western blot of immunoprecipitated FLAG-tagged RHAU and its fragments. *Right panel*, bound RNA. The amount of associated RNA was normalized to the expression level of proteins. ▶, FLAG-tagged RHAU and its fragments. *B*, the N terminus of RHAU binds to RNA *in vitro*. 0, 25, 50, 100, and 150 nM GST or GST-Nter (1–200 aa) was incubated with 10,000 cpm 5′-end-labeled total RNA (*left panel*) or poly(rU) (*right panel*) for 30 min at 37 °C and filtered through nitrocellulose and nylon membranes. Note that the dose-dependent RNA binding, both total RNA and poly(rU), was detected only with GST-Nter.

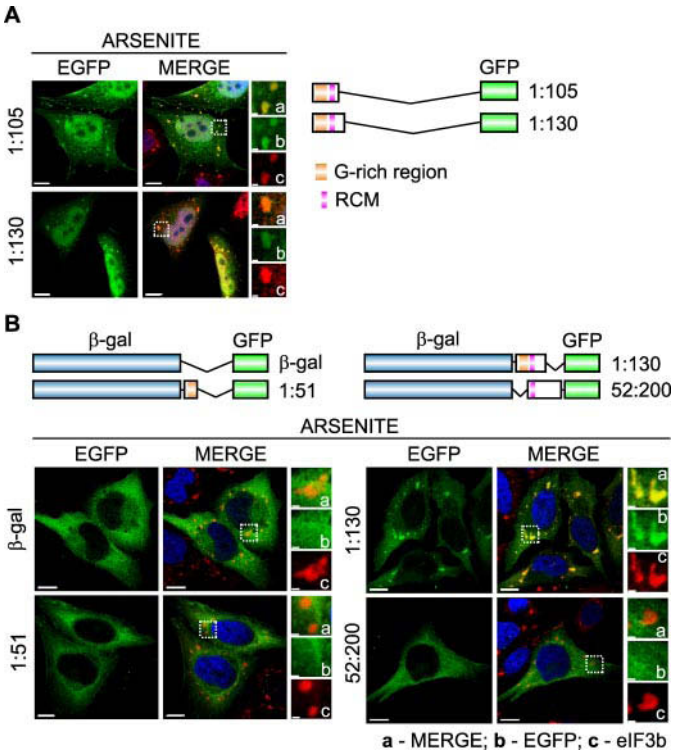


FIGURE 7. **The N-terminal RNA-binding domain is also essential and sufficient for SG localization of RHAU.** *A*, immunofluorescent images of HeLa cells transfected with vectors expressing EGFP-tagged N-terminal fragments of RHAU, (1–105) and (1–130). After 48 h of transfection, cells were treated with arsenite (0.5 mM, 45 min) to induce SGs. The merge shows the co-localization of EGFP-tagged RHAU fragments (*green*) and TIA-1 (*red*). DAPI (*blue*) stains nuclei. *B*, immunofluorescent images of HeLa cells expressing EGFP-β-galactosidase double tagged RHAU N-terminal fragments. Cells were treated as above to induce SGs. The co-localization of EGFP (*green*) and eIF3b (*red*) is shown in the merge together with DAPI (*blue*). Note that only the (1–130) fragment containing an intact RNA-binding domain recruits EGFP-β-galactosidase to SGs. *Enlargements* of *boxed regions* are shown in *small panels* indicating merge (*a*), EGFP-fused RHAU N-terminal fragment (*b*), and eIF3b (*c*). *Bar*, 10 μm (1 μm in *enlargements*). *GFP*, green fluorescent protein.

radiolabeled total RNA or with poly(rU) and then filtered through nitrocellulose and nylon membranes. As shown in Fig. 6*B*, the higher dose-dependent amount of RNA, both total RNA and poly(rU), was associated with GST-Nter rather than GST. The apparent affinities of GST-Nter for total RNA and poly(rU) were 45 and 80 nM, respectively. These data confirmed the *in vivo* data that the N terminus is essential and sufficient for RHAU association with RNA.

To determine whether the N terminus is also sufficient for RHAU recruitment to SG, we transfected HeLa cells with N-terminal mutants fused with the EGFP tag at their C termini as shown in Fig. 7*A*. Immunofluorescence analysis revealed that both tested mutants, (1–105)-EGFP and (1–130)-EGFP, co-localized with TIA-1 protein in SGs (Fig. 7*A*). Further we tested whether the N terminus can bring about SG localization of a protein that does not normally associate with SGs in response to stress. We fused different fragments of the RHAU N-terminal domain to β-gal-EGFP constructs (Fig. 7*B*) and transfected them into HeLa cells. In response to arsenite treatment, β-gal-EGFP by itself did not accumulate in SGs but was diffusely distributed in the cytoplasm as in normal conditions. However, when β-gal-EGFP was fused to the N-terminal 1–130 amino acids, including the complete putative RNA-binding domain with the G-rich region and the RSM, the fusion protein localized to SGs. In contrast, SG localization was not detected when β-gal-EGFP was fused with either the G-rich region (β-gal-EGFP(1–51 aa)) or the RSM (β-gal-EGFP(52–200 aa)) alone, indicating that only the complete RNA-binding domain is required for RHAU recruitment to SGs.

ATP Hydrolysis Plays a Role in RNA Binding and Kinetics of RHAU in SGs—Previously we showed that an ATPase-deficient RHAU mutant with a DEIH to DAIH point mutation is unable to accelerate uPA mRNA decay (26) and is excluded from the nucleus (30). This suggested that the ATPase activity of RHAU has profound effects on its biological activity and subcellular localization. Therefore, we examined the role of RHAU ATPase activity on its RNA binding activity and its localization in SGs. To elucidate the RNA binding potential of ATPase-deficient RHAU mutants, we performed CLIP analyses. After UV cross-linking and subsequent immunoprecipitation of EGFP-WT and EGFP-DAIH with an anti-RHAU antibody, the amount of associated RNA was normalized to corresponding protein levels. As shown in Fig. 8*A*, a higher amount of RNA was pulled down with the ATPase-deficient DAIH mutant than with EGFP-WT. Based on results of Tanaka and Schwer (34) where Prp22, a yeast DEAH box helicase, dissociated from the RNA upon ATP hydrolysis resulting in apparent RNA binding activity reduction, we concluded that ATPase activity of RHAU is

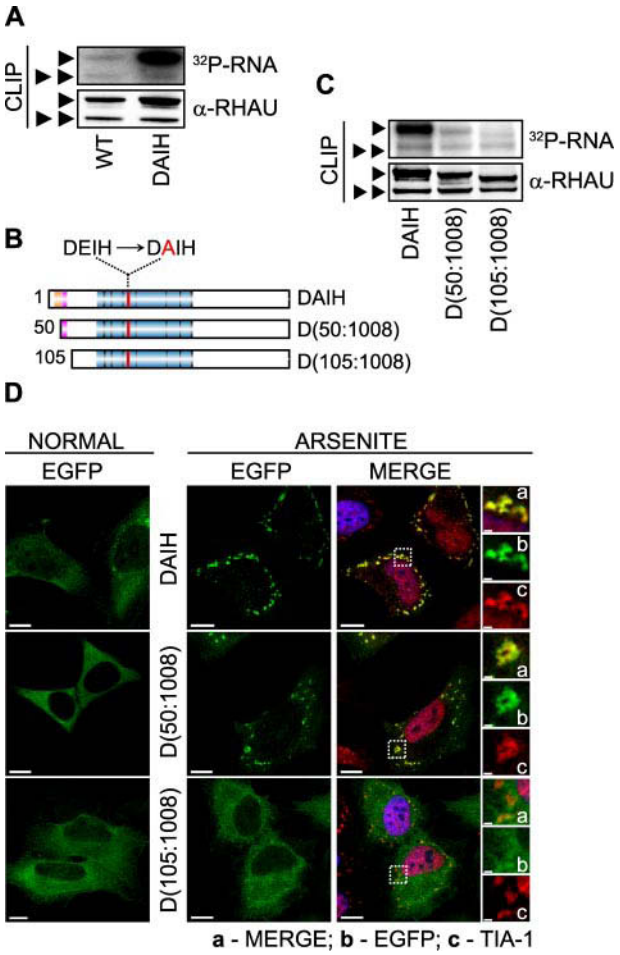


FIGURE 8. **Role of RHAU ATPase activity in RNA binding and retention of protein in SGs.** *A*, RNA binding activity of WT and the ATPase-deficient RHAU mutant DAIH. Levels of RHAU-associated RNA were measured by CLIP as described under "Experimental Procedures" 24 h after transfecting HeLa cells with EGFP-fused WT or DAIH. The level of RNA bound to the DAIH mutant was much higher than that bound to WT. Note that the level of RNA bound to endogenous RHAU is comparable in cells expressing exogenous WT or DAIH. *B*, schematic representation of N-terminal deletion, DAIH mutants: EGFP-DAIH, EGFP-DAIH(50–1008) (*D(50:1008)*), and EGFP-DAIH(105–1008) (*D(105:1008)*). *C*, comparison of RNA binding between mutants shown in *B*. HeLa cells were transfected with vectors expressing EGFP-DAIH, EGFP-DAIH(50–1008) (*D(50:1008)*), or EGFP-DAIH(105–1008) (*D(105:1008)*), and 24 h later the levels of RNA associated with these mutants and endogenous RHAU were assessed by CLIP as in *A*. Note that N-terminal deletion mutants show stepwise reduction of RNA binding compared with DAIH full length (such as WT and its deletion mutants in Fig. 4*E*). *D*, intracellular localization of EGFP-tagged DAIH mutants listed in *B* in control and arsenite-treated cells. Transfected HeLa cells were treated without (*NORMAL*) or with 0.5 mM sodium arsenite for 45 min. Images denoted *MERGE* are the merger of EGFP (*green*), TIA-1 (*red*), and DAPI (*blue*). *Enlargements* of *boxed regions* show merge (*a*), EGFP-tagged DAIH mutants (*b*), and TIA-1 (*c*). *Bar*, 10 μm (1 μm in *enlargements*). ▶, EGFP-tagged DAIH and its deletion mutants; ▶▶, endogenous RHAU.

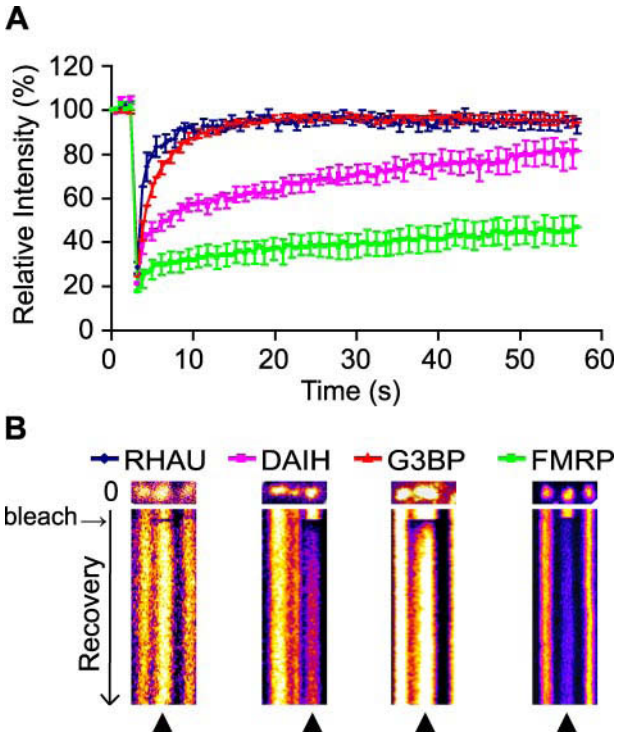


FIGURE 9. **ATP hydrolysis takes part in dynamic RHAU shuttling into and out of SGs.** *A*, fluorescent recovery patterns of RHAU, ATP-deficient RHAU mutant (DAIH), G3BP, and FMRP. HeLa cells transfected with EGFP-tagged RHAU, DAIH, G3BP, or FMRP were treated with arsenite and analyzed by the FRAP method. Bleaching of selected SGs was performed with three scan iterations for a total of 1.5 s. Fluorescence recovery was monitored at 0.5-s intervals for 50 s, and results were analyzed as described under "Experimental Procedures." Data were collected from at least 10 independent experiments, and each curve represents average fluorescence intensity ±S.E.M. (standard errors of means) over time. *B*, representative result of FRAP analysis of EGFP-tagged RHAU, DAIH, G3BP, or FMRP. SGs were photobleached, and subsequent fluorescence recoveries are shown in a fire mode in the time scale of recovery. ▲ indicates SGs bleaching.

involved in releasing RHAU from RNA rather than in binding to RNA.

Considering the possibility that the N terminus and ATP hydrolysis contribute to the RNA interaction of RHAU, we investigated whether the N-terminal RNA-binding domain is also required for RNA binding of the DAIH mutant and its accumulation in SGs. For this, we introduced a point mutation (DEIH to DAIH) into motif II of the helicase core region to generate DAIH(50–1008) and DAIH(105–1008) fused with EGFP at their N termini (Fig. 8*B*). First the truncated fragments were used in the CLIP analysis. While being aware of CLIP limitations in quantitative analysis, stepwise deletion of the N-terminal domain of the DAIH mutant to DAIH(50–1008) and DAIH(105–1008) nevertheless decreased the amount of associated RNA to ~50 and 15%, respectively, compared with the DAIH full-length mutant (Fig. 8*C*). The stepwise decrease in RNA interaction seen with the two DAIH deletion mutants corresponds to that seen with RHAU deletion mutants (see Fig. 4*E*). Thus, enhanced RNA binding of the DAIH mutant still required the N terminus.

Having shown that RHAU recruitment to SGs is dependent on its RNA interaction (see Figs. 3*C* and 4*E*), we tested whether the ATPase-deficient DAIH mutant accumulates in SGs more than wild type. Indeed the DAIH mutant localized in SGs (Fig. 8*D*) to a greater extent than the wild type, increasing the number of transfected cells with EGFP-positive SGs from 81 to 100%. Unlike DAIH full length and DAIH(50–1008), DAIH(105–100) was not detected in SGs. Therefore, RHAU recruitment to SGs is independent of its ATPase activity.

Having established that both RHAU wild type and the DAIH mutant accumulate in SGs, we used the FRAP method to examine whether the dynamics of shuttling into and out of SGs differs between the two proteins. Cells were transfected with EGFP-WT, EGFP-DAIH, EGFP-G3BP, or EGFP-FMRP, and

the FRAP analyses were performed 48 h after transfection. It has been shown that TIA-1 and G3BP recover rapidly and completely in SGs within 30 s after bleaching (6, 13). In contrast, PABP-1, another marker for mRNA shuttling, showed only 60% fluorescence recovery after 30 s (3, 6). In our study, we used EGFP-G3BP as a control for the rapid fluorescence recovery (Fig. 9, *A* and *B*). Although FMRP has never been tested by FRAP analysis, we included this protein as a control as it is known to localize in SGs (43). As shown in Fig. 9*A*, the fluorescence signal of G3BP recovered rapidly and completely within 10 s as reported previously (6), but the FMRP signal did not recover fully within 60 s, suggesting that FMRP is a scaffold SG component. In this analysis, the wild type and DAIH signals showed different recovery patterns. The former displayed rapid recovery (≥90%) of fluorescence after 8 s similar to G3BP with the same recovery after 11 s. However, the recovery of the EGFP-DAIH mutant fluorescent signal was much slower with only 60% of fluorescence recovery after 30 s of bleaching similar to PABP-1. The recovery rate reflects the strength of binding whereby slower recoveries correspond to tighter binding. Thus, in comparison with the wild type, the ATPase-deficient mutant, which cannot release RNA because of its inability to hydrolyze ATP, is retained longer in SGs, resulting in slower recovery kinetics. Combining these results, we conclude that the ATPase activity of RHAU is required for its active shuttling into and out of SGs and may also be important for remodeling and/or recruitment of RNPs in SGs.

DISCUSSION

In the present study, we have shown that the DEAH box helicase RHAU is a novel SG-associated protein that is recruited to SGs via RNA. Its N terminus is necessary and sufficient for RHAU localization in SGs and simultaneously for its binding to RNA. Bioinformatics analysis of the RHAU peptide sequence corroborated the experimental data, revealing that the N terminus of RHAU harbors a unique RNA-binding domain consisting of two abutting motifs: the G-rich region and the RSM. We have also shown that ATPase activity is involved in RHAU association with RNA and markedly influences the kinetics of RHAU recovery in SGs after photobleaching.

The list of proteins identified as components of SGs, co-localizing with known SG-associated proteins such as TIA-1, is still expanding. However, a method for isolating SGs to a significantly pure level has not yet been established, biochemical analysis of SGs is very difficult, and thus there is no consensus on their detailed structure and molecular composition. Furthermore it has not been systematically examined whether SGs induced by different stresses are qualitatively different. The presence of qualitatively different SGs is suggested by the fact that the recruitment of TTP to SGs is induced by carbonyl cyanide *m*-fluorophenylhydrazone but not by arsenite, whereas TIA-1 is recruited to SGs by both stimuli (39). Our results clearly demonstrate that RHAU is one of the SG-associated proteins and that its recruitment to SGs does not depend on the type of stress. Thus, RHAU appears to be a general SG component like TIA-1 or eIF3b. Although RHAU is recruited to SGs, it does not participate actively in SG formation as RNA interference-mediated down-regulation of RHAU had no effect on arsenite-induced SG formation (supplemental data). Similarly a further SG-associated protein, ZBP1, did not induce or inhibit SG formation when overexpressed or knocked down, respectively (7). Although ZBP1 does not modulate SG formation, its significance for SGs is attributed to its function in mRNA turnover regulation. Interestingly the stability of ZBP1 target mRNAs is not regulated by their recruitment to SGs but by their retention in SGs to prevent their rapid translocation and subsequent decay in processing bodies or by the exosome (7). Knowing that RHAU is a *cis*-acting factor involved in ARE-mediated decay of uPA mRNA and that its association with SG is regulated by RNA interaction, we speculate that RHAU acts in SGs in a manner similar to ZBP1.

Given that several RNA-binding proteins such as FMRP, TIA-1, G3BP, ZBP1, or Caprin-1 are mostly targeted to SGs via mRNA (7, 41, 43, 44), we performed CLIP, nitrocellulose filter binding assays, and bioinformatics analyses to elucidate whether the N terminus of RHAU, which is necessary and sufficient for RHAU accumulation in SGs, physically interacts with RNA. The analyses revealed that RHAU does associate with RNA both *in vitro* and *in vivo* and that only the N-terminal 105 amino acids harbor significant RNA binding activity. In addition, this region of RHAU contains the computationally predicted RNA-binding domain (3–75 aa) composed of a unique, previously unidentified RNA-binding motif. The domain is located in an unstructured and flexible sequence composed of two adjacent motifs, the G-rich region (10–50 aa) and the RSM (54–66 aa). The latter is highly conserved among RHAU orthologs, but it was not found in other proteins, suggesting that this motif confers a unique function on RHAU. Interestingly a potential RGG box motif (47–49 aa), which has been reported previously to be involved in the RNA interaction of several proteins, is present in the G-rich region of RHAU. For example, two RGG box sequences identified at the C terminus of the FMRP protein were shown to be necessary for an RNA and G-quadruplex interaction of FMRP and also for the induction of SG formation by EGFP-FMRP overexpression (42, 43). Moreover Caprin-1, an SG-associated protein, contains three RGG box motifs in its C terminus that are also important for its specific RNA interaction and SG induction (41). Our analyses of the RHAU N terminus shows that the deletion of the G-rich region together with the RGG box motif reduces RNA binding activity of RHAU and its accumulation in SGs (42%) suggesting that the RGG box motif may be involved in the RNA interaction and RHAU localization in SGs. However, the G-rich region with the RGG box motif alone was not sufficient for the recruitment of a fused β-galactosidase reporter protein to SGs (see Fig. 7*B*). Only the complete RNA-binding domain enables RHAU localization in SGs and RNA interaction, indicating that a domain other than the G-rich region takes part in RHAU recruitment to SGs. The detailed nature of the cooperation and the underlying mechanism between the G-rich region and the RSM remain interesting questions for the future.

According to the classification of putative RNA helicases by Abdelhaleem *et al.* (20), RHAU belongs to the DEAH box family containing a highly conserved C terminus next to the conserved helicase core region. Studies of DEAH proteins in yeast have

shown that the specific function of each member is determined mainly by its N terminus and that helicase and ATPase activities are essential for their function during splicing. For example, Prp16, the ortholog of mammalian DHX38, localizes in the nucleus and recognizes the spliceosome through its N terminus (25). Its C terminus enhances spliceosome binding, whereas the helicase core region drives the process by binding and hydrolyzing ATP. Similarly Prp22 (DHX8), another splicing factor, requires the N-terminal domain for its ATPase and helicase function (24, 45). Our results show that the RHAU helicase follows the same principle because its specific localization in SGs and RNA interaction is governed by its N-terminal domain where the RNA-binding domain is located.

Importantly CLIP analysis confirmed that RHAU is an RNA-binding protein, which was suggested by our previous work when RHAU was first identified as an ARE-associated protein (26) but was not formally tested. Although we do not have direct evidence that RHAU-associated RNAs are mRNAs, the correlation between RHAU RNA binding activity and RHAU accumulation in SGs, which are aggregates of protein-mRNA complexes, strongly suggests this to be the case. Furthermore results showing that the abrogation of RHAU ATPase activity increases the amount of associated RNA (see Fig. 8*A*) are in good agreement with the accepted model of the closed and open conformation of DE*X*(H/D) box helicases. The structural and fluorescence resonance energy transfer information of complete DEAD box helicases has shown that the cooperative binding of ATP and RNA induces a closed conformation of helicases that is relaxed upon ATP hydrolysis, leading to the release of bound RNA (46–49). A similar effect has been reported for the Prp22 splicing factor that releases bound RNA upon ATP hydrolysis, thereby reducing its apparent RNA binding affinity (34). In this context, it is noteworthy that the ATPase-deficient mutant is excluded from the nucleus despite its strong RNA interaction indicating that it may associate mostly with mRNA and that the presence of RHAU in the nucleus is not essential for its localization to SGs. The nuclear exclusion of the ATPase-deficient mutant DAIH is not a consequence of it being trapped in the cytoplasm by a strong, irreversible interaction with RNA as we proposed previously (30). The N terminus-deleted ATPase-deficient mutant DAIH(105–1008), which showed markedly reduced RNA interaction and localization in SGs, remained mainly in the cytoplasm (Fig. 8*D*). Therefore, the nuclear-cytoplasmic translocation of RHAU is most probably regulated by its ATPase activity. A similar role for ATPase activity in nuclear export via the CRM1 protein was reported previously for the DEAD box helicase An3 (27).

Many SG-associated proteins are RNA-binding proteins involved in different intracellular processes mediated by or acting upon mRNA. Several of these proteins, including TIA-1, TTP, heterogeneous nuclear RNP A1, PCBP-2, MLN51, and G3BP (6, 9, 11–13), have been shown by FRAP analysis to rapidly shuttle into and out of SGs. This suggests that SGs are not static storage centers for untranslated mRNA but rather dynamic structures that sort individual transcripts for storage, reinitiation, or decay (4, 10). FRAP analysis also revealed that proteins behave with differing kinetics. For instance the PABP-1 protein exhibits a slower and only partial recovery (60%) 30 s after bleaching compared with TIA-1 (6, 11, 13). Like G3BP, wild-type RHAU shows very rapid shuttling into and out of SGs. In striking contrast, the kinetics of the ATPase-deficient mutant were different from RHAU and G3BP: green fluorescent protein signal recovery in SGs was as slow as for PABP-1. Thus, ATP hydrolysis affects the shuttling kinetics of RHAU into SGs. Furthermore our analysis showed that the FMRP protein is not a member of the group of rapidly shuttling RNA-binding proteins. Rather it exhibits slow shuttling like the Fas-activated serine/threonine kinase protein and, thus, may play a scaffolding role in SGs (6). Although RHAU exhibits the rapid mobility of G3BP, TIA-1, and TTP, it may be involved in RNP remodeling or their recruitment to SGs as suggested for TIA-1 and TTP (13).

With the detection of RHAU in SGs, we are the first to show the association of a DEAH box RNA helicase with SGs. Our FRAP analysis shows the striking difference in SG shuttling kinetics between fully active RHAU protein and its ATPase-deficient mutant, tendering the hypothesis that ATPase activity takes part in the dynamic remodeling of RNPs in SGs. Furthermore the sorting and remodeling of RNPs in SGs presumably requires energy, which may be provided by RNA helicases that utilize energy from ATP hydrolysis to rearrange inter- or intramolecular RNA structures or to dissociate protein-RNA complexes. To our knowledge, three DEAD box RNA helicases, rck/p54 (DDX6), DDX3, and eIF4A, have been shown to localize in SGs (32, 50, 51). Considering that various remodeling processes take place in these foci, it will be no surprise if the number of RNA helicases found associated with SGs is revealed to be much higher.

Acknowledgments—We thank J. Tanaka (University of the Ryukyus, Nishihara, Japan), D. Schmitz Rohmer, and B. A. Hemmings (Friedrich Miescher Institute, Basel, Switzerland) for kindly providing valuable materials and P. King, C. Du Roure (Friedrich Miescher Institute), P. Svoboda (Institute of Molecular Genetics AS CR, Prague, Czech Republic), and G. Stoecklin (German Cancer Research Center, Heidelberg, Germany) for critical comments on the manuscript. We also thank S. Thiry and S. Pauli (Friedrich Miescher Institute) and M. Yazdani and L. Jaskiewicz (Biozentrum, Basel, Switzerland) for excellent technical assistance and advice. This work was supported in part by CREST grant (to Y. Fujiki) from the Science and Technology Agency of Japan.

REFERENCES

1. Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007) *Nat. Rev. Mol. Cell Biol.* **8,** 113–126
2. Anderson, P., and Kedersha, N. (2002) *J. Cell Sci.* **115,** 3227–3234
3. Anderson, P., and Kedersha, N. (2002) *Cell Stress Chaperones* **7,** 213–221
4. Anderson, P., and Kedersha, N. (2006) *J. Cell Biol.* **172,** 803–808
5. Kedersha, N., Chen, S., Gilks, N., Li, W., Miller, I. J., Stahl, J., and Anderson, P. (2002) *Mol. Biol. Cell* **13,** 195–210
6. Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M. J., Scheuner, D., Kaufman, R. J., Golan, D. E., and Anderson, P. (2005) *J. Cell Biol.* **169,** 871–884
7. Stohr, N., Lederer, M., Reinke, C., Meyer, S., Hatzfeld, M., Singer, R. H., and Huttelmaier, S. (2006) *J. Cell Biol.* **175,** 527–534
8. Kedersha, N. L., Gupta, M., Li, W., Miller, I., and Anderson, P. (1999) *J. Cell Biol.* **147,** 1431–1442
9. Guil, S., Long, J. C., and Caceres, J. F. (2006) *Mol. Cell. Biol.* **26,** 5744–5758

10. Anderson, P., and Kedersha, N. (2008) *Trends Biochem. Sci.* **33,** 141–150
11. Baguet, A., Degot, S., Cougot, N., Bertrand, E., Chenard, M. P., Wendling, C., Kessler, P., Le Hir, H., Rio, M. C., and Tomasetto, C. (2007) *J. Cell Sci.* **120,** 2774–2784
12. Fujimura, K., Kano, F., and Murata, M. (2008) *RNA* **14,** 425–431
13. Kedersha, N., Cho, M. R., Li, W., Yacono, P. W., Chen, S., Gilks, N., Golan, D. E., and Anderson, P. (2000) *J. Cell Biol.* **151,** 1257–1268
14. Fujimura, K., Kano, F., and Murata, M. (2008) *Exp. Cell Res.* **314,** 543–553
15. Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002) *Nat. Rev. Mol. Cell Biol.* **3,** 195–205
16. Linder, P. (2006) *Nucleic Acids Res.* **34,** 4168–4180
17. Jankowsky, E., and Jankowsky, A. (2000) *Nucleic Acids Res.* **28,** 333–334
18. Jankowsky, E., Gross, C. H., Shuman, S., and Pyle, A. M. (2001) *Science* **291,** 121–125
19. Bleichert, F., and Baserga, S. J. (2007) *Mol. Cell* **27,** 339–352
20. Abdelhaleem, M., Maltais, L., and Wain, H. (2003) *Genomics* **81,** 618–622
21. Jankowsky, E., and Bowers, H. (2006) *Nucleic Acids Res.* **34,** 4181–4188
22. Mohr, G., Del Campo, M., Mohr, S., Yang, Q., Jia, H., Jankowsky, E., and Lambowitz, A. M. (2008) *J. Mol. Biol.* **375,** 1344–1364
23. Valgardsdottir, R., and Prydz, H. (2003) *J. Biol. Chem.* **278,** 21146–21154
24. Schneider, S., and Schwer, B. (2001) *J. Biol. Chem.* **276,** 21184–21191
25. Wang, Y., and Guthrie, C. (1998) *RNA* **4,** 1216–1229
26. Tran, H., Schilling, M., Wirbelauer, C., Hess, D., and Nagamine, Y. (2004) *Mol. Cell* **13,** 101–111
27. Askjaer, P., Rosendahl, R., and Kjems, J. (2000) *J. Biol. Chem.* **275,** 11561–11568
28. Schmitt, C., von Kobbe, C., Bachi, A., Pante, N., Rodrigues, J. P., Boscheron, C., Rigaut, G., Wilm, M., Seraphin, B., Carmo-Fonseca, M., and Izaurralde, E. (1999) *EMBO J.* **18,** 4332–4347
29. Wagner, J. D., Jankowsky, E., Company, M., Pyle, A. M., and Abelson, J. N. (1998) *EMBO J.* **17,** 2926–2937
30. Iwamoto, F., Stadler, M., Chalupnikova, K., Oakeley, E., and Nagamine, Y. (2008) *Exp. Cell Res.* **314,** 1378–1391
31. Vaughn, J. P., Creacy, S. D., Routh, E. D., Joyner-Butt, C., Jenkins, G. S., Pauli, S., Nagamine, Y., and Akman, S. A. (2005) *J. Biol. Chem.* **280,** 38117–38120
32. Mazroui, R., Sukarieh, R., Bordeleau, M. E., Kaufman, R. J., Northcote, P., Tanaka, J., Gallouzi, I., and Pelletier, J. (2006) *Mol. Biol. Cell* **17,** 4212–4219
33. Ule, J., Jensen, K., Mele, A., and Darnell, R. B. (2005) *Methods* **37,** 376–386
34. Tanaka, N., and Schwer, B. (2005) *Biochemistry* **44,** 9795–9803
35. Terribilini, M., Sander, J. D., Lee, J. H., Zaback, P., Jernigan, R. L., Honavar, V., and Dobbs, D. (2007) *Nucleic Acids Res.* **35,** W578–W584
36. Tong, J., Jiang, P., and Lu, Z. H. (2008) *Comput. Methods Programs Biomed.* **90,** 148–153
37. Jeong, E., Chung, I. F., and Miyano, S. (2004) *Genome Inform.* **15,** 105–116
38. Do, C. B., Mahabhashyam, M. S., Brudno, M., and Batzoglou, S. (2005) *Genome Res.* **15,** 330–340
39. Stoecklin, G., Stubbs, T., Kedersha, N., Wax, S., Rigby, W. F., Blackwell, T. K., and Anderson, P. (2004) *EMBO J.* **23,** 1313–1324
40. Gallouzi, I. E., Brennan, C. M., Stenberg, M. G., Swanson, M. S., Eversole, A., Maizels, N., and Steitz, J. A. (2000) *Proc. Natl. Acad. Sci. U. S. A* **97,** 3073–3078
41. Solomon, S., Xu, Y., Wang, B., David, M. D., Schubert, P., Kennedy, D., and Schrader, J. W. (2007) *Mol. Cell. Biol.* **27,** 2324–2342
42. Ramos, A., Hollingworth, D., and Pastore, A. (2003) *RNA* **9,** 1198–1207
43. Mazroui, R., Huot, M. E., Tremblay, S., Filion, C., Labelle, Y., and Khandjian, E. W. (2002) *Hum. Mol. Genet.* **11,** 3007–3017
44. Tourriere, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J. M., Bertrand, E., and Tazi, J. (2003) *J. Cell Biol.* **160,** 823–831
45. Schneider, S., Hotz, H. R., and Schwer, B. (2002) *J. Biol. Chem.* **277,** 15452–15458
46. Theissen, B., Karow, A. R., Kohler, J., Gubaev, A., and Klostermeier, D. (2008) *Proc. Natl. Acad. Sci. U. S. A* **105,** 548–553
47. Caruthers, J. M., Johnson, E. R., and McKay, D. B. (2000) *Proc. Natl. Acad. Sci. U. S. A* **97,** 13080–13085
48. Cheng, Z., Coller, J., Parker, R., and Song, H. (2005) *RNA* **11,** 1258–1270
49. Shi, H., Cordin, O., Minder, C. M., Linder, P., and Xu, R. M. (2004) *Proc. Natl. Acad. Sci. U. S. A* **101,** 17628–17633
50. Wilczynska, A., Aigueperse, C., Kress, M., Dautry, F., and Weil, D. (2005) *J. Cell Sci.* **118,** 981–992
51. Lai, M. C., Lee, Y. H., and Tarn, W. Y. (2008) *Mol. Biol. Cell* **19,** 3847–3858

Published online 14 May 2010 Nucleic Acids Research, 2010, Vol. 38, No. 18 6219–6233
doi:10.1093/nar/gkq372

Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU

Simon Lattmann[1], Banabihari Giri[2], James P. Vaughn[2], Steven A. Akman[2] and Yoshikuni Nagamine[1,*]

[1]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, 4058 Basel, Switzerland and [2]Department of Cancer Biology and the Comprehensive Cancer Center, Wake Forest University School of Medicine, Winston-Salem, North Carolina, NC 27157, USA

Received January 25, 2010; Revised April 12, 2010; Accepted April 27, 2010

ABSTRACT

Under physiological conditions, guanine-rich sequences of DNA and RNA can adopt stable and atypical four-stranded helical structures called G-quadruplexes (G4). Such G4 structures have been shown to occur *in vivo* and to play a role in various processes such as transcription, translation and telomere maintenance. Owing to their high-thermodynamic stability, resolution of G4 structures *in vivo* requires specialized enzymes. RHAU is a human RNA helicase of the DEAH-box family that exhibits a unique ATP-dependent G4-resolvase activity with a high affinity and specificity for its substrate *in vitro*. How RHAU recognizes G4-RNAs has not yet been established. Here, we show that the amino-terminal region of RHAU is essential for RHAU to bind G4 structures and further identify within this region the evolutionary conserved RSM (RHAU-specific motif) domain as a major affinity and specificity determinant. G4-resolvase activity and strict RSM dependency are also observed with CG9323, the *Drosophila* orthologue of RHAU, in the amino terminal region of which the RSM is the only conserved motif. Thus, these results reveal a novel motif in RHAU protein that plays an important role in recognizing and resolving G4-RNA structures, properties unique to RHAU among many known RNA helicases.

INTRODUCTION

In cells, the strong inclination of RNA for mis-folding or adopting non-functional conformations is overcome by the presence of RNA chaperones that facilitate conformational transitions of RNA. Among these chaperones are the RNA helicases, which couple NTP hydrolysis with structural and functional rearrangement of the RNA. The DEAD/H-box proteins constitute a widely spread subgroup of RNA helicases that have been identified in all forms of life, including viruses. DEAD/H-box proteins have been shown to catalyse the disruption of RNA–RNA interactions (1,2), to remodel ribonucleoprotein (RNP) complexes (3,4) and to assist the correct folding of RNA (5,6). In this regard, DEAD/H-box proteins are essential cellular components that take part in many, if not all, aspects of RNA metabolism, ranging from transcription to RNA decay [for review see (7–9)].

Structurally, DEAD/H-box proteins consist of a highly conserved catalytic core composed of two RecA-like domains that couples NTP hydrolysis with the helicase activity. The helicase core domain is often flanked by ancillary N- and C-terminal regions of variable length and sequence. While the core domain of RNA helicases has been extensively studied, much less is known about the biological role of the N- and C-terminal regions. Because of the high degree of amino acid sequence conservation within the helicase core of DEAD/H-box proteins, this region may not contribute directly to the substrate specificity of the enzyme. In contrast to the helicase core, the N- and C-terminal flanking regions are usually unique, with the exception of certain identifiable sequence features. These regions have been shown to provide substantial substrate specificity through their interaction with RNAs or with protein partners that modulate the activity and/or the specificity of the helicase (1,10).

Genetic studies in yeast have demonstrated that DEAD/H-box proteins perform highly specific tasks *in vivo*. In most cases, they are required at a specific stage of RNA

*To whom correspondence should be addressed. Tel: +41 61 697 6669; Fax: +41 61 697 3976; Email: Yoshikuni.Nagamine@fmi.ch

6220 Nucleic Acids Research, 2010, Vol. 38, No. 18

processing and most of them are highly specific for their substrates. As revealed by the lethality of null mutants in yeast, most DEAD/H-box proteins are essential, suggesting tight target specificity for each protein (10,11). However, with the exception of DbpA and related proteins (12,13) and to a lesser extent Prp5 (14,15), DEAD/H-box proteins show little or even no RNA specificity when analysed *in vitro* (8). This apparent non-discrimination of target RNA by *in vitro* analysis may be due to the absence of essential co-factors that would direct the helicase to its physiological RNA substrate or, more likely, due to the use of biologically non-relevant RNA substrates. It is, therefore, a prerequisite to identify the naturally occurring substrates of RNA helicases in order to characterize them in an *in vitro* context.

Unlike most of the RNA helicases that have been investigated biochemically, the human DEAH-box protein RHAU (alias DHX36 or G4R1) exhibits a unique ATP-dependent guanine-quadruplex (G4) resolvase activity with a high affinity and specificity for its substrate *in vitro* (16,17). G4–nucleic acid structures result from the propensity of guanine-rich sequences of DNA and RNA to form atypical and thermodynamically stable four-stranded helical structures under physiological conditions [for review see (18,19)]. Formation of G4 structures *in vivo* is related to impairment of cellular DNA replication, transcription or translation initiation (20). G4 structures have also been shown to play a role in immunoglobulin gene rearrangement, promoter activation and telomere maintenance (19). Owing to their high-thermodynamic stability, resolution of G4 structures *in vivo* requires specialized enzymes (21). RHAU binds G4-RNA with sub-nanomolar affinity (16) and unwinds G4 structures much more efficiently than double-stranded nucleic acid [(17) and Tran, H., unpublished data]. Consistent with these biochemical observations, RHAU was also identified as the major source of tetramolecular RNA-resolving activity in HeLa cell lysates (16). Despite these advances, we still lack a corresponding understanding of the mechanism by which RHAU recognizes its substrate.

Structurally, RHAU consists of a ~400-amino acid helicase core comprising all signature motifs of the DEAH-box family of helicases (Figure 1A). The core region is flanked by N- and C-terminal regions of ~200 and ~400 amino acid, respectively. Previous work showed that RHAU associates with mRNAs and re-localizes to stress granules (SGs) upon translational arrest induced by various environmental stresses (22). Deletion analysis of the N-terminal region of RHAU revealed that a region of the first 105 amino acid was critical for RNA binding and re-localization of RHAU to SGs. Importantly, this 105 amino acid-long region alone had the ability *in vivo* to bind RNA and re-localize to SGs. The apparent significance of the first 105 amino acid for RHAU interaction with RNA prompted us to determine whether the N-terminal domain of RHAU also contributes to the recognition of G4 structures *in vitro*. To address this question, a series of N-terminal deletion mutants of RHAU was generated as shown in Figure 1A. Through biochemical analysis of these mutants, we have uncovered the functional importance of the first 105 amino acids of RHAU for interaction with G4 structures and further revealed among this region that the previously identified RSM (RHAU-specific motif, amino acids 54–66) domain (22) is essential to the high affinity for G4 structures shown by RHAU. Here, we show that the N-terminal region of RHAU alone can bind to G4-RNA structures, albeit with lower affinity than the full-length protein. We also show that the G4-resolving activity of RHAU is conserved in higher eukaryotes, insofar as CG9323, the *Drosophila* orthologue of RHAU, readily unwinds G4 structures in the presence of ATP. Finally, we show that a variant form of CG9323 harbouring mutations in the RSM domain manifests reduced G4-binding activity, suggesting that RSM functions similarly in *Drosophila* and human RHAU proteins.

MATERIALS AND METHODS

Plasmid constructs, cloning and mutagenesis

The plasmid used for the expression of GST-RHAU(1–200) protein in bacteria and the baculoviral expression vector used for the expression of GST-RHAU have been described previously (22,23). Plasmids for the expression of C-terminal FLAG-tagged recombinant RHAU proteins: RHAU-FLAG, RHAU(ΔGly)-FLAG, RHAU(ΔGly–RSM)-FLAG, RHAU(DAIH)-FLAG and RHAU(1–105)-FLAG have been described previously (22). CG9323 cDNA was obtained from the Drosophila Genomics Resource Center (Bloomington, IN, USA) and was sub-cloned by PCR amplification into pIRES.EGFP-FLAG-N1 (22). For the RHAU(ΔRSM)-FLAG plasmid, the RSM (amino acid 54–66) excision was performed using a standard overlapping PCR method (24,25) The resulting PCR product was sub-cloned into pIRES.EGFP-FLAG-N1. For RHAU(RSM-mut*x*)-FLAG and CG9323(RSM-mut), the RSM-coding sequence was mutagenized using a variation of the classical QuickChange (Stratagene) site-directed mutagenesis PCR method (26). Construction of all these plasmids was confirmed by sequencing. Sequences of oligonucleotides used in this work and detailed descriptions of the plasmid constructions are available upon request.

Cell culture

Human embryonic kidney HEK293T cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS) and 2 mM L-glutamine at 37°C in a humidified 5% CO_2 incubator. *Spodoptera frugiperda Sf9* cells were maintained in Grace's insect cells medium (Invitrogen) supplemented with 10% FCS, 4.11 mM L-glutamine, 3.33 g l^{-1} lactalbumin hydrolysate and 3.33 g l^{-1} yeastolate at 27°C.

Expression and purification of recombinant RHAU proteins

GST-RHAU(1–200) protein was expressed in *Escherichia coli* strain BL21-CodonPlus (DE3)-RIPL (Stratagene).

Nucleic Acids Research, 2010, Vol. 38, No. 18 6221

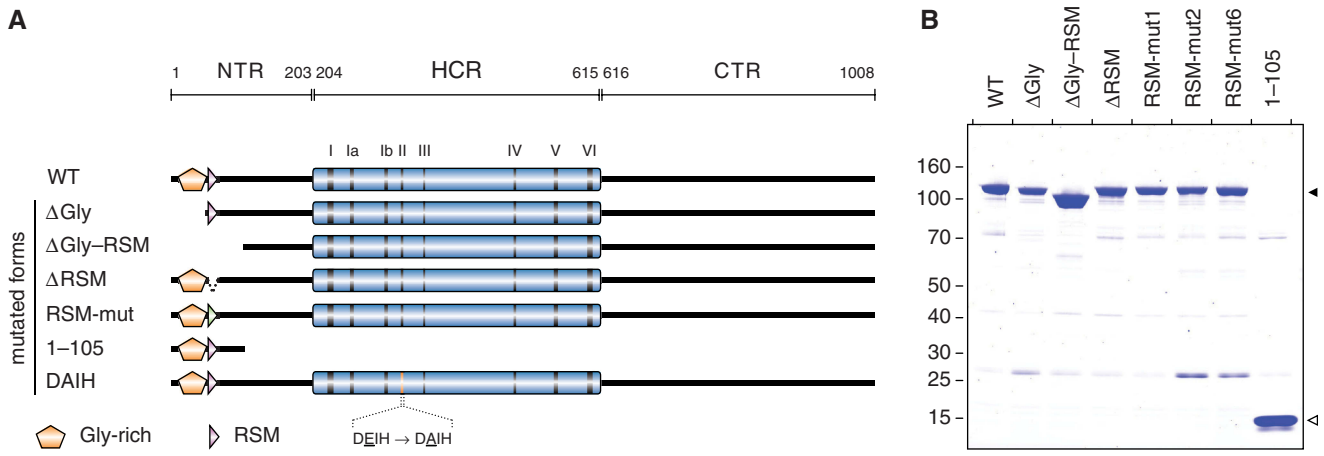


Figure 1. N-terminal deletion mutants of RHAU. (**A**) Schematic representation of the 1008 amino acid RHAU protein and its N-terminal truncated mutants. The conserved ATPase/helicase motifs I–VI of the DEAD/H-box family are indicated within the helicase core region (HCR) by vertical bars. The HCR is flanked by N-terminal (NTR) and C-terminal (CTR) regions of 203 and 393 amino acid, respectively. The N-terminal Gly-rich (amino acid 10–51) and RSM (RHAU-specific motif, amino acids 54–66) domains are indicated. WT, amino acids 1–1008; ΔGly, amino acids 50–1008; ΔGly–RSM, amino acids 105–1008; ΔRSM, RHAU harbouring a deletion of amino acids 54–66; RSM-mut, RHAU with mutagenized RSM (for details about amino acid substitutions, see Figure 6A); 1–105, amino acids 1–105; DAIH, D335A ATPase deficient mutant. (**B**) SDS–PAGE separation and Coomassie staining of purified FLAG-tagged recombinant wild-type (WT) and mutant RHAU proteins (2 µg protein per lane). The positions and sizes (kDa) of marker proteins are indicated at the left. The filled and open arrows on the right indicate the positions of the purified proteins.

Cultures were inoculated from a single colony of freshly transformed cells and maintained in logarithmic growth at 37°C in 2× YT medium supplemented with ampicillin (0.1 mg ml^{-1}) to a final volume of 1 l. When the OD_{600} reached 0.3, the temperature was adjusted to 25°C and cells were cultured to an OD_{600} of 0.6. Isopropyl β-D-thiogalactopyranoside was added to 0.4 mM and the cultures were incubated for 12 h at 25°C with constant shaking. Cells were harvested by centrifugation and the pellets stored at −80°C. All subsequent operations were performed at 4°C. The cell pellets were resuspended in 50 ml of lysis buffer [1× PBS supplemented with NaCl to a final concentration of 300 mM, 1% Triton X-100, 5 mM EDTA, 5 mM DTT, 1× protease inhibitor cocktail (Complete EDTA-free, Roche)]. Lysozyme was added to 1 mg ml^{-1} and the suspensions were lysed for 30 min, followed by sonication (3× 10 s) to reduce viscosity. Insoluble material was removed by centrifugation (39 000*g*, 30 min, 4°C) in a Beckman JA-17 rotor and the resulting supernatant was filtered through a 0.22-µm Express PLUS Membrane (Millipore). The resulting filtrate was applied to an 1-ml GSTrap 4B column (GE Healthcare). The column was washed with 30 ml washing buffer (1× PBS supplemented with NaCl to a final concentration of 300 mM, 5 mM DTT) and the recombinant protein was recovered in elution buffer [50 mM Tris–HCl (pH 8.0 at 4°C), 10 mM reduced glutathione, 5 mM DTT)]. The fraction containing the recombinant protein were pooled and elution buffer was exchanged to storage buffer [20 mM HEPES-KOH (pH 7.7 at room temperature), 50 mM KCl, 0.01% Nonidet P-40, 0.5 mM EDTA, 5 mM DTT, 10% glycerol, 2 mM AEBSF] by repeated concentration and dilution steps in Amicon Ultra-15 centrifugal filters (Millipore). Recombinant proteins were stored at −80°C. Purity of protein preparations was assessed by SDS–PAGE and protein concentrations were determined photometrically at 280 nm using the calculated extinction coefficient $\varepsilon = 71\,905\ M^{-1}\,cm^{-1}$.

GST-RHAU protein was expressed in *Sf9* cells according to the supplier's instructions (PharMingen). Three days post-baculoviral infection, cells were harvested by centrifugation and the pellets were stored at −80°C. All subsequent operations were performed at 4°C. The cell pellets were resuspended in insect cell lysis buffer [10 mM Tris–HCl (pH 7.5 at 4°C), 10 mM sodium phosphate, 300 mM NaCl, 1% Triton X-100, 10 mM sodium pyrophosphate, 10% glycerol, 5 mM EDTA, 5 mM DTT, 1× protease inhibitor cocktail (Complete EDTA-free, Roche)] and lysed for 30 min. All subsequent purification steps were carried out as described above for the purification of GST-RHAU(1–200) protein. Purity of protein preparations was assessed by SDS–PAGE and protein concentrations were determined photometrically at 280 nm using the calculated extinction coefficient $\varepsilon = 166\,615\ M^{-1}\,cm^{-1}$.

All FLAG-tagged recombinant RHAU and CG9323 proteins [RHAU-FLAG, RHAU(ΔGly)-FLAG, RHAU(ΔGly–RSM)-FLAG, RHAU(ΔRSM)-FLAG, RHAU(RSM-mut*x*)-FLAG, RHAU(DAIH)-FLAG, RHAU(1–105)-FLAG, CG9323-FLAG and CG9323(RSM-mut2)-FLAG] were transiently expressed in HEK293T. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 24–36 h post-transfection, washed with ice-cold PBS and resuspended in lysis buffer [1× PBS supplemented with NaCl to a final concentration of 600 mM, 1% Nonidet P-40, 2 mM EDTA, 2 mM AEBSF (4-(2-aminoethyl)-benzenesulphonyl fluoride hydrochloride), 1× protease inhibitor cocktail (Complete EDTA-free, Roche) for 30 min. Cell lysates were sonicated (1× 10 s) to reduce the viscosity and insoluble material removed by centrifugation

6222 Nucleic Acids Research, 2010, Vol. 38, No. 18

(39 000*g*, 20 min, 4°C) in a Beckman JA-17 rotor. The soluble lysates were mixed for 5 h with 200 µl of a 50% slurry of anti-FLAG M2-agarose affinity gel (Sigma) that had been equilibrated in lysis buffer. The resin was recovered by centrifugation, washed 3× with 1 ml of lysis buffer follow by three washes with 1 ml of IP-washing buffer [50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 0.1% Nonidet P-40, 5 mM EDTA]. The bound proteins were eluted with 600 µl elution solution [0.1 mg ml^{-1} FLAG peptide, 10 mM Tris–HCl (pH 7.5), 150 mM NaCl] for 10 min at 37°C and stored at −20°C. Purity of protein preparations was assessed by SDS–PAGE and protein concentrations were determined by Bradford assay with BSA as the standard.

Tetraplex G4-RNA preparation

Unlabelled and 5′-TAMRA-labelled 35-mer oligoribonucleotides 5′-A_{15}–G_5–A_{15}-3′ were used to form tetramolecular G4-RNA and are referred to hereafter as 'rAGA'. rAGA were purchased from Dharmacon Research and were dissolved in RNase-free solution [100 mM KCl, 10 mM Tris–HCl (pH 7.5), 1 mM EDTA] to a final concentration of 500 µM. To form tetramolecular quadruplex by annealing of rAGA, the solution was aliquoted into PCR tubes and incubated in a PCR thermocycler at 98°C for 10 min and then held at 80°C. EDTA was added immediately to a final concentration of 25 mM and the solution was allowed to cool slowly to room temperature. rAGA aliquots were pooled together and stored at 4°C for 2–3 days. By this procedure, the conversion of monomeric rAGA to a stable tetramolecular quadruplex form was almost complete as judged by native PAGE (Supplementary Figure S1A and C). Circular dichroism (CD) analysis of the purified, annealed rAGA oligomers revealed a typical spectrum of a parallel G4 structure with positive and negative peaks at 263 and 245 nm, respectively (Supplementary Figure S1B). The above prepared tetramolecular G4-rAGA were stored at −20°C.

CD spectropolarimetry

CD experiments were performed with an AVIV Model 202 spectrophotometer equipped with a thermoelectrically controlled cell holder. G4-rAGA at a concentration of 1 µM were prepared in RNase-free solution [10 mM Tris–HCl (pH 7.5), 1 mM EDTA] supplemented with 50 mM KCl, NaCl or LiCl. Quartz cells with 1 cm path length were used for all experiments. CD spectra were recorded at 25°C in the UV region (200–350 nm) with 1 nm increments and an averaging time of 2 s.

Thermodynamic analysis of the stability of tetramolecular G4-rAGA structures

Preformed 5′-TAMRA-labelled tetramolecular G4-rAGA at a concentration of 100 nM were prepared in RNase-free solution [10 mM Tris–HCl (pH 7.5), 1 mM EDTA) supplemented with 50 mM KCl, NaCl or LiCl. Tetramolecular G4-rAGA structures were incubated for 5 min at various temperatures ranging from 20°C to 99°C and were immediately subjected to separation by electrophoresis for 3 h on a pre-electrophoresed 10% polyacrylamide native gel (19:1 acrylamide:*bis* ratio) in 0.5× TBE at 25°C. After electrophoresis, gels were scanned on a Typhoon 9210 Imager (GE Healthcare) and analysed with Multi Gauge software (Fuji). The fraction of undenatured G4-rAGA was quantitated as the ratio of the signal from the tetramolecular form to the sum of the tetramolecular and the denatured ssRNA. The apparent temperature of mid-transition (T_m) was determined by representing the fraction of undenatured G4-rAGA as a function of the temperature. Reported T_m values are representative of three independent experiments.

Electromobility shift assay and apparent K_d determination

Recombinant RHAU proteins at concentrations from 1 to 1000 nM were incubated with 100 pM 5′-^{32}P-labelled G4-RNA in K-Res buffer [50 mM Tris–acetate (pH 7.8), 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10% glycerol], supplemented with 10 mM EDTA and 0.2 U µl^{-1} SUPERase-In (Ambion) in a 15-µl reaction. The reactions were incubated at 37°C for 30 min. RNA–protein complexes were resolved on a pre-electrophoresed 6% polyacrylamide native gel (37.5:1 acrylamide:*bis* ratio) in 0.5× TBE at 4°C for 90 min. After electrophoresis, gels were fixed for 1 h in 10% isopropanol/7% acetic acid. RNA–protein complexes were detected by Phosphor-Imaging, scanned on a Typhoon 9400 Imager (GE Healthcare) and analysed with ImageQuant TL software (Nonlinear Dynamics).

G4-RNA resolvase assay

Recombinant RHAU proteins at concentrations from 1 to 100 nM were incubated with 4 nM 5′-^{32}P-labelled G4-RNA in K-Res buffer supplemented with 1 mM ATP and 0.2 U µl^{-1} SUPERase-In in a 15-µl reaction. Reactions were allowed to proceed at 30°C for 30 min, stopped by transfer to ice and addition of 1:10 volume of 10× loading buffer [33 mM Tris–HCl (pH 8.0), 25% (w/v) Ficoll-400, 110 mM EDTA, 0.17% SDS]. Reaction products were resolved on a pre-electrophoresed 10% polyacrylamide native gel (19:1 acrylamide:*bis* ratio) in 0.5× TBE at 4°C for 90 min. After electrophoresis, gels were fixed for 1 h in 10% isopropanol/7% acetic acid and exposed to a Phosphor-Imaging screen.

ATPase assay

Recombinant RHAU proteins at concentrations from 25 to 200 nM were incubated with 1 µl [γ-^{32}P]ATP (3000 Ci mmol^{-1}, 0.4 mCi ml^{-1}) in ATPase assay buffer [50 mM Tris–HCl (pH 8.0), 100 mM KCl, 3 mM $MgCl_2$, 1 mM ATP, 1 mM DTT] supplemented with 0.4 U µl^{-1} RNasin (Promega) and 1 µg µl^{-1} homopolymeric poly(U) RNA (Sigma) in a 20-µl reaction. Reactions were allowed to proceed at 37°C for 15 min. The reactions were stopped by addition of 1 ml of a 5% (w/v) suspension of activated charcoal (Sigma) in 20 mM phosphoric acid. The samples were incubated on ice for 10 min and the charcoal containing the adsorbed unhydrolysed ATP was pelleted by centrifugation (21 000*g*, 15 min, 4°C). The supernatants

Nucleic Acids Research, 2010, Vol. 38, No. 18 6223

(containing free γ-$^{32}P_i$) were transferred to new tubes and the radioactivity was quantified by liquid scintillation counting (Cerenkov counts).

RESULTS

The first 105 amino acids of RHAU are required for binding and resolving G4 structures

We demonstrated previously *in vivo* and *in vitro* that amino acids 1–105 of RHAU delineate a functional domain of prime importance for the interaction of RHAU with RNA (22). To assess the significance of this region for the G4-resolvase activity of RHAU, we constructed a series of FLAG-tagged N-terminal truncated mutants of RHAU (Figure 1A). These truncated forms were expressed in HEK293T cells and purified from soluble lysates by anti-FLAG immunoaffinity chromatography. The purity of these preparations as judged by Coomassie staining after SDS–PAGE was very similar (Figure 1B). To test whether the truncated form of RHAU lacking the first 105 amino acids [RHAU(ΔGly–RSM)] could resolve G4 structures, RHAU protein was incubated with ^{32}P-labelled tetramolecular rAGA and ATP. The products were analysed by native PAGE after disruption of RNA–protein interactions by addition of SDS. The free single-stranded ^{32}P-labelled oligos migrated faster than the quadruplex substrate. As shown previously, wild-type (WT) RHAU efficiently resolved G4 structures into single-stranded oligos (Figure 2A). The labelled product of the resolvase reaction co-migrated during electrophoresis with the single-stranded species released by thermal denaturation of the substrate and was proportional to the level of input protein. In contrast to RHAU(WT), RHAU(ΔGly–RSM) failed to resolve the G4 substrate, suggesting that the N-terminal region of RHAU is essential for its G4-resolvase activity.

Being critical for the enzymatic activity and containing an atypical RNA-binding domain (22), the N-terminal region of RHAU is most likely to play a role in substrate recognition. Therefore, we examined whether this N-terminal region was also essential for the recognition of G4 structures. To address this question, we performed RNA electromobility shift assays (REMSA) using G4 substrate as the ligand. RHAU protein was incubated with ^{32}P-labelled G4 in the absence of ATP (to prevent G4 resolution) and the mixtures were analysed by native PAGE. In the absence of protein, the ^{32}P-labelled G4 structures migrated as a single species in the gel (Figure 2B). Addition of increasing amounts of RHAU(WT) protein resulted in the appearance of a protein–G4 complex with reduced mobility. The bound complexes appeared as two main band regions in the gel that might reflect multiple forms of the complexes. In contrast, the RHAU(ΔGly–RSM) mutant protein failed to form a stable complex with G4 structures. Taken together, these results demonstrate that the N-terminal region encompassing residues 1–105 of RHAU protein is indispensable for both binding and the resolving of G4 structures by RHAU. These results are in agreement with our previous report that the N-terminal region of RHAU is critical for its RNA binding *in vivo* and its relocalization to SGs (22).

The N-terminal region of RHAU binds but cannot alone resolve G4 structures

In view of their uniqueness, the N-terminal regions of DEAH-box proteins may have regulatory functions such

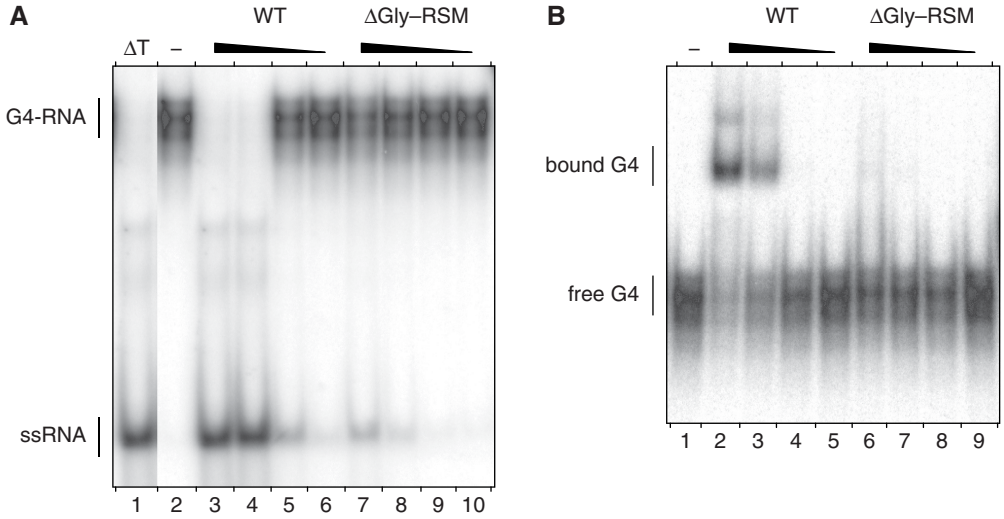


Figure 2. G4-RNA unwinding and binding by N-terminal truncated RHAU proteins. (**A**) G4-RNA unwinding assay: radio-labelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (1, 3, 10 and 30 nM) of WT RHAU or ΔGly–RSM mutant. The reaction products were resolved by native PAGE after disrupting RNA–protein interactions with SDS. An autoradiogram of the gel is shown. The positions of the tetraplex substrate RNA (G4-RNA) and the unwound single-stranded product (ssRNA) are denoted on the left. An aliquot of the tetraplex substrate that was heat-denatured (95°C, 5 min) and then quenched (ΔT) serves as a marker for the position of ssRNA. (**B**) G4-RNA binding assay: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (1, 3, 10 and 30 nM) of WT RHAU or ΔGly–RSM mutant in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. The positions of the free tetramolecular RNA substrate and the protein–RNA complex are indicated on the left.

6224 Nucleic Acids Research, 2010, Vol. 38, No. 18

as sub-cellular localization and/or interaction with RNAs or other proteins (1,10). With *prima facie* evidence that the N-terminal region of RHAU is important for both binding to and resolving G4 structures, we examined whether this region has G4-binding activity. As shown by REMSA, increasing amounts of RHAU(1–105) protein formed proportionately slow-migrating complexes with G4-RNA (Figure 3A). That the observed mobility shift was due to binding of RHAU(1–105) protein was confirmed by addition of anti-FLAG antibody to binding reactions. As shown in Figure 3B, anti-FLAG antibody caused a super shift of the protein–probe complex but not of the free probe.

Given that some G4-binding proteins destabilize G4 structures in a non-catalytic fashion without requiring ATP hydrolysis (27), we examined whether the N-terminal region of RHAU may also destabilize G4 structures in a similar manner. However, as shown in Figure 3C, binding of the N-terminal region to G4 structures is not sufficient to destabilize it. In agreement with the observation that RHAU cannot resolve G4 structures in the absence of ATP (17), this result demonstrates that the N-terminal region alone lacks G4-destabilizing activity and that the G4-resolving activity is likely a property of the central catalytically active core domain.

The helicase core domain, together with the N-terminal region, contributes to tight G4 binding of RHAU

Considering that RHAU has a high affinity for G4 structures (16), we next examined whether the N-terminal region itself accounts for the high affinity for G4 structures of the whole protein. N-terminal region [RHAU(1–200)] and full-length RHAU(WT) were expressed as GST fusion proteins in bacteria and *Sf9* insect cells, respectively, and purified to homogeneity as shown in Supplementary Figure S2. Their apparent dissociation constants (K_d) for G4 structures were determined by REMSA. Titration of RHAU(WT) and RHAU(1–200) proteins gave half-saturation points for G4 binding of 14 nM and 440 nM, respectively (Figure 4A and B). This 30-fold difference between the two proteins indicates that the N-terminal region does not constitute by itself an independent and high-affinity G4-RNA binding domain. We propose that the helicase core domain provides RHAU with substantial additional binding activity through interactions with the RNA phosphate backbone, as already shown for many DEAD/H-box proteins (28–34). It should be mentioned that the observed K_d value obtained here is higher than that previously reported (16), which may reflect the difference of the type and location of tags attached to RHAU (N-GST versus C-His_6) and of the way how proteins were purified.

The RSM domain, but not the Gly-rich sequence, in the N-terminal region is crucial for the recognition and resolution of G4 structures by RHAU

The functional N-terminal region encompassing residues 1–105 of RHAU protein consists of two abutting domains (Figure 1A): a 41 amino acid-long low-complexity Gly-rich domain followed by the evolutionary conserved

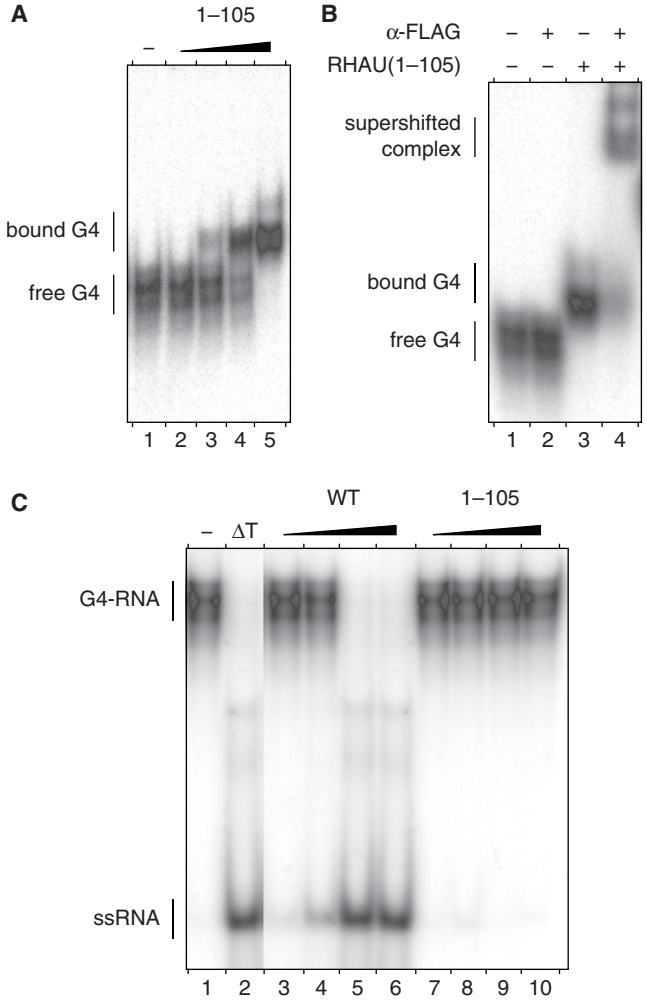


Figure 3. G4-RNA binding and resolving activities of the N-terminal domain of RHAU. (**A**) Gel mobility shift assay for G4-RNA binding: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (30, 100, 300 and 1000 nM) of RHAU(1–105) in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**B**) Super shift of G4-RNA-binding complex by a specific antibody: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated with RHAU(1–105) at 300 nM in the presence or absence of anti-FLAG antibodies (2.5 μg per lane), as indicated. The reaction mixtures were electrophoresed on a native gel. An autoradiogram of the gel is shown. The positions of the free tetramolecular RNA substrate, the protein–RNA complex and the super-shifted complex are indicated on the left. (**C**) G4-RNA unwinding assay: radio-labelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (1, 3, 10 and 30 nM) of WT RHAU or ΔGly–RSM mutant (30, 100, 300 and 1000 nM). The reaction products were resolved by native PAGE after disrupting of RNA–protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the tetraplex substrate that was heat-denatured (95°C, 5 min) and then quenched (ΔT) serves as a marker for the position of ssRNA.

13 amino acid-long RSM (amino acids 54–66). We demonstrated previously *in vivo* that deletion of the Gly-rich region [RHAU(ΔGly)] significantly impinged on the relocalization of the protein to SGs and on its association with RNA (22). To characterize the individual contributions of each of the two domains to the

Nucleic Acids Research, 2010, Vol. 38, No. 18 6225

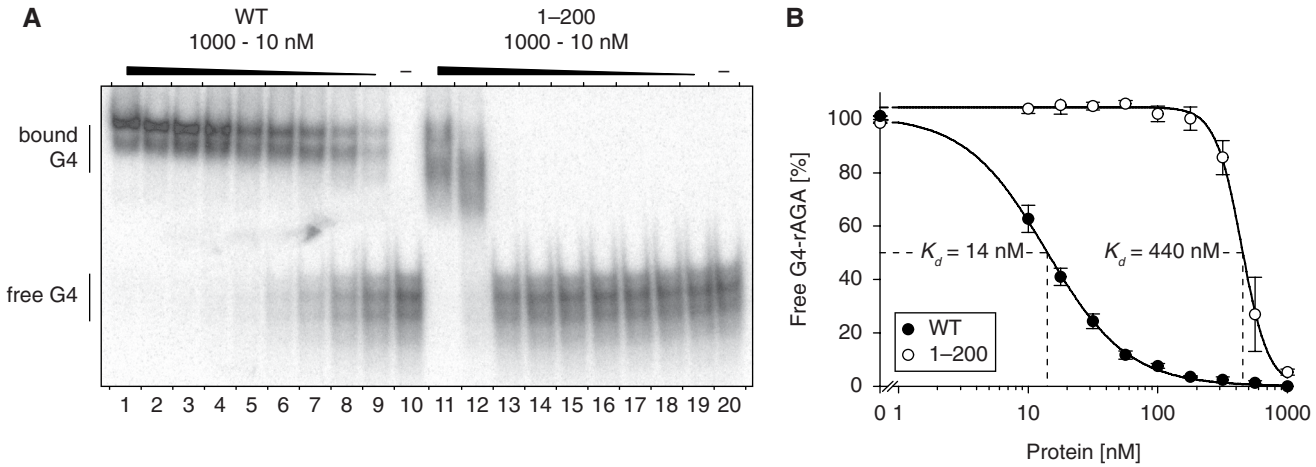


Figure 4. G4-RNA-binding properties of RHAU and the N-terminal region. (**A**) Gel mobility shift assay for G4-RNA binding: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts of either GST-tagged RHAU(WT) or RHAU(1–200) in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. At concentrations from 10 to 1000 nM, GST protein alone had no effect on G4-RNA mobility (data not shown). (**B**) Quantification of gel mobility shift assays for WT RHAU (WT, filled circle) and RHAU N-terminal region (1–200, open circle) binding to tetramolecular rAGA. The data represent means ± SEM from three independent experiments.

interaction of RHAU with G4 structures, we performed REMSA experiments using corresponding deletion mutants (Figure 5A). As already shown in Figure 2B, RHAU(ΔGly–RSM) lacking the first 105 amino acids failed to form a stable complex with G4 structures. In contrast, RHAU(ΔGly) still harbouring the RSM was able to bind to G4 structures with the same efficiency as RHAU(WT). Contrary to the deletion of the Gly-rich sequence, a RHAU mutant with an RSM deletion [RHAU(ΔRSM)] abrogated the interaction with G4 structures. We further assessed the consequence of loss of G4 recognition on the *in vitro* resolvase activity of RHAU (Figure 5B). As expected, RHAU(ΔRSM), like RHAU(ΔGly–RSM), failed to resolve G4 structures, while RHAU(ΔGly), like RHAU(WT), did. Similar results with strict RSM-dependency could also be reproduced with other G4-RNA structures (data not shown). Together, these results indicate that the presence of the RSM but not further sequences of the N-terminal region is a prerequisite for the recognition and the resolution of G4 structures by RHAU *in vitro*.

Conserved residues within the RSM domain are essential for the recognition of G4 structures by RHAU

Multiple-sequence alignments of RHAU orthologues of various species from choanoflagelates to humans unveiled a highly conserved cluster of 13 amino acids presenting the RSM domain embedded in a moderately conserved region of about 50 amino acids (amino acids 54–100) (Figure 6A). In contrast, the rest of the N-terminus surrounding this region is poorly conserved. The RSM consensus sequence as determined from 40 different RHAU orthologue sequences is: P–π–x–L–ζ–G–[+]–ζ–I–G–Ψ–Ω–Ω [consensus RSM is listed according to the Seefeld convention (37), symbol nomenclature stands for: π (pi) = small side chain, ζ (zeta) = hydrophilic, [+] = basic, Ψ (Psi) = aliphatic, Ω (Omega) = aromatic] (Supplementary Figure S3). The motif consists of five invariant amino acids (Pro-54, Leu-57, Gly-59, Ile-62 and Gly-63) and seven highly conserved residues of similar biochemical properties (π-55, ζ-58, [+]-60, ζ-61, Ψ-64, Ω-65, Ω-66). In addition, computation-based secondary structure prediction of the sequences used for alignment suggested that RSM is partially structured, sitting astride an unstructured loop (amino acids 54–59) and an α-helix (amino acids 60–66) (Figure 6A; Supplementary Figure S4). We further examined by site-specific mutagenesis *in vitro* the contribution of conserved amino acids within the RSM domain to the recognition and resolution of G4 structures (Figure 6A). To this end, conserved residues of the RSM were substituted with alanines, prolines or glycines. We focused on Pro-54, Gly-59 and Gly-63 because they may provide the RSM with substantial structural rigidity (Pro-54) or flexibility (Gly-59 and Gly-63). As shown in Figure 6B, mutation of the five invariant residues of the RSM (RSM-mut2) considerably reduced the binding affinity of RHAU for G4-RNA, albeit to a lesser extent than the RSM-deleted form of RHAU (ΔRSM). Interestingly, substitution of the invariant Gly-59 and Gly-63 residues with prolines (RSM-mut6) caused a stronger reduction of G4-binding activity than the RHAU(RSM-mut2) in which these two amino acids are mutated to alanines. Since prolines may induce a structural constraint on the RSM, this result suggests that recognition of G4 structures may depend on the conformational organization of the RSM. Finally, as observed with RHAU(RSM-mut1) and RHAU(RSM-mut3) mutants, Pro-54 as well as the polar residues Lys-58, Arg-60 and Glu-61 appear to be dispensable.

As expected, RHAU proteins with mutated RSM and displaying strong G4-RNA binding deficiency (RSM-mut2 and 6) also had reduced G4-resolving activity (Figure 6C). In contrast, despite the slight reduction of G4-binding activity observed for

6226 Nucleic Acids Research, 2010, Vol. 38, No. 18

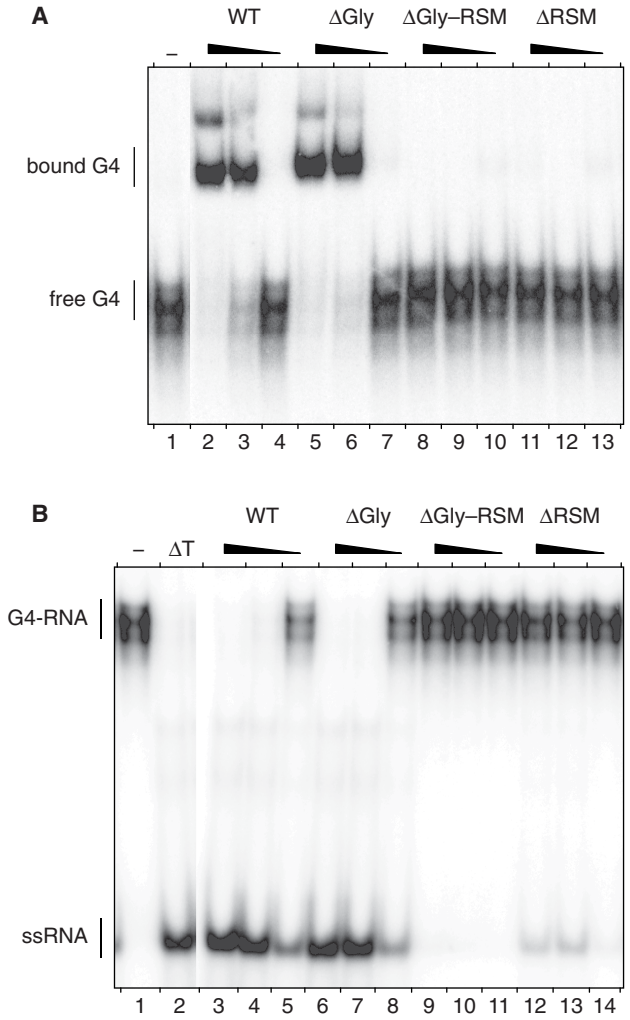


Figure 5. G4-RNA binding and unwinding by Gly-rich and RSM-truncated RHAU proteins. (**A**) Gel mobility shift assay for G4-RNA binding: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU, or ΔGly, ΔGly–RSM or ΔRSM RHAU mutants in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**B**) G4-RNA unwinding assay: radio-labelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU, or ΔGly, ΔGly–RSM or ΔRSM RHAU mutants. The reaction products were resolved by native PAGE after disrupting of RNA–protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the tetraplex substrate that was heat denatured (95°C, 5 min) and then quenched (ΔT) serves as a marker for the position of ssRNA.

RHAU(RSM-mut1) and RHAU(RSM-mut3), these two mutants unwound G4 structures with an efficiency comparable to WT RHAU. We concluded from these *in vitro* mutagenesis experiments that the highly conserved residues in the RSM domain are essential for RHAU G4-structure binding and resolving activities.

ATPase activity of RHAU N-terminal truncated mutants

Hitherto, it could not be excluded that the introduced mutations caused significant conformational changes in the helicase core domain that resulted in enzymatically defective proteins. To define whether loss of G4-resolvase activity was due to loss of G4 binding by the N-terminal domain or to impairment of the basic activity of the helicase core domain, we compared the ATPase activities of RHAU mutants and WT RHAU. As shown previously (23), RHAU exhibited significant ATPase activity in the absence of nucleic acids, which was stimulated substantially by the presence of homopolymeric poly(U) (Figure 7A). Both the RNA-dependent and -independent ATPase activities were proportional to the amount of input RHAU protein. To exclude the possibility that the RNA-independent ATPase activity observed was due to contaminants derived from HEK293T cells, we substituted the Glu-335 residue with Ala within the Walker B site (DEIH → DAIH), which abolishes RHAU ATPase activity (23). RHAU(DAIH) protein was prepared according to the protocol employed for WT RHAU protein, with comparable yield and purity (data not shown). In contrast to WT RHAU, RHAU(DAIH) showed no ATPase activity, even in the presence of poly(U) (Figure 7B). Thus, we concluded that RHAU harbours intrinsic RNA-independent ATPase activity, which can be further stimulated by homopolymeric RNA.

To clarify the biochemical basis of the absence of G4-resolvase activity in the RHAU(ΔRSM) mutant, we measured the rate of ATP hydrolysis by this mutant and compared it to that of the WT RHAU and the RHAU(ΔGly) mutant that was still proficient in G4 unwinding. In the absence of RNA co-factor, the ATPase activity of RHAU(ΔRSM) protein was the same as RHAU(WT) and RHAU(ΔGly) proteins (Figure 7C), suggesting that the deletion of the Gly-rich (ΔGly) or the RSM (ΔRSM) domains does not affect the basal ATPase activity of RHAU. However, the extent of poly(U)-dependent stimulation of ATPase activity of RHAU(ΔRSM) was ~60% of that of RHAU(WT) and similar to that of RHAU(ΔGly). Taken together, these results indicate that deletion of the N-terminal region of RHAU does not cause any significant conformational changes to the helicase core, but renders the RHAU protein less responsive to RNA in the stimulation of its ATPase activity. Given that RHAU(ΔRSM) retained significant RNA-dependent ATPase activity, equivalent to the G4-resolvase proficient RHAU(ΔGly) mutant, we concluded that the lack of G4-resolving activity in the RHAU(ΔRSM) mutant resulted from the loss of G4 binding by the N-terminal domain.

CG9323, the *Drosophila* ortholog of RHAU efficiently unwinds G4-RNA

Based on sequence analysis, RHAU has clear orthologs in almost all species of the animal kingdom ranging from choanoflagellates to humans. A multiple sequence alignment between eight of these orthologs showed the helicase core (amino acids 204–615) together with the C-terminal region (amino acids 616–1008) of RHAU to be evolutionary conserved (Figure 8A). In contrast, sequences of the N-terminal region (amino acid 1–203) show little

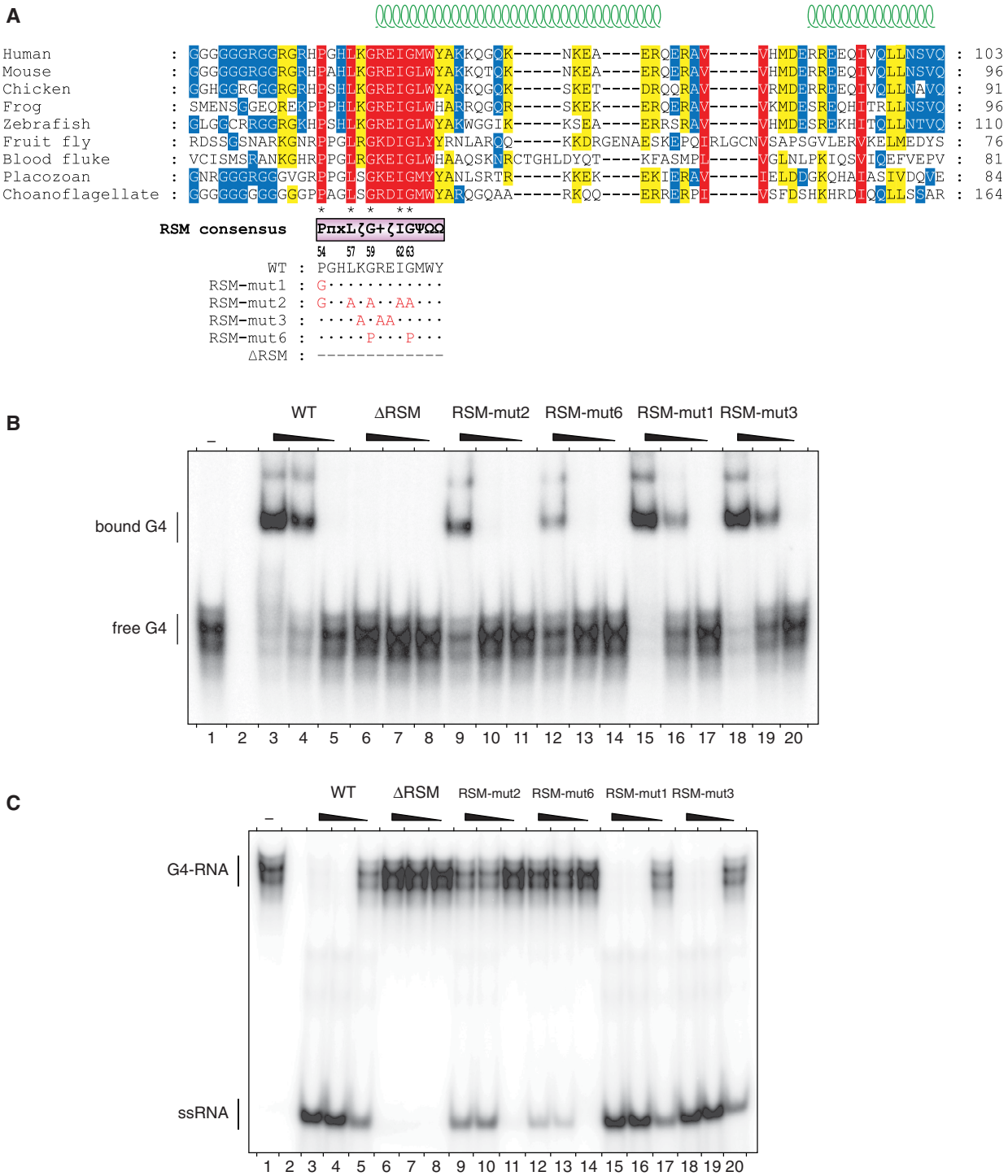


Figure 6. G4-RNA binding and unwinding by RSM-mutated forms of RHAU. (**A**) Conservation of the RSM among RHAU orthologues throughout evolution. Multiple sequence alignment was carried out with MAFFT (version 6) (35). Similarity analysis was made by GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Similarity is shown in red for 100%, yellow for 99–80% and blue for 79–60%. Amino acids that are identical between the nine sequences are indicated by asterisks below the alignment. Secondary structure prediction was performed by JPRED (36) and is indicated on the top. The RSM and its consensus sequence derived from 40 different RHAU orthologues are shown below the alignment [π (pi) = small side chain, ζ (zeta) = hydrophilic, [+] = basic, Ψ (Psi) = aliphatic, Ω (Omega) = aromatic]. The site-directed substitutions of the RSM mutants employed in this study are listed with the amino acid changes indicated underneath. Species and accession numbers of RHAU orthologues listed are: human (*Homo sapiens*, NP_065916), mouse (*Mus musculus*, NP_082412), chicken (*Gallus gallus*, XP_422834), frog (*Xenopus tropicalis*, ENSXETP00000016958), zebrafish (*Danio rerio*, NP_001122016), fruit fly (*Drosophila melanogaster*, NP_610056), blood fluke (*Schistosoma mansoni*, XP_002577014), placozoan (*Trichoplax adhaerens*, XP_002110272), choanoflagellate (*Monosiga brevicollis*, XP_001747335). (**B**) Gel mobility shift assay for G4-RNA binding: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU or the indicated RSM mutants in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**C**) G4-RNA unwinding assay: radio-labelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP without protein (−) or with increasing amounts (6, 20 and 60 nM) of either WT RHAU or the indicated RSM mutants. The reaction products were resolved by native PAGE after disrupting RNA–protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the duplex substrate that was heat denatured (95°C, 5 min) and then quenched (ΔT) serves as a marker for the position of ssRNA.

6228 Nucleic Acids Research, 2010, Vol. 38, No. 18

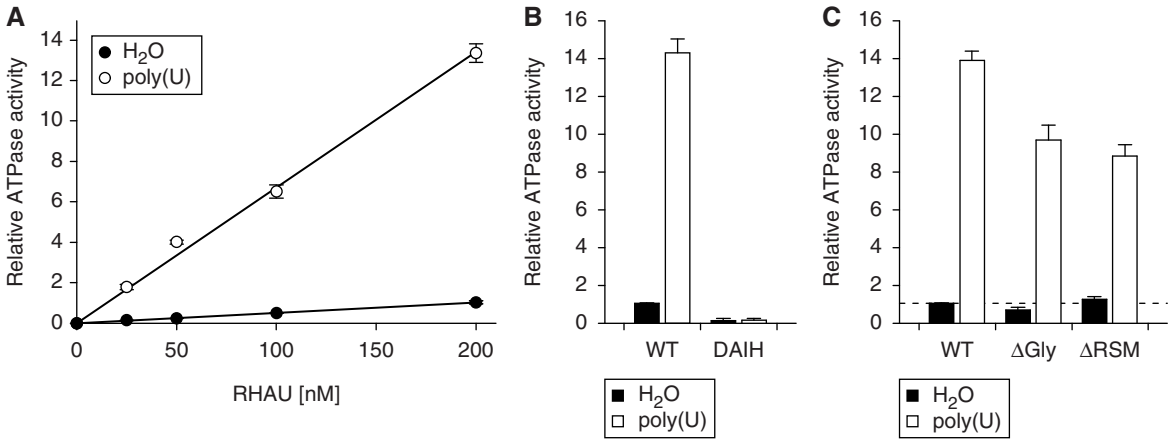


Figure 7. ATPase activity of WT RHAU and N-terminal mutants. (**A**) The extent of ATP hydrolysis by WT RHAU in the presence (open circle) or absence (filled circle) of poly(U) plotted as a function of input RHAU. The ATPase activity is expressed as a function of the activity obtained with WT RHAU at a concentration of 200 nM when poly(U) was omitted. (**B**) Influence of poly(U) on the ATPase activity of WT RHAU and DAIH ATPase-deficient mutant (200 nM). The ATPase activity was represented relatively to that of WT RHAU in the absence of poly(U) that was set to 1. (**C**) Influence of poly(U) on the ATPase activity of RHAU (WT) and ΔGly or ΔRSM RHAU mutants (200 nM). The basal ATPase activity (without nucleic acid co-factor) of WT RHAU is indicated (dashed line). The data represent the means ± SEM from three independent experiments.

similarity with the exception of the highly conserved RSM (amino acids 54–66) and its surrounding region (amino acids 54–100). Considering the apparent importance of the RSM in the recognition of G4 structures and the high conservation of its sequence together with the rest of RHAU protein, we surmised that the G4-binding and resolving activity of the RHAU protein are conserved among higher eukaryotes. To validate this hypothesis, we cloned and characterized CG9323, the *Drosophila* ortholog of RHAU.

CG9323 is a 942 amino acid-long protein, which like RHAU comprises all the typical signature motifs of the DEAH-box family of RNA helicases (Figure 8B; Supplementary Figure S5). The overall identity and similarity between CG9323 and RHAU are 34% and 52%, respectively, but not evenly distributed along the entire sequence. Sequence homology with RHAU is particularly high (41% identity and 60% similarity) for the helicase core and the C-terminal regions of CG9323. Nevertheless, apart from the conserved RSM, the CG9323 N-terminal region lacks evident sequence homology with RHAU. In addition, the CG9323 N-terminal region is 25% shorter than that of RHAU, mainly due to the absence of the Gly-rich sequence upstream of the RSM.

To check whether the *Drosophila* ortholog of RHAU retains G4 binding and resolving activities, CG9323 was expressed as a FLAG-tagged recombinant protein and immunopurified to homogeneity as shown in Supplementary Figure S6. The ability of purified recombinant CG9323 protein to unwind and to bind G4 structures was assessed under the conditions previously employed for RHAU. As shown in Figure 9A, CG9323 efficiently unwound the G4-RNA substrate. As for RHAU, the extent of products resolved by CG9323 was proportional to the input protein. Furthermore, CG9323 failed to resolve G4 substrates in the presence of the non-hydrolysable ATP analogue AMP-PNP, indicating that G4 unwinding by CG9323 requires the hydrolysis of nucleosides triphosphate.

Similar to RHAU, CG9323 also formed a stable complex with G4 substrates in the absence of hydrolysable rNTPs (Figure 9B). In binding experiments, CG9323 turned out to be more specific for G4-RNA relative to ssRNAs, since CG9323 bound tetramolecular rAGA with a 10-fold higher affinity than monomeric single-stranded rAGA oligoribonucleotides of the same sequence. This observation, by analogy, is consistent with earlier observations that RHAU has poor sequence-specific recognition of ssRNAs (16,23). Finally, a variant form of CG9323, in which highly conserved residues of the RSM were mutagenized, displayed reduced G4-binding activity, suggesting that the RSM is essential for the recognition of G4-structures by CG9323 (Figure 9C). In conclusion, these results provide compelling evidence that the G4 binding and resolving activities of RHAU have been conserved from *Drosophila* to human, with the RSM playing a pivotal role.

DISCUSSION

Owing to their bulky and thermodynamically stable features, G4 structures have been shown in many respects to impede normal nucleic acid metabolism (38–45). To cope with this problem, proteins are produced that mitigate effects of these atypical stable structures. Four human helicases, including RHAU, have been shown so far to harbour G4-resolving activity *in vitro*. These include the RecQ family proteins BLM (46,47) and WRN (48), as well as FANCJ (41), a Rad3-like helicase. The latter are all DNA helicases that have been clearly implicated in the maintenance of genome integrity (49–51). RHAU, on the other hand, belongs to the DEAH-box family of RNA helicases and has very little sequence similarity with the above mentioned helicases. Curiously, RHAU was

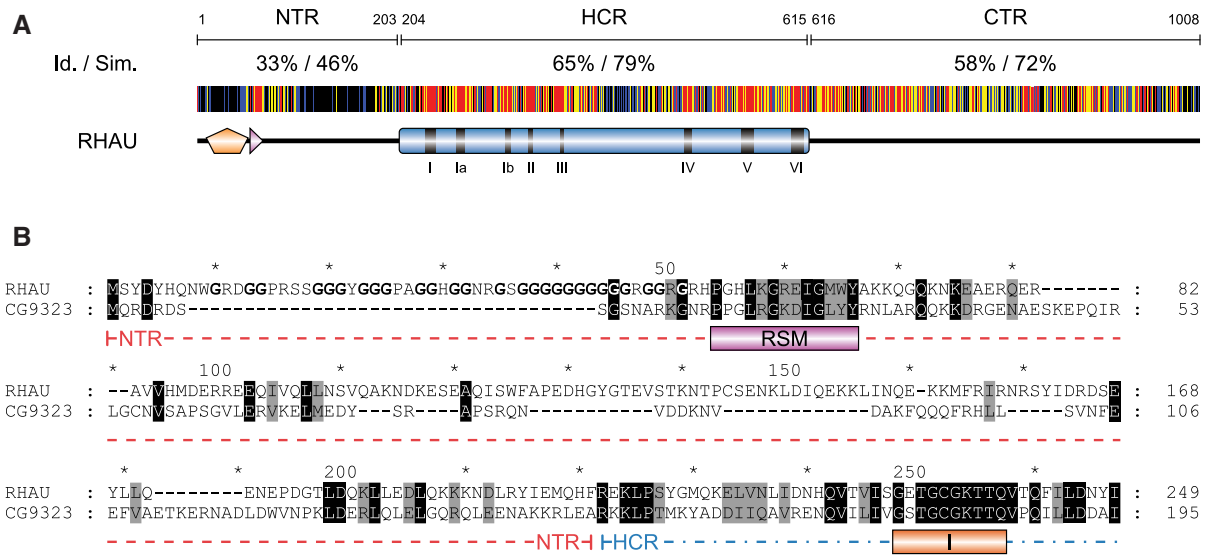


Figure 8. Amino acid conservation of the N-terminal region of RHAU. (**A**) Schematic representation of amino acid conservation of human RHAU throughout evolution. The human RHAU sequence was aligned with eight RHAU orthologues (the sequences shown in Figure 6) by MAFFT (version 6) (35). Each residue of RHAU is represented with a colour code that indicates its level of conservation amongst the eight orthologous sequences. Similarity is shown in red for 100%, yellow for 99–80% and blue for 79–60%. Similarity analysis was made by GeneDoc (version 2.7) using the BLOSUM62 scoring matrix. Average values of identity (Id) and similarity (Sim) for N-terminal (NTR), helicase core (HCR) and C-terminal (CTR) regions are indicated. (**B**) Sequence alignment of the N-terminal region of RHAU with its *Drosophila* ortholog CG9323. Amino acids that are identical or similar between the two sequences are shaded in black and grey, respectively. The RSM domain as well as helicase motif I are indicated below the sequences. Gly residues of the Gly-rich domain (amino acid 10–51) of RHAU are bolded. N-terminal region and a part of the helicase core domain are delineated with a coloured dashed line in red and blue, respectively. For complete alignment between RHAU and CG9323, see Supplementary Figure S5.

initially identified as a G4-DNA resolvase enzyme (17). The possibility that G4-RNA structures are targets of RHAU emerged from *in vivo* UV cross-linking results showing RHAU binding mainly to RNA (22). Subsequent characterization of RHAU demonstrated its aptitude to resolve G4-RNA better than G4-DNA (16). This finding was a remarkable breakthrough, since RHAU was the first and still the only known helicase to possess G4-RNA resolvase activity. Furthermore, RHAU is one of the rare DEAD/H-box proteins that exhibit high affinity and specificity for its substrate *in vitro* independent of accessory proteins. Since this finding, efforts have been made to understand the mechanism underlying recognition of G4-structures by RHAU. The present study shows that the N-terminal region of RHAU is essential and responsible for binding of RHAU to G4-structures. Further investigations dissecting the N-terminal region, coupled with site-directed mutagenesis, have demonstrated that the RSM makes a decisive contribution to the high affinity of RHAU for G4-structures. Sequence comparisons of RHAU orthologs from various species showed the RSM to be the unique highly conserved part of the N-terminal region. Hence, we predicted that all orthologs forms should possess G4-resolving activity based on the functional significance and sequence conservation of the RSM domain and the catalytic region. This hypothesis was supported by the robust ATP-dependent G4-RNA resolvase activity found for CG9323, the *Drosophila* form of RHAU. As expected, and also shown for RHAU, the binding of CG9323 to G4-RNAs depended on RSM integrity, which further indicated similar functions for this motif in both proteins.

Recognition of G4-RNA by RHAU depends on the N-terminal RSM

RHAU shares with most helicases a global scheme of modular architecture that combines a conserved central helicase core domain with peripheral regions of various lengths and sequences (52). Together with previous findings (22,53–55), our results indicate that the helicase core alone cannot account for the high specificity of function usually attributed to DEAD/H-box proteins. In this regard, these data also agree with numerous structural observations that the HCR of DEAD-box proteins interacts essentially in a non-sequence-specific manner with the phosphoribose backbone of single-stranded nucleic acid (28–32). Such contacts suffice to discriminate RNA from DNA by means of the 2′-hydroxyl groups of the ribose moieties, but not to distinguish between sequences of varying nucleotide composition. In the present work, we have shown the importance of the unique N-terminal flanking region in adapting the conserved catalytic core to a specific function. The present investigation has also shown clearly that, although necessary, ATPase activity by itself is not sufficient for RHAU to unwind G4 structures. Our data strongly suggest that the establishment of a stable complex between RHAU and its G4 substrate is a prerequisite for the subsequent ATPase-dependent unwinding of the G4 structure. We propose that the N-terminal RSM endows the enzyme with specificity by binding the G4 substrate, thereby positioning the helicase core in close proximity to the substrate. Likewise, critical roles for N- and C-terminal flanking regions have been reported for several DEAD/H-box proteins, exerting

6230 Nucleic Acids Research, 2010, Vol. 38, No. 18

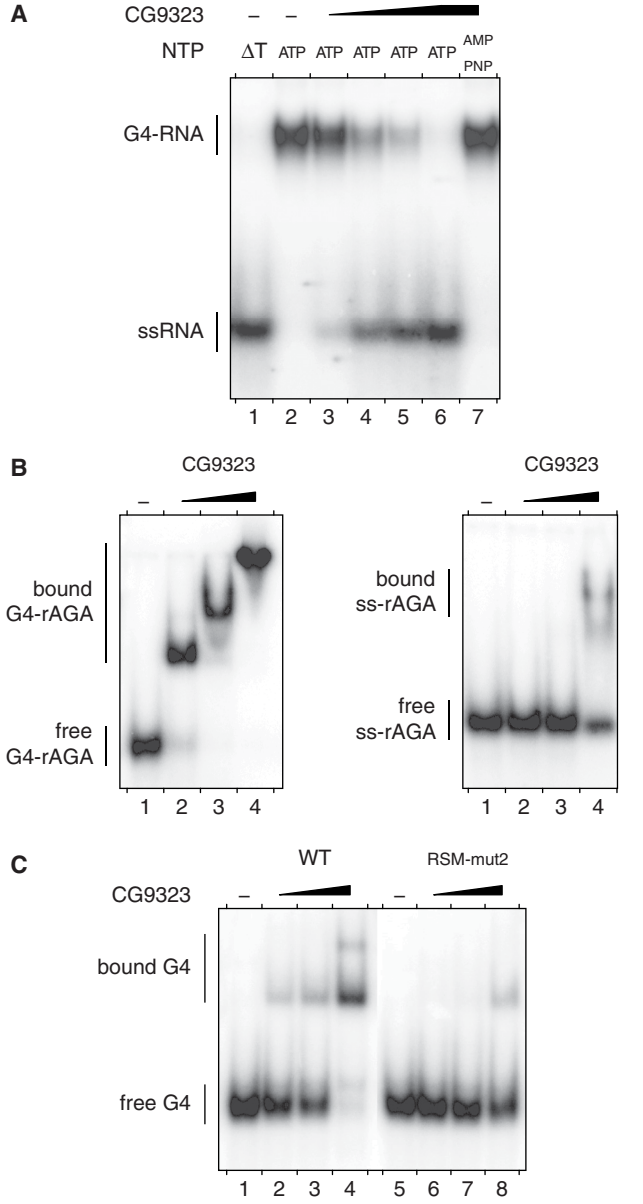


Figure 9. G4-RNA unwinding and binding by CG9323. (**A**) G4-RNA unwinding assay: radio-labelled tetramolecular rAGA at a concentration of 4 nM was incubated in the presence of ATP or AMP-PNP (as indicated) without protein (−) or with increasing amounts (2, 6, 20 and 60 nM) of purified recombinant CG9323 protein. The reaction products were resolved by native PAGE after disrupting RNA–protein interactions with SDS. An autoradiogram of the gel is shown. An aliquot of the duplex substrate that was heat-denatured (95°C, 5 min) and then quenched (ΔT) serves as a marker for the position of ssRNA. (**B**) G4- and ssRNA-binding assay: tetramolecular (left) or monomeric (right) radio-labelled rAGA at a concentration of 100 pM were incubated without protein (−) or with increasing amounts (10, 30 and 100 nM) of CG9323 in the absence of ATP. The reaction mixtures were analysed by native PAGE. An autoradiogram of the gel is shown. (**C**) G4-RNA-binding assay: radio-labelled tetramolecular rAGA at a concentration of 100 pM was incubated without protein (−) or with increasing amounts (3, 10 and 30 nM) of WT or RSM-mut2 CG9323 proteins in the absence of ATP. The reaction mixture were analysed by native PAGE. An autoradiogram of the gel is shown.

specific functions by interacting with particular RNA species/structures or with other regulatory proteins (53,56–58). For example, the prototypical yeast DEAH-box proteins Prp16 and Prp22 and the human ortholog HRH1 (alias DHX8 or hPrp22) have been shown to associate with the spliceosome via their non-conserved N-terminal regions (59–61). The *Drosophila* maleless (MLE) protein, more closely related to RHAU, harbours two copies of dsRNA-binding motifs in its N-terminal region (62). Their deletion, as with the N-terminal truncation of RHAU, caused the loss of RNA binding and unwinding activities *in vitro* and subcellular mis-localization of the protein *in vivo* (63). Together, these examples emphasize the role of the peripheral domains of DEAD/H-box proteins in adapting a common catalytic core to a broad spectrum of specific functions.

Potential role of the RSM in RHAU re-localization to SGs

The present work underscores the essential nature of the conserved RSM, which endows RHAU with a high affinity for G4 structures *in vitro*. Determining the mechanisms whereby RHAU resolve G4 substrates *in vivo* is an important issue to be addressed in future. It is tempting to draw a parallel between the present observation and a previous report that the stress-induced recruitment of RHAU to SGs is mediated by interactions of RHAU with RNA (22). Similar to our observations, the region identified as essential for this activity included the RSM together with the upstream Gly-rich domain, underscoring the functional relevance of the RSM and its surrounding region for the biological activity of RHAU. The N-terminal region of RHAU alone was also found to be sufficient to drive RNA binding *in vivo* as well to re-localize to SGs (22). This observation, however, contrasts with the data presented here in that, although required for functional specificity, the N-terminal region alone does not constitute an independent and high-affinity G4-RNA-binding domain. In this regard, our data rather suggest that both the N-terminal and HCRs are required for the productive interaction of RHAU with G4 structures. Since the nucleic acid binding motifs Ia, Ib, IV and V of DEAH-box proteins have been proposed to contact RNA (33,34), we surmise that, for RHAU, the helicase core may also provide the protein with substantial binding activity by interacting in a non-specific manner with the phosphoribose backbone of the single-stranded tail flanking the tetramolecular G4 structure. Further investigations into the functional contribution of the RNA-binding site of the helicase core are needed to further challenge this hypothesis. In addition, our previous finding that the ATPase-deficient form of RHAU [RHAU(DAIH)] stalls in SGs (22) agrees with the observation that, once bound to G4-RNA, RHAU(DAIH) cannot dissociate itself from its substrate, even in the presence of ATP (data not shown). This indicates that the G4-unwinding reaction requires ATP hydrolysis, rather than ATP binding *per se*, and further

suggests that the release of the RNA substrate occurs only after G4 unwinding. SGs may constitute a favourable environment for the formation of intermolecular G4-RNA structures, as they are temporary sites of accumulation of stress-induced stalled translation initiation RNP complexes (64,65). However, an important question that we have not yet addressed is whether RHAU relocalizes to SGs upon binding to G4-RNA structures. Interestingly, however, we noticed that RSM-mutated forms of RHAU that are deficient for G4-structure recognition *in vitro*, manifest reduced association with RNA *in vivo* concomitantly with reduced relocalization to SGs (Chalupnikova, K., unpublished data). Thus, this finding raises the possibility that at least a fraction of RHAU is recruited to SGs via interactions with G4 structures.

The mechanism by which RHAU recognizes G4 structures is an important issue that needed to be addressed to improve our understanding of RHAU. So far, however, little is known about its biological function as a potential G4-resolvase enzyme. Recently, G4 structures in RNA have attracted considerable attention as a plausible means of regulating gene expression (42). Formation of G4 structures in the 5′-untranslated region has been shown to affect mRNA translation (43–45) and bioinformatics studies have identified more than 50 000 potential G4 structures near splicing and polyadenylation sites of various human and mouse genes. This raises the possibility that G4 formation impedes RNA metabolism at many different stages (66). At the moment, we lack a corresponding understanding of how cells negotiate with G4 structures in various RNA molecules. However, from the findings presented here, RHAU emerges as a promising regulator of G4 structure-based RNA metabolism and merits future investigation of its potential roles in different biological contexts.

ACCESSION NUMBER

NP_610056.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Stéphane Thiry for excellent technical assistance and Patrick King, Janice Ching Lai and Sandra Pauli for critical comments on the manuscript

FUNDING

Novartis Research Foundation. Funding for open access charge: Novartis Research Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Hilbert,M., Karow,A.R. and Klostermeier,D. (2009) The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *Biol. Chem.*, **390**, 1237–1250.
2. Silverman,E., Edwalds-Gilbert,G. and Lin,R.J. (2003) DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene*, **312**, 1–16.
3. Jankowsky,E. and Bowers,H. (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res.*, **34**, 4181–4188.
4. Linder,P. (2006) DEAD-box proteins: a family affair–active and passive players in RNP-remodeling. *Nucleic Acids Res.*, **34**, 4168–4180.
5. Lorsch,J.R. (2002) RNA chaperones exist and DEAD box proteins get a life. *Cell*, **109**, 797–800.
6. Russell,R. (2008) RNA misfolding and the action of chaperones. *Front. Biosci.*, **13**, 1–20.
7. Rocak,S. and Linder,P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol.*, **5**, 232–241.
8. Cordin,O., Banroques,J., Tanner,N.K. and Linder,P. (2006) The DEAD-box protein family of RNA helicases. *Gene*, **367**, 17–37.
9. Bleichert,F. and Baserga,S.J. (2007) The long unwinding road of RNA helicases. *Mol. Cell*, **27**, 339–352.
10. Tanner,N.K. and Linder,P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell*, **8**, 251–262.
11. de la Cruz,J., Kressler,D. and Linder,P. (1999) Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. *Trends Biochem. Sci.*, **24**, 192–198.
12. Fuller-Pace,F.V., Nicol,S.M., Reid,A.D. and Lane,D.P. (1993) DbpA: a DEAD box protein specifically activated by 23s rRNA. *EMBO J.*, **12**, 3619–3626.
13. Kossen,K. and Uhlenbeck,O.C. (1999) Cloning and biochemical characterization of Bacillus subtilis YxiN, a DEAD protein specifically activated by 23S rRNA: delineation of a novel sub-family of bacterial DEAD proteins. *Nucleic Acids Res.*, **27**, 3811–3820.
14. O'Day,C.L., Dalbadie-McFarland,G. and Abelson,J. (1996) The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J. Biol. Chem.*, **271**, 33261–33267.
15. Perriman,R., Barta,I., Voeltz,G.K., Abelson,J. and Ares,M. Jr. (2003) ATP requirement for Prp5p function is determined by Cus2p and the structure of U2 small nuclear RNA. *Proc. Natl Acad. Sci. USA*, **100**, 13857–13862.
16. Creacy,S.D., Routh,E.D., Iwamoto,F., Nagamine,Y., Akman,S.A. and Vaughn,J.P. (2008) G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **283**, 34626–34634.
17. Vaughn,J.P., Creacy,S.D., Routh,E.D., Joyner-Butt,C., Jenkins,G.S., Pauli,S., Nagamine,Y. and Akman,S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.
18. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
19. Maizels,N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
20. Wong,H.M., Payet,L. and Huppert,J.L. (2009) Function and targeting of G-quadruplexes. *Curr. Opin. Mol. Ther.*, **11**, 146–155.
21. Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci.*, **12**, 4336–4351.
22. Chalupnikova,K., Lattmann,S., Selak,N., Iwamoto,F., Fujiki,Y. and Nagamine,Y. (2008) Recruitment of the RNA helicase

RHAU to stress granules via a unique RNA-binding domain. *J. Biol. Chem.*, **283**, 35186–35198.
23. Tran,H., Schilling,M., Wirbelauer,C., Hess,D. and Nagamine,Y. (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol. Cell*, **13**, 101–111.
24. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
25. Horton,R.M., Hunt,H.D., Ho,S.N., Pullen,J.K. and Pease,L.R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61–68.
26. Zheng,L., Baumann,U. and Reymond,J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
27. Khateb,S., Weisman-Shomer,P., Hershco,I., Loeb,L.A. and Fry,M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
28. Kim,J.L., Morgenstern,K.A., Griffith,J.P., Dwyer,M.D., Thomson,J.A., Murcko,M.A., Lin,C. and Caron,P.R. (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure*, **6**, 89–100.
29. Sengoku,T., Nureki,O., Nakamura,A., Kobayashi,S. and Yokoyama,S. (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell*, **125**, 287–300.
30. von Moeller,H., Basquin,C. and Conti,E. (2009) The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat. Struct. Mol. Biol.*, **16**, 247–254.
31. Collins,R., Karlberg,T., Lehtio,L., Schutz,P., van den Berg,S., Dahlgren,L.G., Hammarstrom,M., Weigelt,J. and Schuler,H. (2009) The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J. Biol. Chem.*, **284**, 10296–10300.
32. Del Campo,M. and Lambowitz,A.M. (2009) Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Mol. Cell*, **35**, 598–609.
33. Schneider,S., Campodonico,E. and Schwer,B. (2004) Motifs IV and V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *J. Biol. Chem.*, **279**, 8617–8626.
34. Tanaka,N. and Schwer,B. (2006) Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function. *Biochemistry*, **45**, 6510–6521.
35. Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.*, **9**, 286–298.
36. Cole,C., Barber,J.D. and Barton,G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.
37. Aasland,R., Abrams,C., Ampe,C., Ball,L.J., Bedford,M.T., Cesareni,G., Gimona,M., Hurley,J.H., Jarchau,T., Lehto,V.P. *et al.* (2002) Normalization of nomenclature for peptide motifs as ligands of modular protein domains. *FEBS Lett.*, **513**, 141–144.
38. Qin,Y. and Hurley,L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.
39. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
40. London,T.B., Barber,L.J., Mosedale,G., Kelly,G.P., Balasubramanian,S., Hickson,I.D., Boulton,S.J. and Hiom,K. (2008) FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.*, **283**, 36132–36139.
41. Wu,Y., Shin-ya,K. and Brosh,R.M. Jr. (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell Biol.*, **28**, 4116–4128.
42. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
43. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
44. Arora,A., Dutkiewicz,M., Scaria,V., Hariharan,M., Maiti,S. and Kurreck,J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA*, **14**, 1290–1296.
45. Morris,M.J. and Basu,S. (2009) An unusually stable G-quadruplex within the 5′-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry*, **48**, 5313–5319.
46. Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **273**, 27587–27592.
47. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.
48. Fry,M. and Loeb,L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **274**, 12797–12802.
49. Chu,W.K. and Hickson,I.D. (2009) RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer*, **9**, 644–654.
50. Wu,Y., Suhasini,A.N. and Brosh,R.M. Jr. (2009) Welcome the family of FANCJ-like helicases to the block of genome stability maintenance proteins. *Cell Mol. Life Sci.*, **66**, 1209–1222.
51. Wang,W. (2007) Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat. Rev. Genet.*, **8**, 735–748.
52. Singleton,M.R. and Wigley,D.B. (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.*, **184**, 1819–1826.
53. Karginov,F.V., Caruthers,J.M., Hu,Y., McKay,D.B. and Uhlenbeck,O.C. (2005) YxiN is a modular protein combining a DEx(D/H) core and a specific RNA-binding domain. *J. Biol. Chem.*, **280**, 35499–35505.
54. Grohman,J.K., Del Campo,M., Bhaskaran,H., Tijerina,P., Lambowitz,A.M. and Russell,R. (2007) Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry*, **46**, 3013–3022.
55. Mohr,G., Del Campo,M., Mohr,S., Yang,Q., Jia,H., Jankowsky,E. and Lambowitz,A.M. (2008) Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. *J. Mol. Biol.*, **375**, 1344–1364.
56. Diges,C.M. and Uhlenbeck,O.C. (2001) Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J.*, **20**, 5503–5512.
57. Tsu,C.A., Kossen,K. and Uhlenbeck,O.C. (2001) The Escherichia coli DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA. *RNA*, **7**, 702–709.
58. Kossen,K., Karginov,F.V. and Uhlenbeck,O.C. (2002) The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA. *J. Mol. Biol.*, **324**, 625–636.
59. Wang,Y. and Guthrie,C. (1998) PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. *RNA*, **4**, 1216–1229.
60. Schneider,S. and Schwer,B. (2001) Functional domains of the yeast splicing factor Prp22p. *J. Biol. Chem.*, **276**, 21184–21191.

61. Ohno,M. and Shimura,Y. (1996) A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes Dev.*, **10**, 997–1007.
62. Gibson,T.J. and Thompson,J.D. (1994) Detection of dsRNA-binding domains in RNA helicase A and Drosophila maleless: implications for monomeric RNA helicases. *Nucleic Acids Res.*, **22**, 2552–2556.
63. Izzo,A., Regnard,C., Morales,V., Kremmer,E. and Becker,P.B. (2008) Structure-function analysis of the RNA helicase maleless. *Nucleic Acids Res.*, **36**, 950–962.
64. Anderson,P. and Kedersha,N. (2006) RNA granules. *J. Cell Biol.*, **172**, 803–808.
65. Anderson,P. and Kedersha,N. (2008) Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.*, **33**, 141–150.
66. Kostadinov,R., Malhotra,N., Viotti,M., Shine,R., D'Antonio,L. and Bagga,P. (2006) GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119–D124.

9390–9404 Nucleic Acids Research, 2011, Vol. 39, No. 21 Published online 16 August 2011
doi:10.1093/nar/gkr630

The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme

Simon Lattmann[1,*], Michael B. Stadler[1], James P. Vaughn[2], Steven A. Akman[2] and Yoshikuni Nagamine[1,*]

[1]Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation, Maulbeerstrasse 66, 4058 Basel, Switzerland and [2]Department of Cancer Biology and the Comprehensive Cancer Center, Wake Forest University School of Medicine, Winston-Salem, North Carolina, NC 27157, USA

Received December 31, 2010; Revised July 11, 2011; Accepted July 19, 2011

ABSTRACT

Guanine-quadruplexes (G4) consist of non-canonical four-stranded helical arrangements of guanine-rich nucleic acid sequences. The bulky and thermodynamically stable features of G4 structures have been shown in many respects to affect normal nucleic acid metabolism. *In vivo* conversion of G4 structures to single-stranded nucleic acid requires specialized proteins with G4 destabilizing/unwinding activity. RHAU is a human DEAH-box RNA helicase that exhibits G4-RNA binding and resolving activity. In this study, we employed RIP-chip analysis to identify *en masse* RNAs associated with RHAU *in vivo*. Approximately 100 RNAs were found to be associated with RHAU and bioinformatics analysis revealed that the majority contained potential G4-forming sequences. Among the most abundant RNAs selectively enriched with RHAU, we identified the human telomerase RNA template TERC as a true target of RHAU. Remarkably, binding of RHAU to TERC depended on the presence of a stable G4 structure in the 5′-region of TERC, both *in vivo* and *in vitro*. RHAU was further found to associate with the telomerase holoenzyme via the 5′-region of TERC. Collectively, these results provide the first evidence that intramolecular G4-RNAs serve as physiologically relevant targets for RHAU. Furthermore, our results suggest the existence of alternatively folded forms of TERC in the fully assembled telomerase holoenzyme.

INTRODUCTION

DNA and RNA sequences containing tandem repeats of guanine tracts can form G-quadruplex (G4) structures, non-canonical and thermodynamically stable four-stranded helical arrangements (1,2). The building block of G4 structures is the G-quartet, a square-planar assembly of four guanines held together by Hoogsteen hydrogen bonding. G4 structures result from the consecutive stacking of G-quartets and are further stabilized by alkali metal ions such as Na^+ or K^+ that position along the helix axis and coordinate the O6 keto oxygens of the tetrad-forming guanines. G4 scaffolds are extremely polymorphic in the relative orientation of strands (parallel, anti-parallel or mixed configurations), length, sequence and conformation of the loops connecting the G-tracts. In addition, G4 structures *in vitro* can result from the assembly of one (intramolecular G4) or multiple (bi- or tetramolecular-G4) nucleic acid strands.

Biochemical and biophysical studies have shown *in vitro* that stable G4 structures can form spontaneously from G-rich regions of single-stranded nucleic acid under near physiological conditions. Genome-wide computational surveys have identified more than 300 000 potential intramolecular G4-forming sequences in the human genome (3,4) and revealed a higher prevalence of these sequences in functional genomic regions such as telomeres, promoters (5,6), untranslated regions [UTRs (7,8)] and first introns (9). Taken together, these observations suggest that G4 structures participate in regulating various nucleic acid processes, such as telomere maintenance or the control of gene expression. Nevertheless, because of the lack of evidence that such structures really exist *in vivo*, G4 structures have often been considered as a structural curiosity

*To whom correspondence should be addressed. Email: simon.lattmann@fmi.ch
Correspondence may also be addressed to Yoshikuni Nagamine. Email: yoshikuni.nagamine@fmi.ch

Nucleic Acids Research, 2011, Vol. 39, No. 21 9391

without relevance for living organisms. There is now reason to believe that G4 structures are not merely an *in vitro* artefact, as several recent findings concur with their existence in cells. First, both a G4-specific dye and antibodies raised against telomeric G4-DNA specifically stained telomeres in human and ciliate cells, respectively (10–12). In addition, several potential G4-forming sequences in promoters were shown to form intramolecular G4 structures *in vitro* and to affect gene expression *in vivo* (13,14). A possible contribution of G4 to regulating promoter activity was indicated by impairment of the transcriptional activity of several genes by G4-stabilizing ligands (14) or a single-chain antibody specific for intramolecular G4-DNA (15), in a manner correlating with the occurrence of predicted G4 structures in the control regions (16).

Like DNA, RNA can also form G4 structures. Although, to date, G4-RNAs have not attracted as much attention as their DNA counterparts, the formation of G4 structures in RNA is emerging as a plausible regulatory factor in gene expression. RNA is more prone than DNA to form G4 structures due to its single-strandedness, and G4-RNAs have also proved to be more stable than their cognate G4-DNA under physiological conditions (17–19). Bioinformatics analyses of human 5′-UTR sequences revealed potential G4-forming motifs in as many as 3000 different RNAs (7,20). Moreover, the formation of G4 structures in 5′-UTR was shown to impede translation initiation (7,21–23). Given that potential G4 sequences have been identified near splicing and polyadenylation sites (24–26), G4 formation may also affect RNA metabolism at several different stages. Furthermore, formation of parallel G4-RNA structures has also been reported *in vitro* for telomeric RNA repeats [TERRA, (27–29)] and for the human telomerase template RNA [TERC, (30)], suggesting that G4-RNA formation also plays a part in regulatory processes at telomeres.

The discovery of proteins that positively or negatively stabilize such G4 structures is further indirect evidence for the existence of such structures *in vivo*. Among the many proteins from various organisms that bind to G4-DNAs *in vitro* (31), several helicases show ATP-dependent G4-resolving activity (32–36) and have been clearly implicated in the maintenance of genome integrity (37–40). RHAU (alias DHX36 or G4R1), a member of the human DEAH-box family of RNA helicases, exhibits *in vitro* G4-RNA binding with high affinity for its substrate, and unwinds G4 structures much more efficiently than double-stranded nucleic acid (41,42). Consistent with these biochemical observations, RHAU was also shown to bind to mRNAs *in vivo* (43) and was identified as the main source of tetramolecular G4-RNA-resolving activity in HeLa cell lysates (42).

Although considerable information is available on the enzymatic activity of RHAU *in vitro*, almost nothing is known about its biological function as a G4-binding/resolving enzyme *in vivo*. To address this question, we sought RNAs bound by RHAU in living cells, surmising that the identification of RHAU-bound RNAs and understanding the effects of RHAU should provide important clues to its function *in vivo*. To this end, we employed high-throughput gene array technologies to identify and quantify the co-purified RNAs on a genome-wide scale. With this screen, we identified about 100 RNAs significantly enriched by RHAU. Computational analysis of RNA sequences for potential intramolecular G4 structures revealed the preferential association of RHAU with transcripts bearing G4-forming motifs, suggesting direct targeting of G4-RNA by RHAU. Amongst the RNAs with the potential to form G4 structures and selectively enriched with RHAU, we identified TERC as a *bona fide* target of RHAU. Characterization of the RHAU-TERC interaction *in vivo* and *in vitro* showed binding of TERC by RHAU to be strictly dependent on the formation of a G4 structure on the 5′-extremity of TERC RNA. Finally, we have demonstrated that RHAU not only interacts with TERC *stricto sensu*, but is also part of the fully assembled telomerase complex through direct interaction with TERC G4 structure. Taken together, these data demonstrate that intramolecular G4-RNAs are naturally occurring substrates of RHAU *in vivo*. Moreover, they provide indirect but strong support for the existence of a G4-RNA structure in the telomerase RNP.

MATERIALS AND METHODS

Plasmid constructs, cloning and mutagenesis

The baculoviral expression vector employed for the expression of GST-RHAU and the plasmids used for the expression of the C-terminal FLAG-tagged recombinant RHAU proteins RHAU-FLAG and RHAU(ΔRSM)-FLAG were previously described (43–45). The pIRES.EGFP-myc-N1 vector expressing the C-terminal myc-tagged recombinant RHAU protein was derived from the previously described pIRES.EGFP-FLAG-N1/RHAU expression vector (43) by substituting the FLAG sequence with the myc epitope sequence. Human TERT cDNA (clone: IOH36343, mapped sequence: NM_198253) was obtained from ImaGenes GmbH (Berlin, Germany) and subcloned by PCR amplification into the pSL1-FLAG-N1 mammalian expression vector. The pSL1-FLAG-N1 vector was derived from pEGFP-N1 (Clontech) by replacing the EGFP open reading frame with the FLAG epitope sequence. The human TERC genomic region was PCR-amplified from the genomic DNA of HEK293T cells to yield a 2.1-kb fragment encompassing 1080 bp of the TERC promoter, its coding sequence, and 553 bp of the 3′ flanking genomic region. This fragment was blunt-cloned into pGEM-T Easy vector (Promega) at the HincII/EcoRV sites. The G4 motif sequence of TERC was mutagenized using a variation of the classical QuikChange® (Stratagene) site-directed mutagenesis PCR method (46). To prepare templates for *in vitro* run-off transcription, the T7 or SP6 phage promoters were inserted upstream of the TERC coding sequence by PCR. The resulting PCR products were cloned into the pSL1-FLAG-N1 vector at the NheI/AgeI sites. Following linearization with NarI or AgeI, *in vitro* transcription of these templates yielded the TERC (1–71 nt) and full-length TERC (1–451 nt) RNA fragments, respectively. Constructions of all these plasmids were confirmed by sequencing. Sequences of

9392 Nucleic Acids Research, 2011, Vol. 39, No. 21

oligonucleotides used in this work and detailed descriptions of the plasmid constructs are available upon request.

Cell culture and transfection

Human cervical carcinoma HeLa and embryonic kidney HEK293T cell lines were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS) and 2 mM L-glutamine at 37°C in a humidified 5% CO_2 incubator. Transient transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Transfected cells were cultured for 24–36 h prior to testing for transgene expression.

RIP-chip assay

Cells were harvested 24–36 h post-transfection, washed with ice-cold PBS and resuspended in lysis buffer {1× PBS, 1% $v \cdot v^{-1}$ Nonidet P-40, 2 mM EDTA, 2 mM AEBSF [4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride], 1× protease inhibitor cocktail (Complete EDTA-free, Roche), 0.2 $U \cdot \mu l^{-1}$ RNasin® Plus (Promega)} for 30 min. All subsequent operations were performed at 4°C. The lysates were cleared by centrifugation (21 000*g*, 15 min) and mixed with 10 µl of a 50% slurry of anti-FLAG M2-agarose affinity gel (Sigma) that had been equilibrated in lysis buffer. After gentle agitation for 5 h, the resin was recovered by centrifugation, washed 3× with 500 µl lysis buffer, followed by three washes with 500 µl IP-washing buffer [50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 0.1% $v \cdot v^{-1}$ Nonidet P-40, 5 mM EDTA, 0.4 $U \cdot \mu l^{-1}$ RNasin® Plus]. The resin was resuspended in 1 ml TRIzol (Invitrogen) for protein analysis and RNA extraction. For microarray analysis, 100 ng RNA was converted to cRNA according to the manufacturer's guidelines and the reaction products hybridized to GeneChip® Human Gene 1.0 ST arrays (Affymetrix). The resulting raw expression values were RMA-normalized using R/BioConductor (47) and the Oligo Package [version 1.14.0 (48)]. Probesets were linked to Entrez Gene entries using Affymetrix annotation (NetAffx release 28, 11 March 2009), retaining a single probeset per gene. Genes not clearly detected in input samples [average $\log_2$(expression value) < 6.0] were discarded. Differential expression was determined using the Limma package (49), selecting genes with a minimal fold-change of 2.0 and an FDR-adjusted $P < 0.01$. RHAU targets were identified as transcripts enriched in RHAU-FLAG IP versus RHAU-FLAG input samples but not enriched in control IP versus control RHAU-myc input samples.

G4-RNA structure prediction and bioinformatics analysis

Human RNA sequences were retrieved from Entrez Nucleotide database. Non-overlapping putative intramolecular G4-forming sequences (PQS) and the corresponding G4-score values were computed with the QGRS Mapper algorithm (50) using the default parameters (window size = 30 nt, Min. G-group = 2, loop size = 0–36). For each transcript, the ∑(G4-score) value was calculated as the sum of all non-overlapping G4-scores computed by QGRS Mapper. To obtain a normalized ∑(G4-score) value, the ∑(G4-score) value was divided by the RNA length (kb). Randomization of RNA sequences by single- or dinucleotide shuffling was performed using the Altschul-Erickson algorithm (51).

Protein immunoprecipitation assay

Protein immunoprecipitation experiments were performed under the same conditions employed for the RIP-chip assay. Cleared cell lysates were mixed with rProtein A or Protein G Sepharose™ Fast Flow beads (GE Healthcare) and appropriate antibodies {mouse mAb anti-RHAU [12F33 (43)], rabbit pAb anti-dyskerin (H-300, Santa Cruz Biotechnology)}. After gentle agitation for 5 h, the beads were washed. For telomere repeat amplification protocol (TRAP) assays, beads were resuspended in 40 µl TRAP lysis buffer [10 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM $MgCl_2$, 1 mM EDTA (pH 8.0), 1% $v \cdot v^{-1}$ Nonidet P-40, 0.25 mM Na-deoxycholate, 10% $v \cdot v^{-1}$ glycerol, 5 mM 2-mercaptoethanol, 0.1 mM AEBSF]. For protein and RNA analysis, beads were resuspended in 1 ml TRIzol and extraction performed according to the manufacturer's instructions. Input and co-purified RNA samples were analysed by RT–qPCR. For protein analysis only, beads were directly resuspended in sodium dodecyl sulphate (SDS)-gel loading buffer. Input and immunoprecipitated protein samples were separated by SDS–PAGE and analysed by western blotting.

RNA analysis by quantitative (RT–qPCR) and semi-quantitative RT–PCR

Reverse transcription was performed using the ImProm-II™ Reverse Transcription System (Promega) with oligo$(dT)_{15}$ or random hexamer primers, according to the manufacturer's instructions. For monitoring first-strand cDNA synthesis, the reverse transcription reaction was performed in the presence of 20 µCi [α-^{32}P]dATP (3000 $Ci \cdot mmol^{-1}$). The reaction products were separated by agarose gel electrophoresis and visualized by PhosphorImaging. Quantitative and semi-quantitative PCR reactions were performed in technical duplicates using the ABsolute™ QPCR SYBR® Green ROX Mix (Thermo Fisher Scientific), according to the manufacturer's instructions, on an ABI Prism 7000 Sequence Detection System (SDS) and analysed with the ABI Prism 7000 SDS 1.0 Software (Applied Biosystems). Relative transcript levels were determined using the $2^{-\Delta Ct}$ method (52). For each primer pair (Supplementary Table S1), the efficiency of amplification was determined to be equal or superior to 1.8. Control reactions lacking the reverse transcriptase or template RNA confirmed the specificity of the amplification reactions.

TRAP assays

Immunopurified ribonucleoprotein (RNP) complexes were assayed for telomerase activity by the TRAP assay (53). Four microliters of bead-immobilized RNP complexes in TRAP lysis buffer were incubated (30 min, 25°C) in 50 µl TRAP reaction buffer [20 mM Tris–HCl (pH 8.3 at room temperature), 63 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTP

mix (50 µM each of dATP, dTTP, dGTP and dCTP), 0.05% v•v^{-1} Tween®-20, 1 mM EGTA (pH 8.0), 4 ng•µl^{-1} Cy5-TS primer, 2 ng•µl^{-1} ACX primer (Supplementary Table S1), 1 U•µl^{-1} RNasin® Plus, 0.4 mg•ml^{-1} BSA, 0.04 U•µl^{-1} Thermo-Start™ *Taq* DNA Polymerase (Thermo Fisher Scientific)]. The reaction was followed by a 15-min incubation step at 95°C, followed by 15–17 cycles of amplification (30 s at 95°C, 30 s at 52°C, 45 s at 72°C). The reaction products were resolved on a pre-electrophoresed 10% non-denaturing polyacrylamide gel (19:1 acrylamide:*bis* ratio) in 0.5× TBE at 4°C for 90 min. After electrophoresis, gels were fixed [500 mM NaCl, 50% v•v^{-1} ethanol, 40 mM Na-acetate (pH 4.2)] for 30 min, scanned on a Typhoon 9400 Imager (GE Healthcare) and analysed with ImageQuant TL software (Nonlinear Dynamics). For immunodepletion experiments of RHAU, the residual telomerase activity in cell extracts was quantified by the quantitative TRAP assay [qTRAP, (53)]. An aliquot of 250 ng protein from immunodepleted HEK293T cell extracts in TRAP lysis buffer was incubated (30 min, 25°C) in 25 µl qTRAP reaction buffer [1× ABsolute™ QPCR SYBR® Green ROX Mix, 1 mM EGTA (pH 8.0), 4 ng•µl^{-1} TS primer, 4 ng•µl^{-1} ACX primer (Supplementary Table S1), 0.2 U•µl^{-1} RNasin® Plus]. The reaction was followed by a 15-min incubation step at 95°C, followed by 40 cycles of amplification (15 s at 95°C, 60 s at 60°C) on an ABI Prism 7000 Sequence Detector. The relative telomerase activity was determined using a standard curve and linear equation model (53).

Expression and purification of recombinant RHAU protein

Recombinant wild-type and ATPase deficient [RHAU(DAIH)] N-terminal GST-tagged RHAU proteins were expressed in *Sf9* cells according to the supplier's instructions (PharMingen) and purified to homogeneity as described previously (45). Purified recombinant GST-RHAU proteins were stored at −80°C. Purity of protein preparations was assessed by SDS–PAGE (Supplementary Figure S1) and protein concentrations determined photometrically at 280 nm using the calculated extinction coefficient ε = 166 615 M^{-1}•cm^{-1}.

In vitro synthesis of ^{32}P-labeled TERC transcripts and intramolecular G4-RNA preparation

Synthetic radio-labeled wild-type and mutant (G4-MT) telomerase RNAs were prepared by *in vitro* transcription using 50 µCi [α-^{32}P]UTP (3000 Ci•mmol^{-1}) and T7 or SP6 RNA polymerases (Promega), respectively, according to the manufacturer's instructions. The transcripts were purified by denaturing PAGE, ethanol precipitated and recovered by centrifugation. The purified RNAs were re-suspended in potassium- or lithium-based storage buffer [10 mM Li-cacodylate (pH 7.4), 100 mM KCl or LiCl] and annealed by heating at 95°C for 2 min, followed by slow cooling to room temperature. G4-annealed radio-labeled RNAs were stored at −80°C.

RNA electromobility shift assay

Purified recombinant GST-RHAU protein at concentrations from 1 to 320 nM was incubated with 100 pM ^{32}P-labeled G4-RNA in K-Res buffer [50 mM Tris–acetate (pH 7.8), 100 mM KCl, 10 mM NaCl, 3 mM $MgCl_2$, 70 mM glycine, 10% glycerol], supplemented with 10 mM EDTA and 0.2 U•µl^{-1} SUPERase-In (Ambion) in a 10-µl reaction. The reactions were equilibrated at 22°C for 30 min. RNA–protein complexes were resolved on a pre-electrophoresed 6% non-denaturing polyacrylamide gel (37.5:1 acrylamide:*bis* ratio) in 0.5× TBE at 4°C for 90 min. After electrophoresis, gels were fixed for 1 h in 10% isopropanol/7% acetic acid. RNA–protein complexes were detected by Phosphor-Imaging, scanned on a Typhoon 9400 Imager (GE Healthcare) and analysed with ImageQuant TL software (Nonlinear Dynamics).

RESULTS

Microarray identification of RHAU-associated RNAs

To identify endogenous RNAs associated with RHAU *in vivo* on a genome-wide scale, we designed a RIP-chip assay (RNA immunoprecipitation coupled to microarray analysis). Subsets of RHAU target RNAs were isolated by immunoprecipitation (IP) assays under optimized conditions that preserved RNA–protein complexes. Briefly, HeLa cells were transfected with a vector expressing FLAG-tagged or myc-tagged RHAU. Immunoprecipitations of non-cross-linked whole-cell extracts were carried out using anti-FLAG antibodies. Anti-FLAG IP from cells expressing myc-tagged RHAU was employed as an IP control (IP_{ctrl}) to assess non-specific interactions that may occur during RIP. Following IP, co-fractionated RNAs were recovered and purified by standard phenol–chloroform extraction and converted to cRNA. Products were subsequently hybridized to human oligonucleotide arrays.

Western blot analysis of immunoprecipitated proteins revealed that RHAU-FLAG, but not RHAU-myc, was efficiently enriched from HeLa cell extracts following IP with anti-FLAG antibody (Figure 1A, bottom). Oligo(dT)-primed reverser-transcription of co-immunoprecipitated RNAs showed the presence of polyadenylated RNAs with sizes similar to the input (Figure 1A, top, compare lane 2 to lanes 3 and 4). With regard to the RHAU-FLAG IP fraction, 10 times less RNA signal was detected in the control immunoprecipitate (lane 1), demonstrating specific co-immunoprecipitation of mRNAs with RHAU. The association between RHAU and target RNAs was deemed specific as the anti-FLAG antibody exhibited no obvious cross-reactivity with other cellular proteins (data not shown).

Microarray analysis of the purified RNAs recovered from input and immunoprecipitated fractions revealed that 9354 (49%) of the 19 089 total genes available on the chip were significantly expressed in HeLa cells overexpressing RHAU (data not shown). In order to maximize the chance of identifying true RNA targets of RHAU as well as to minimize the occurrence of false positives, only those RNAs were chosen that were significantly (adjusted $P < 0.01$) enriched and were at least 2-fold more abundant in the RHAU-FLAG IP fraction than in the control. Of these potential RHAU targets, we discarded those that were also significantly enriched in the

9394 Nucleic Acids Research, 2011, Vol. 39, No. 21

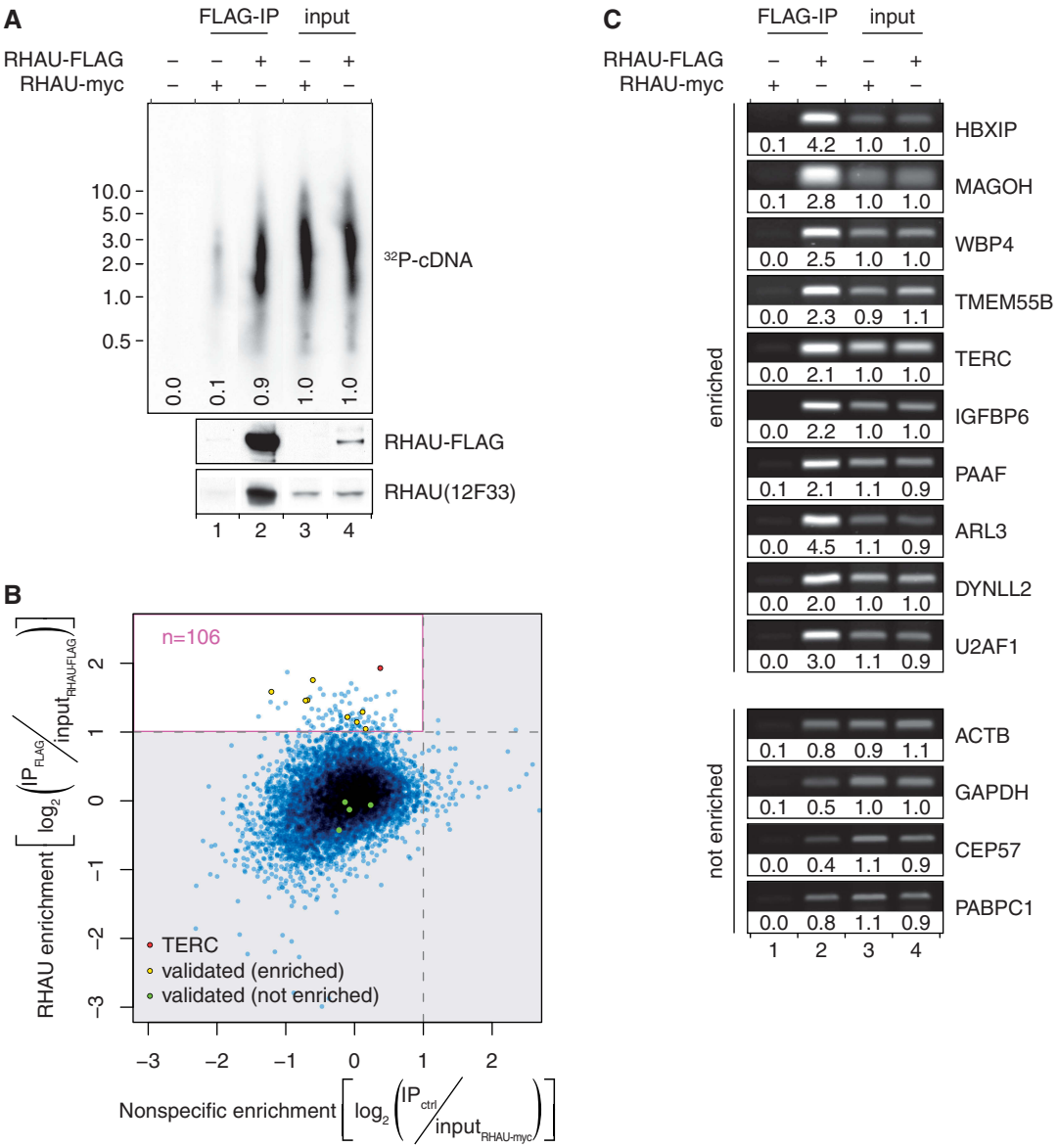


Figure 1. Analysis of the RNAs co-immunoprecipitated with RHAU RNPs. (**A**) RHAU associates with poly(A)$^+$ RNAs in HeLa cells. First strand cDNA synthesis from input (lanes 3 and 4) and co-immunoprecipitated (lanes 1 and 2) RNAs was monitored by incorporation of [α-^{32}P]dATP. The RT–PCR reaction products were separated by agarose gel electrophoresis. An autoradiogram of the gel is shown. The positions and sizes (kb) of marker DNAs are indicated at the left. Signal intensities are expressed relative to the average signal intensity of the input fractions (lanes 3 and 4). Expression and specific enrichment of RHAU-FLAG in input and FLAG-IP fractions were verified by Western blot analysis. (**B**) Scatter plot representation of the differential enrichment of RHAU-bound versus non-specific RNAs as quantified by microarray analysis. The pink rectangle delineates the area of RNA specifically enriched by RHAU. RNAs whose enrichment was further validated by semi-quantitative RT–PCR are indicated. (**C**) Validation of novel RHAU-bound target RNAs. The abundance of 10 potential RHAU targets and four non-targeted RNAs in total input RNA (lanes 3 and 4), control IP (lane 1) and RHAU-FLAG IP (lane 2) fractions was monitored by semi-quantitative RT–PCR. Reaction products were separated by agarose gel electrophoresis and visualized by SYBR Green staining. Band intensities are expressed relative to the average signal intensity of the input fractions (lanes 3 and 4).

RHAU-myc expressing cells and might thus be RHAU-independent RNAs binding non-specifically to the anti-FLAG antibody matrix.

Thus, of the 9354 genes expressed in HeLa cells, 108 RNAs (1.2%) were found to be significantly enriched in RHAU-FLAG IP fraction compared with total input RNA (Figure 1B). Finally, after subtraction of two non-target RNAs based on the above-mentioned criteria that were associated with the antibody or the beads, 106 RNAs were judged to be specifically enriched by RHAU (Supplementary Table S2). The most abundant transcripts of these potential RHAU targets were selected for further analysis.

Validation of potential RHAU target RNAs

To assess independently the validity and reproducibility of the identified RHAU-associated transcripts, RNA abundance in total input RNA, control IP and RHAU-FLAG IP fractions from fresh whole-cell extracts were

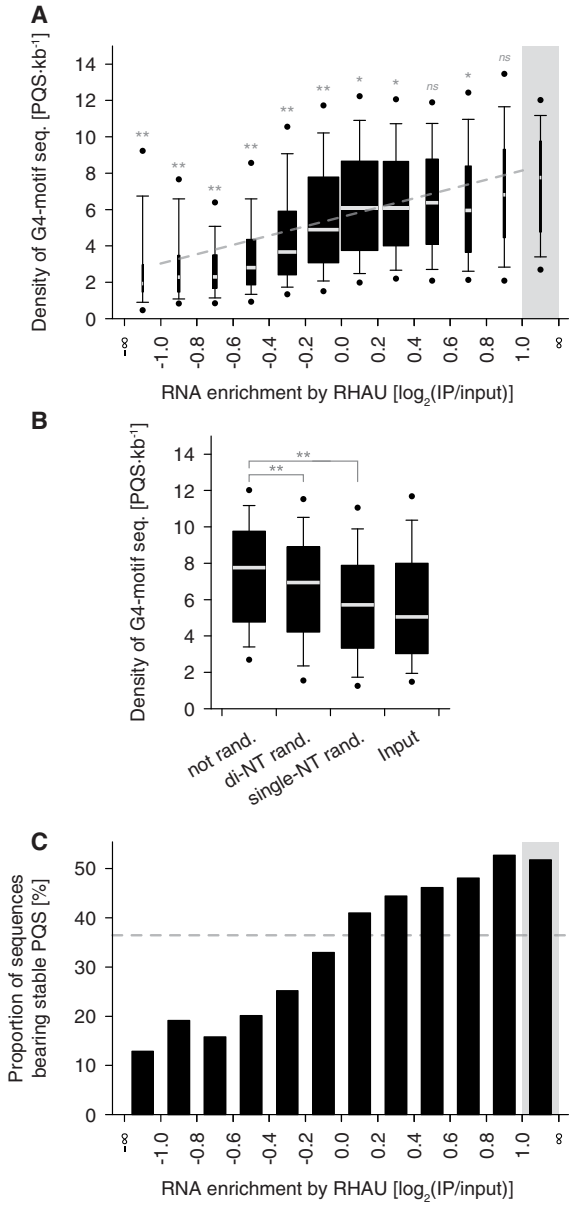


Figure 2. Computational analysis of potential intramolecular G4-forming sequences (PQS) among RNAs enriched by RHAU. (**A**) Box plot representation of the density of potential intramolecular G4-forming sequences per transcript ($PQS \cdot kb^{-1}$) as a function of their level of enrichment by RHAU. The lower and upper boundaries of the boxes indicate the 25th and 75th percentiles, respectively. The white lines within the boxes mark the median. The width of the boxes is proportional to the number of transcripts found within a group. The error bars indicate the 10th and 90th percentiles and the filled circles the 5th and 95th percentiles. The grey rectangle refers to the group of 106 RNAs specifically enriched by RHAU and serves as a control group for multiple comparison analysis. Statistical significance was determined by Kruskal–Wallis one-way ANOVA on ranks and the Dunn's test. $^{ns}P \geq 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$. The correlation between the two variables was estimated by linear regression analysis ($y = 2.56x + 5.60$, $R^2 = 0.10$, $P < 0.001$) and is indicated as a grey line on the graph. (**B**) PQS analysis among randomized RHAU target sequences. The sequences of the 106 transcripts specifically enriched by RHAU were shuffled, retaining single- or di-nucleotide base composition of the original transcripts. The box plot represents the density of PQS ($PQS \cdot kb^{-1}$) among the original (not rand.), di-nucleotide shuffled (di-NT rand.) and single-nucleotide shuffled (single-NT rand.) sequences. Statistical significance was determined by one-way repeated measures ANOVA and the Dunnett's test. (**C**) Proportions of RNAs bearing stable PQS as a function of their level of enrichment by RHAU. The grey rectangle refers to the group of 106 RNAs specifically enriched by RHAU. The dashed grey line denotes the proportion of sequences showing stable PQS in the input fraction. Significant ($P < 0.001$) association between the magnitudes of the two variables was estimated by the chi-square-test-for-trend.

analysed by semi-quantitative RT–PCR. As shown in Figure 1C, 10 RNAs randomly selected from the group of plausible RHAU targets were confirmed to be enriched >2-fold in the RHAU-FLAG IP fraction than in the control IP or input fractions. Besides, four non-targeted RNAs were included as negative control to monitor RHAU binding to non-specific RNAs. None of them were found to be enriched in the RHAU-FLAG IP fraction despite their relatively high abundance in the input fraction. Taken together, these independents results validate the previous RIP-chip data. We assume that the remaining RNA species are also part of RHAU RNPs, although additional experiments would be necessary to confirm this supposition.

G4-content analysis for RNAs enriched by RHAU

As RHAU shows a high affinity for G4-RNA structures *in vitro* (42), we next carried out a bioinformatics search for RNAs with potential intramolecular G4 structures. The rationale was that if RHAU binds G4-RNA *in vivo*, the proportion of potential G4-forming sequences should be higher among RNAs enriched by RHAU than among non-enriched RNAs. Of the various available methods for predicting intramolecular G4 motif sequences, we used QGRS Mapper that identifies and scores each potential G4-forming sequence according to their predicted stabilities (50). Being aware that only limited experimental data so far support the scoring method of QGRS Mapper, we employed the algorithm to identify potential G4-forming sequences, but we also included the G4-score-based analysis as Supplementary Data (Supplementary Figure S2), since the two approaches provided similar results. In fact, with almost eight potential intramolecular G4-forming sequences (PQS) per kilobase, the occurrence of G4 motif sequences was higher in the fraction enriched by RHAU than in any other fraction (Figure 2A and Supplementary Figure S2A). In addition, there was a weak ($R^2 = 0.10$) but significant positive correlation between the predicted G4 motif density per transcript and the magnitude of RNA enrichment by RHAU. Randomization of RHAU target sequences, retaining single- or even dinucleotide frequencies of the original transcripts, significantly reduced the proportion of potential G4-forming sequences (Figure 2B and Supplementary Figure S2B). Thus, the proportion of G4 motif sequences among RHAU-associated transcripts cannot be explained by a mere bias in nucleotide composition.

It should be mentioned that our prediction is likely to overestimate the number of G4-forming motif sequences per transcript. In fact, we also considered G4 structures consisting of only two successive G-tetrads as these metastable structures have also been demonstrated to have

9396 Nucleic Acids Research, 2011, Vol. 39, No. 21

biologically relevant roles (54). However, after discarding these less-stable G4 structures from the analysis, we still observed a significant relationship between the proportion of sequences bearing potentially stable G4-forming sequences (composed of three or more stacked G-quartets) and the magnitude of RNA enrichment by RHAU (Figure 2C). Indeed, with more than half of the sequences presenting potentially stable G4-forming sequences, the group of 106 RNAs identified as true RHAU targets presents one of the highest incidences of potentially stable G4-forming motifs. In summary, the present data are consistent with preferential association of RHAU with transcripts containing potential G4-forming sequences, suggesting direct recognition of G4-RNA structures by RHAU. To further test this hypothesis, we selected one of the identified RHAU targets and addressed the molecular basis of its interaction with RHAU.

RHAU associates with TERC through its G4 motif sequence

Within the pool of transcripts enriched by RHAU and showing a high potential to form stable G4, TERC is the only RNA reported previously to form a stable and parallel G4 structure *in vitro* (Figure 3A) (30). TERC was

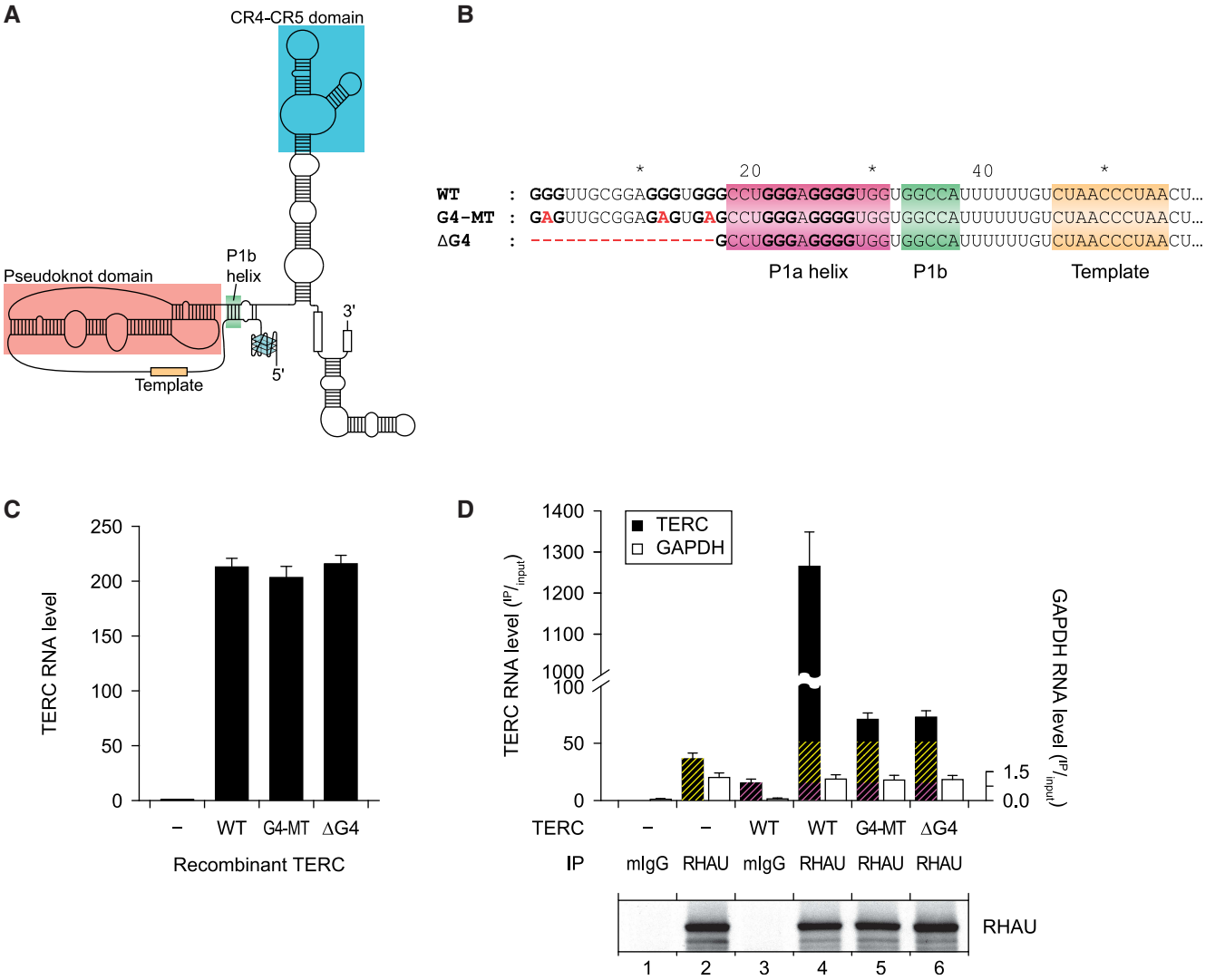


Figure 3. RHAU associates with TERC through its G4-motif sequence. (**A**) Schematic representation of the secondary structure of human TERC bearing a parallel G4 structure in the 5′-extremity as described by Mergny et *al.* (30). (**B**) Nucleotide sequence of the WT and G4 motif mutant (G4-MT and ΔG4) forms of TERC. Guanine tracts that are predicted to form a stable G4-structure are shown bold. Nucleotide substitutions or deletions in mutant forms of TERC are marked in red. The P1 helix subdomains as well as the template region are indicated. (**C**) RNA expression levels of endogenous and recombinant WT and G4 motif mutant forms of TERC in HEK293T cells. Expression was quantified by RT–qPCR, normalized to GAPDH expression and endogenous TERC levels set to 1. Data represent the mean ± SEM of three independent experiments. (**D**) RT–qPCR analysis of the abundance of WT and G4 motif mutant forms of TERC that co-immunoprecipitated with endogenous RHAU protein. RNA levels in IP fractions are represented in function of their respective abundance in the input fraction. Immunoprecipitation with mouse IgGs (mIgG) served as a control to assess non-specific interactions. In lanes 4, 5 and 6, yellow hatches represent the fraction related to endogenous TERC signal and violet hatches represent the fraction of TERC RNA that non-specifically interacts with the antibody matrix or the beads. Data represent the mean ± SEM of five independent experiments. Comparable efficiency of RHAU immunoprecipitation in the various fractions was verified by western blot analysis with anti-RHAU antibodies.

also found to be one of the most abundant RNAs enriched by RHAU in several independent RIP-chip assays using various cell lines (Supplementary Table S2 and unpublished data). To address the role of the TERC G4 structure in its efficient co-immunoprecipitation by RHAU, we cloned the TERC gene, including its promoter and 3′ flanking genomic sequence. The 5′ extremity of the TERC sequence was subsequently mutated (G4-MT) or truncated (ΔG4) to prevent G4 formation (Figure 3B). To avoid effects on the structurally conserved P1 helix by the introduced substitutions, guanine residues that are part of both the predicted G4 structure and the P1a helix region were left intact. The recombinant forms of TERC were transiently transfected into HEK293T cells and the accumulation of stable TERC transcripts was monitored by RT–qPCR (Figure3C). After 24 h, HEK293T cells transiently transfected with the TERC(WT) construct but not with the vector alone showed a substantial (~200-fold) increase in TERC abundance relative to endogenous TERC levels. Importantly, mutations of the G4 motif (G4-MT or ΔG4) did not influence the steady-state level of exogenous TERC expression in these cells since recombinant wild-type, G4-MT and ΔG4 forms of TERC accumulated to comparable levels.

To examine the significance of the G4 motif sequence for co-precipitation of TERC with RHAU, we repeated the RIP assay using HEK293T cells transiently overexpressing the wild-type or G4-MT and ΔG4 mutated forms of TERC. Immunoprecipitations were carried out on endogenous RHAU using a monoclonal antibody against RHAU and TERC RNA abundance was analysed by RT–qPCR (Figure 3D). As already shown for endogenous TERC, a substantial amount of overexpressed TERC(WT) was also recovered with endogenous RHAU, as evidenced by a 25-fold enrichment of TERC transcript over the mock transfection (Figure 3D, compare lanes 4 and 2). In marked contrast, both G4 motif mutated forms of TERC were barely co-immunoprecipitated with RHAU, judged by the 95% reduction in TERC abundance in the corresponding fractions compared with the WT control (Figure 3D, compare lanes 5 and 6 to lane 4). The drastic reduction in immunoprecipitation of the TERC mutants was not due to non-specific RNA degradation in these fractions. Indeed, comparable levels of non-specifically co-immunoprecipitated GAPDH RNA were found to contaminate all of these fractions. Taken together, these results indicate that the intact G4 motif sequence is a prerequisite for TERC association with RHAU *in vivo*.

RHAU binds TERC through a G4 structure in the TERC 5′-region

In the absence of Mg•NTPs, RHAU specifically binds to tetramolecular G4-RNA structures with high affinity (42,45). The strict requirement for the G4 motif sequence for effective recovery of TERC by RHAU *in vivo* suggested that RHAU may also form a stable complex with TERC through direct binding of the G4-RNA forming structure. To address this question *in vitro*, we performed RNA electromobility shift assays (REMSA) using purified recombinant GST-tagged RHAU protein and *in vitro* transcribed radio-labeled full-length (1–451 nt) and 5′-end (1–71 nt) TERC fragments. In the absence of protein, the ^{32}P-labeled probes migrated as a single species in the gel (Figure 4A and B). Addition of increasing amounts of RHAU to both full-length and 5′-end TERC fragments resulted in the appearance of a high-affinity (estimated K_d of 10 nM) ribonucleoprotein complex of reduced mobility. In contrast, the stability of the RHAU-TERC interaction was strongly impaired (~20-fold reduction) when formation of the G4 structure was prevented by mutation of the G4 motif sequence (G4-MT, Figure 4C). Similarly, conditions that are unfavourable to G4 stability (substitution of K^+ for Li^+) impaired the RHAU-TERC interaction to comparable extent (Figure 4D). Replacing K^+ by Li^+ was indeed shown to strongly reduce the thermodynamic stability of the TERC G4 structure (30). Thus, these results demonstrate a direct and specific interaction between RHAU and TERC dependent on RNA folding into a stable G4 structure. This is consistent with the above finding that RHAU can co-immunoprecipitate the wild-type but not the G4 motif mutated forms of TERC. Together these observations argue that a fraction of TERC RNA forms a G4 structure that can be further bound by RHAU, *in vivo*.

RHAU associates with telomerase RNPs by direct interaction with TERC

In cells, the biogenesis of the telomerase holoenzyme follows a stepwise RNP assembly. Upon transcription, nascent TERC is first bound by proteins dyskerin, NHP2 and NOP10 and subsequently joined by GAR1 (55). The co-transcriptional binding of this RNP complex is essential for the processing and accumulation of TERC transcripts. Finally, the catalytic protein component telomerase reverse transcriptase (TERT) assembles with the processed telomerase RNPs and forms the active telomerase holoenzyme. To determine whether RHAU was exclusively interacting with TERC during the transcriptional process or whether it was also part of the telomerase RNP, we performed co-IP assays and examined whether endogenous RHAU co-fractionated with endogenous dyskerin or FLAG-tagged TERT proteins. Immunoblot analysis showed that RHAU was present in both dyskerin and TERT immunoprecipitated fractions but absent in the control IP (Figure 5A and B, compare lanes 1 and 2). The association of RHAU with components of the telomerase holoenzyme proved to be strictly dependent on the ability of RHAU to bind the 5′-end of TERC, since overexpression of wild-type but not the G4 motif mutated forms of TERC resulted in extensive enrichment of co-precipitating RHAU protein (Figure 5A and B, compare lane 2 to lanes 3, 4 and 5). Although mutations of the TERC G4 motif sequence severely impinged on the recruitment of RHAU to the telomerase, they had no apparent repercussions on the steady-state expression level (Figure 5C) or on the assembly of the core components of telomerase. Indeed, as shown in Figure 5D (compare lane 2 to lanes 3, 4 and 5), similar levels of TERT protein co-fractionated with dyskerin

9398 Nucleic Acids Research, 2011, Vol. 39, No. 21

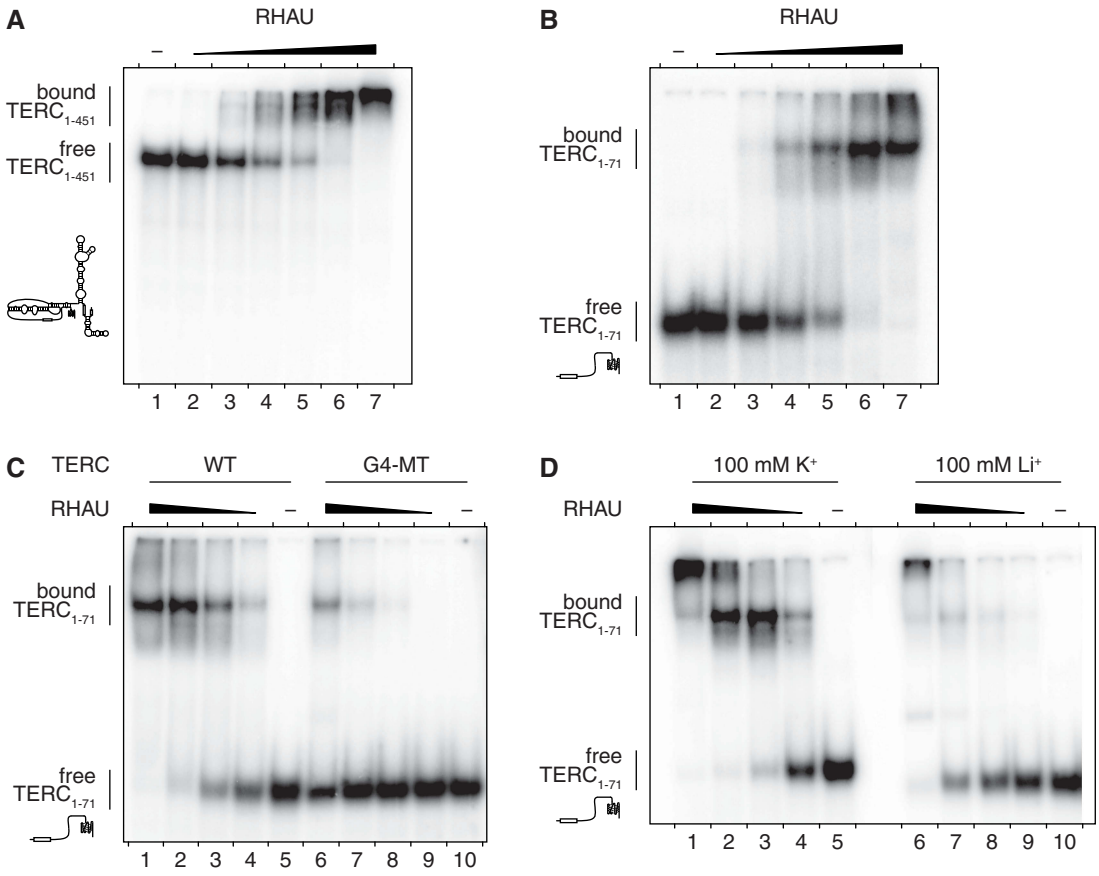


Figure 4. Gel mobility shift assay for TERC G4 binding by RHAU. (**A**) Radio-labeled full-length (1–451 nt) and (**B**) 5′-end (1–71 nt) TERC fragments at a concentration of 100 pM were incubated without protein (−) or with increasing amounts (1, 3.2, 10, 32, 100 and 320 nM) of GST-tagged RHAU in the absence of ATP. The reaction mixtures were analysed by non-denaturing PAGE. An autoradiogram of the gel is shown. The positions of the free RNA substrate and the protein–RNA complex are indicated on the left. At concentrations up to 320 nM, GST protein alone had no effect on TERC mobility (data not shown). (**C**) RNA-binding assay with WT and G4 motif mutant (G4-MT) forms of TERC 5′-end (1–71 nt) fragments. Radio-labeled TERC fragments were incubated without protein (−) or with increasing amounts (3.2, 10, 32 and 100 nM) of GST-tagged RHAU. (**D**) Cation dependency of RHAU interaction with TERC. The 5′-end WT TERC fragment was incubated under standard (100 mM K^+) or lithium-based (100 mM Li^+) REMSA conditions without protein (−) or with increasing amounts (10, 32 100 and 320 nM) of GST-tagged RHAU.

following transient overexpression of either wild-type or G4 motif mutated forms of TERC. These results suggest that RHAU does not strictly bind TERC in the course of its biogenesis but that a fraction of RHAU also associates with the fully assembled telomerase holoenzyme through direct interaction with the G4 motif sequence of TERC.

RHAU associates with telomerase activity

As RHAU binding to TERC requires a G4 structure, the finding that RHAU co-immunoprecipitated with components of the telomerase complex suggested the existence of an alternatively folded form of TERC bearing a G4 structure in a fraction of telomerase RNP. To address this further, immunopurified RHAU RNP complexes were analysed for telomerase activity by TRAP assay. In all the subsequent experiments, parallel immunopurification of dyskerin served as a positive control for telomerase activity (56). In agreement with previous observations, TRAP assays of antibody-bound complexes confirmed that substantial telomerase activity was recovered with RHAU but not with control IgGs (Figure 6A, top). Coomassie staining demonstrated similar yields of immunoprecipitated RHAU and dyskerin proteins (Figure 6A, bottom) and protein identity was subsequently verified by Western blotting (Figure 6B). To further examine the apparent significance of G4 binding by RHAU for its recruitment to the telomerase holoenzyme, we transiently expressed a FLAG-tagged mutant version of RHAU (ΔRSM, Supplementary Figure S3A) that was unable to bind to G4 structures (45). Western blot analysis confirmed that the wild-type and G4-binding-deficient (ΔRSM) forms of RHAU were expressed and recovered to similar levels (Supplementary Figure S3B). Immunopurified RHAU fractions were further assayed for telomerase activity and enrichment of TERC (Supplementary Figure S3C and D). As previously observed for endogenous RHAU, immunopurified FLAG-tagged RHAU(WT) efficiently recovered telomerase activity and co-precipitated TERC. In contrast,

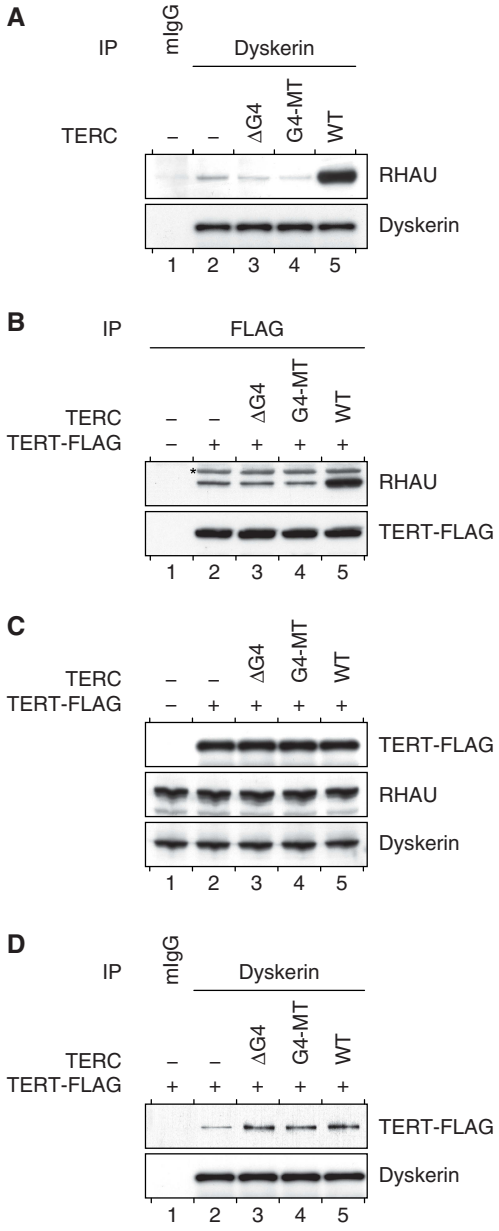


Figure 5. Association of RHAU with telomerase holoenzyme subunits. (**A**) TERC G4 motif dependent association of RHAU with dyskerin. Proteins from whole HEK293T cell lysates of either mock-transfected (−) cells or transiently expressing WT or G4 motif mutant (G4-MT, ΔG4) forms of TERC were immunoprecipitated with either control mouse IgGs (mIgG) or anti-dyskerin antibodies. Immunopurified RNP complexes were separated by SDS–PAGE and probed with anti-RHAU or anti-dyskerin antibodies. (**B**) TERC G4 motif-dependent association of RHAU with TERT. Protein immunoprecipitation experiments with anti-FLAG antibodies were performed with whole cells lysates of HEK293T cells transiently expressing TERT–FLAG protein together with WT or G4 motif mutant (G4-MT, ΔG4) forms of TERC. Immunopurified RNP complexes were separated by SDS–PAGE and probed with anti-RHAU or anti-FLAG antibodies. The asterisk denotes immunoprecipitated TERT–FLAG protein that cross-reacted unspecifically with the horseradish peroxidase-conjugated secondary antibody. (**C**) Western blot analysis of comparable protein expression in input HEK293T cell lysates. (**D**) Protein immunoprecipitation experiments with anti-dyskerin antibodies were performed with whole cells lysates of HEK293T cells transiently expressing TERT–FLAG protein together with WT or G4 motif mutant (G4-MT, ΔG4) forms of TERC. Immunopurified RNP complexes were separated by SDS–PAGE and probed with anti-FLAG or anti-dyskerin antibodies.

negligible telomerase activity and TERC signal were retrieved from the FLAG-tagged RHAU(ΔRSM) mutant, showing that G4 binding by RHAU is a prerequisite for RHAU binding to TERC. Altogether, these findings provide persuasive evidence that a fraction of RHAU associates with a subpopulation of the telomerase holoenzyme *in vivo* through direct interaction with the G4-motif sequence of TERC.

Finally, to determine the proportion of telomerase activity associated with RHAU, we carried out immunodepletion studies of RHAU in HEK293T cell lysates and subsequently quantified the residual TERC level and telomerase activity within these fractions (Figure 6C and D). Depletion with RHAU antibody significantly ablated telomerase activity in an aliquot of the supernatant by 25%, while ~70% of the telomerase activity of the input fraction was lost following depletion of endogenous dyskerin. As a control, immunodepletion with control mouse IgGs produced no significant reduction in telomerase activity. Western blot analysis of the supernatants confirmed that both RHAU and dyskerin proteins were effectively depleted from the corresponding fractions (Figure 6E). Furthermore, the quantification by RT–PCR of RNA levels in the immunodepleted fractions was consistent with the above findings. Indeed, approximately one quarter of input TERC, but not unrelated RNAs such as β-actin mRNA, vanished following depletion of endogenous RHAU. These results not only corroborate the association of RHAU with telomerase, but also show that the fraction of telomerase interacting with RHAU accounts for merely 25% of the total telomerase activity. Nevertheless, these data provide indirect but explicit evidence of the existence of an alternatively folded TERC structure bearing a G4 scaffold in the fully assembled telomerase holoenzyme.

ATPase-dependent interaction of RHAU with TERC

The fact that only one quarter of telomerase activity is linked to RHAU may reflect a dynamic interaction between RHAU and TERC. Indeed, most of the DEAH-box RNA helicases only transiently interact with nucleic acid, because they do not remain bound after the ATPase-dependent remodelling of their substrate (57–60). To gain insight into the role of ATP-hydrolysis by RHAU in its interaction with TERC, we repeated gel mobility shift assays with 5′ TERC fragments in the presence of Mg^{2+} or ATP or both. As shown in Figure 7A, ATP or Mg^{2+} alone had little effect on the binding affinity of RHAU for TERC. In contrast, under conditions supporting ATP hydrolysis (Mg•ATP), RHAU displayed reduced binding affinity for TERC, as evidenced by a 10-fold increase in the apparent equilibrium K_d value (Figure 7C). Such a reduction in the affinity of RHAU for TERC was not observed when Mg•ATP was substituted with the non-hydrolysable ATP analogue Mg•AMP-PNP (Figure 7B and C). Thus, the present data suggest that intrinsic RHAU ATP hydrolysis rather than ATP binding is essential for disrupting the interaction of RHAU with TERC. To further validate this, we turned to the previously described RHAU(DAIH) mutant in which the E335A amino-acid substitution within the Walker B site abolishes

9400 Nucleic Acids Research, 2011, Vol. 39, No. 21

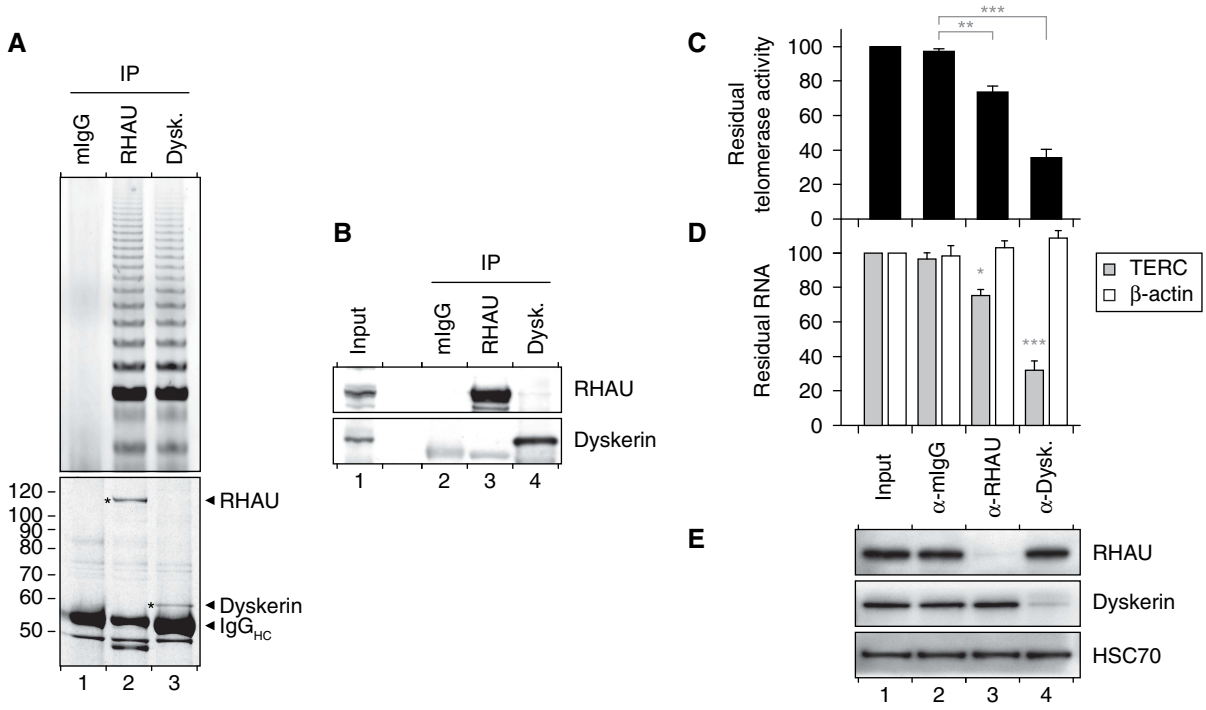


Figure 6. Association of RHAU with telomerase activity. (**A**) TRAP assay of immunopurified endogenous RHAU and dyskerin RNP complexes. RHAU or dyskerin RNP complexes were enriched by immunoprecipitation and a fraction of the immunopurified RNP was essayed for telomerase activity by the TRAP assay. Mouse IgGs (mIgGs) served as a control to assess non-specific interactions to the antibody matrix. A fraction of the immunopurified RNP was separated by SDS–PAGE. The Coomassie stained gel was scanned on a LI-COR Odyssey infrared imaging system. Positions of the immunoprecipitated proteins as well as the immunoglobulin heavy chains (IgG_{HC}) are indicated at the right. Positions and sizes (kDa) of marker proteins are shown at the left. (**B**) Western blot analysis of the immunopurified RNP complexes assayed for telomerase activity. (**C**) qTRAP analysis of the residual telomerase activity present in immunodepleted HEK293T cell lysates. Data represent the mean ± SEM of three independent experiments. Statistical significance was determined by one-way ANOVA and the Bonferroni *t*-test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. (**D**) RT–qPCR analysis of the residual levels of TERC and β-actin RNAs in immunodepleted HEK293T cell lysates. The RNA signal was normalized to GAPDH. Data represent the mean ± SEM of three independent experiments. Statistical significance was determined by one-way ANOVA and the Bonferroni *t*-test. (**E**) Western blot analysis of input and immunodepleted HEK293T cell lysates. Heat shock protein 70 cognate (HSC70) was used as a loading control.

RHAU ATPase activity (45). Unlike wild-type RHAU, the binding affinity of RHAU(DAIH) mutant for TERC was similar in the absence and presence of Mg•ATP (Figure 7B and C). The absence of RHAU helicase activity as a consequence of the loss of ATPase activity probably prevents RHAU dissociation from its substrate. In fact, a similar reduction in *in vivo* RNA binding dynamics after suppression of RHAU intrinsic ATPase activity was reported previously in studies of the shuttling of RHAU in cytoplasmic stress granules (43). Taken together, these data corroborate the idea that RHAU is not permanently associated with TERC but dissociates upon ATP hydrolysis.

Apart from a high affinity for G4 structures, RHAU has also been shown to couple ATP hydrolysis with tetramolecular G4-RNA resolving activity (42,45). These biochemical properties prompted us to test whether RHAU could catalyse the conversion of TERC G4 structure to ssRNA form. However, the faster rate of refolding of intramolecular G4 structures with regard to tetramolecular G4-RNAs made it technically difficult to monitor the G4-resolving activity of RHAU using standard non-denaturing gel electrophoresis.

DISCUSSION

The propensity of nucleic acid guanine-rich sequences to self-assemble into G4 structures *in vitro* has been recognized for several decades. Although RNA is also prone to form such structures, G4-RNA has not attracted as much attention as G4-DNA. Nevertheless, a growing body of evidence indicates a significant role for G4 structure formation during RNA metabolism (20,61). As a consequence, G4 structures emerged as a plausible post-transcriptional means of regulating the function of coding and non-coding RNAs. However, little is known about the mechanisms by which G4-RNA formation is regulated in the cells. Indeed, only a few proteins have been reported so far to interact with G4-RNAs *in vitro* (62–66) and, apart from FMRP and FMR2P, experimental data on their biological activity subsequent to their interaction with G4-RNAs is scarce. Of these G4-RNA binding proteins, RHAU is the only protein that exhibits robust *in vitro* ATPase-dependent G4-RNA resolving activity, in addition to a high affinity and specificity for its target RNAs (42). These particular biochemical features promote RHAU as an ideal candidate for regulating G4-dependent RNA metabolism, although its *in vivo* RNA

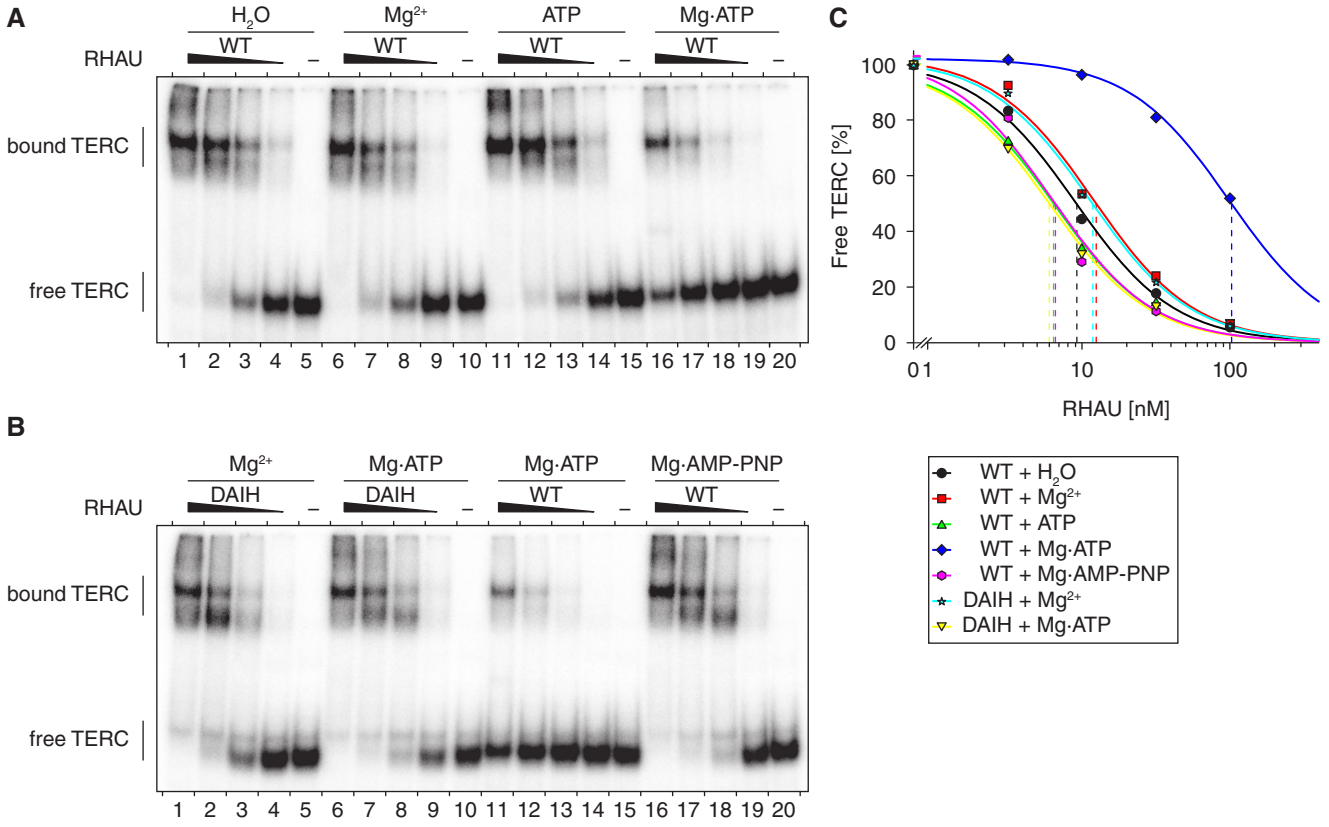


Figure 7. Role of Mg•ATP on the interaction of RHAU with TERC. (**A**) Gel mobility shift assay for TERC G4 binding by RHAU. Radio-labeled 5′-end (1–71 nt) TERC fragment at a concentration of 100 pM was incubated without protein (–) or with increasing amounts (3.2, 10, 32 and 100 nM) of purified recombinant GST-tagged RHAU in the presence of 1 mM Mg^{2+}, ATP or Mg•ATP (as indicated). The reaction mixtures were analysed by non-denaturing PAGE. An autoradiogram of the gel is shown. (**B**) Gel mobility shift assay of TERC G4 binding by the ATPase deficient RHAU(DAIH) mutant. Radio-labeled 5′-end TERC fragment was incubated without protein (–) or with increasing amounts (3.2, 10, 32 and 100 nM) of purified recombinant WT or DAIH mutant GST-tagged RHAU in the presence of 1 mM Mg^{2+}, Mg•ATP or Mg•AMP-PNP (as indicated). (**C**) Quantification of gel mobility shift assays of WT and RHAU(DAIH) mutant binding to TERC 5′-end (1–71 nt) fragment. The data represent the mean of three independent experiments. Error bars for SEM were omitted for clarity.

targets have not yet been determined. The present study aimed to identify naturally occurring RHAU targets with the emphasis on RNAs containing potential G4-forming sequences. By RIP-chip assays, 106 RNAs were found to be significantly enriched with RHAU. Importantly, more than half of these RNAs contained G-rich sequences with potential to form stable G4 structures. In addition, there was a weak but significant correlation between the predicted G4 motif density per transcript and the magnitude of RNA enrichment by RHAU. Nevertheless, because RHAU appears to show significant affinity for RNAs not bearing G4-forming sequences, we do not exclude the possibility that RHAU has affinity to other RNA structural features.

In-depth studies revealed that TERC, one of the identified RNAs, was a *bona fide* target of RHAU. Several independent RIP-chip assays using various cell lines further showed TERC to be one of the most abundant RNAs enriched by RHAU. TERC bears a 5′ G-rich sequence that was previously shown to adopt a stable intramolecular G4 structure *in vitro* (30). Further investigations dissecting the basis of the interaction between RHAU and TERC showed RHAU binding to be strictly dependent on G4 structure formation in the 5′ region of TERC and to require the N-terminal RSM domain of RHAU, an ancillary domain necessary for specific recognition of G4 by RHAU (45). Moreover, RHAU was found to interact transiently with TERC in a manner dependent on its own ATPase activity. Together, these data provide the first evidence of a specific and direct interaction between a G4-resolvase enzyme and a potentially relevant intramolecular G4-RNA substrate. Furthermore, in agreement with our previous reports (42,45), these data attest that G4-RNAs are naturally occurring substrates of RHAU *in vivo*.

The 5′ extremity of TERC folds into an intramolecular G4 structure *in vivo*

In agreement with the observations presented here, Mergny and co-workers reported previously the formation *in vitro* of a G4 structure in the 5′ region of human TERC (30). However, as for the great majority of nucleic acid structures, direct experimental demonstration of G4 structures *in vivo* has proven very difficult. As such, our results

9402 Nucleic Acids Research, 2011, Vol. 39, No. 21

do not directly prove the existence of G4 structures *in vivo*, but strongly support the notion that TERC can form a G4 structure in the cells. Notably, RHAU can only bind TERC provided that all conditions necessary for the formation of an intramolecular G4 structure (RNA sequence and ionic conditions) are met. Modifying any one of the conditions is sufficient to reduce substantially the affinity of RHAU for TERC. In addition, our data argue that TERC G4 structure can also be formed in a fraction of telomerase holoenzyme, since RHAU associates with TERT and dyskerin proteins and significant amounts of telomerase activity were recovered with RHAU. Based on the fraction of telomerase activity co-precipitating with RHAU, we can estimate that at least 25% of TERC in the human telomerase holoenzyme bears a 5′ G4 structure. Moreover, we predict that formation of a G4 structure in TERC is not exclusive to human cells, since the majority of the mammalian orthologues of TERC also harbours a potential 5′ G4-forming sequence (Supplementary Figure S4). Insofar as the G4-RNA binding and resolving activities of RHAU are conserved in higher eukaryotes (45), it would not be surprising to find that RHAU is part of the telomerase holoenzymes of other mammals.

As reported previously (30), folding of the 5′ region of TERC into a G4 structure is likely to interfere with the widely adopted secondary structure of human TERC (Supplementary Figure S5). Indeed, according to the standard model of TERC secondary structure (67), two of the guanine tracts that are part of the intramolecular G4 also form the P1a helix. As a consequence, the P1a helix and G4 represent two mutually exclusive structures and may correspond to different functional states of TERC. Insofar as helix P1 is required for template boundary definition in mammalian telomerase (68), formation of a G4 structure in TERC may be detrimental to telomerase activity. On the other hand, G4 structure formation may protect TERC from degradation during telomerase biogenesis. Such a scenario was suggested quite recently by Collins and co-workers, who found that nucleotide substitutions within the G4-forming sequence markedly reduced accumulation of the mature form of TERC, thereby causing telomere shortening (69). Similar to our data, they also found that RHAU associates with the G4-forming sequence of TERC *in vivo*, but in contrast to our finding, detected negligible telomerase activity with immunopurified RHAU fraction. Thus, further experiments are necessary to clarify the functional impact of G4 formation in TERC on telomerase biogenesis and activity.

In cells, the P1 duplex- and G4-folded forms of TERC are likely to coexist in dynamic equilibrium. However, under physiological conditions, restoration of helix P1 base pairing from the G4-folded conformation of TERC is likely to require a catalyst due to the high thermodynamic stability of the quadruplex (30). The biochemical properties and substrate specificity of RHAU make it a likely candidate for catalysing such a reaction. Quite apart from its high affinity for TERC, RHAU manifests robust tetramolecular G4-RNA resolving activity (42,45) and was also shown to unwind various intramolecular G4-DNA structures *in vitro* (70). These data suggest that RHAU can resolve G4 structures irrespective of the strand stoichiometry of the G4 stem. Although unwinding of TERC G4 by RHAU has not yet been demonstrated experimentally, our observation that RHAU dissociates from TERC upon ATP-hydrolysis, but not upon ATP binding, supports the idea that RHAU couples ATP-hydrolysis to conformational changes of TERC, resulting in the resolution of the G4 and further release of RHAU from its target RNA.

RHAU may target other RNAs containing G4 structures

Apart from TERC, 54 further RNAs found to be associated with RHAU are predicted to form stable G4 structures (Supplementary Table S2). The determination of the modalities and the significance of these interactions are essential issues to be addressed in the future. Nonetheless, the finding that RHAU associates preferentially with transcripts bearing potential G4-forming sequences strongly suggests that RHAU targets G4-RNA structures in cells. Although nearly three-quarters of these G4-forming sequences are located in 5′ or 3′ UTRs of mRNAs (Supplementary Table S2), it is important to stress that, to date, we have not observed any suppressive action of RHAU on translational repression by G4 formation in 5′-UTRs. Therefore, RHAU is unlikely to function as a translational activator but may instead intervene in other aspects of mRNA metabolism, such as pre-mRNA processing or mRNA turnover. Indeed, RHAU localizes predominantly in the nucleus, where it concentrates in nuclear speckles (71). These are sites of high transcriptional activity and mRNA splicing. Furthermore, RHAU relocalizes to cytoplasmic stress granules in response to various cellular stresses (43). Although recruitment of RHAU to stress granules is mediated by interaction with RNA, we have not yet examined whether this phenomenon depends upon binding of RHAU to G4-RNA structures. However, considering that RSM-domain mutated forms of RHAU deficient in G4-RNA binding also show reduced relocalization to stress granules, it is likely that a fraction of RHAU binds G4 structures in stress granules.

Together with our previous findings that RHAU binds and exhibits robust resolvase activity on various types of G4 structures (41,42,70), the data presented here bring forth the idea that G4-RNAs and especially intramolecular G4-RNAs serve as physiologically relevant targets for RHAU. Identification of naturally occurring substrates of RNA helicases is a prerequisite to any further investigation of their biochemical properties *in vitro*. However, although most of the RNA helicases achieve highly specific tasks *in vivo*, they often show only little or even no nucleic acid binding specificity *in vitro*. Currently, our major limitation towards understanding the mechanisms, whereby RNA helicases melt nucleic acid structures stem from the difficulty of identifying such physiologically relevant substrates. Hence, few detailed molecular models available for studying RNA helicases *in vitro*. However, with the finding that the 5′ region of TERC constitutes a biologically relevant substrate, RHAU emerges as a novel and

promising prototype of DEAH-box protein and deserves more investigations to explore its functions as a G4 resolving enzyme.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Stéphane Thiry for excellent technical assistance and Susan Gasser, Patrick King, Janice Ching Lai and Sandra Pauli for critical comments on the article.

FUNDING

Funding for open access charge: Novartis Research Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. König,S., Evans,A. and Huppert,J. (2010) Seven essential questions on G-quadruplexes. *BioMolecular Concepts*, **1**, 197–213.
2. Lane,A.N., Chaires,J.B., Gray,R.D. and Trent,J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
3. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
4. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
5. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
6. Hershman,S.G., Chen,Q., Lee,J.Y., Kozak,M.L., Yue,P., Wang,L.S. and Johnson,F.B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **36**, 144–156.
7. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
8. Huppert,J.L., Bugaut,A., Kumari,S. and Balasubramanian,S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.
9. Eddy,J. and Maizels,N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.*, **36**, 1321–1333.
10. Chang,C.C., Chu,J.F., Kao,F.J., Chiu,Y.C., Lou,P.J., Chen,H.C. and Chang,T.C. (2006) Verification of antiparallel G-quadruplex structure in human telomeres by using two-photon excitation fluorescence lifetime imaging microscopy of the 3,6-Bis(1-methyl-4-vinylpyridinium)carbazole diiodide molecule. *Anal. Chem.*, **78**, 2810–2815.
11. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Pluckthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
12. Paeschke,K., Simonsson,T., Postberg,J., Rhodes,D. and Lipps,H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
13. Grand,C.L., Han,H., Munoz,R.M., Weitman,S., Von Hoff,D.D., Hurley,L.H. and Bearss,D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.*, **1**, 565–573.
14. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
15. Fernando,H., Rodriguez,R. and Balasubramanian,S. (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry*, **47**, 9365–9371.
16. Fernando,H., Sewitz,S., Darot,J., Tavare,S., Huppert,J.L. and Balasubramanian,S. (2009) Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Res.*, **37**, 6716–6722.
17. Joachimi,A., Benz,A. and Hartig,J.S. (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg. Med. Chem.*, **17**, 6811–6815.
18. Arora,A. and Maiti,S. (2009) Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *J. Phys. Chem. B*, **113**, 10515–10520.
19. Mergny,J.L., De Cian,A., Ghelab,A., Sacca,B. and Lacroix,L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
20. Beaudoin,J.D. and Perreault,J.P. (2010) 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.*, **38**, 7022–7036.
21. Arora,A., Dutkiewicz,M., Scaria,V., Hariharan,M., Maiti,S. and Kurreck,J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA*, **14**, 1290–1296.
22. Morris,M.J. and Basu,S. (2009) An unusually stable G-quadruplex within the 5'-UTR of the MT3 matrix metalloproteinase mRNA represses translation in eukaryotic cells. *Biochemistry*, **48**, 5313–5319.
23. Gomez,D., Guedin,A., Mergny,J.L., Salles,B., Riou,J.F., Teulade-Fichou,M.P. and Calsou,P. (2010) A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res.*, **38**, 7187–7198.
24. Gomez,D., Lemarteleur,T., Lacroix,L., Mailliet,P., Mergny,J.L. and Riou,J.F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.*, **32**, 371–379.
25. Kostadinov,R., Malhotra,N., Viotti,M., Shine,R., D'Antonio,L. and Bagga,P. (2006) GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.*, **34**, D119–D124.
26. Didiot,M.C., Tian,Z., Schaeffer,C., Subramanian,M., Mandel,J.L. and Moine,H. (2008) The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.*, **36**, 4902–4912.
27. Randall,A. and Griffith,J.D. (2009) Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets. *J. Biol. Chem.*, **284**, 13980–13986.
28. Martadinata,H. and Phan,A.T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. *J. Am. Chem. Soc.*, **131**, 2570–2578.
29. Xu,Y., Kaminaga,K. and Komiyama,M. (2008) G-quadruplex formation by human telomeric repeats-containing RNA in Na+ solution. *J. Am. Chem. Soc.*, **130**, 11179–11184.
30. Gros,J., Guedin,A., Mergny,J.L. and Lacroix,L. (2008) G-Quadruplex formation interferes with P1 helix formation in the RNA component of telomerase hTERC. *Chembiochem*, **9**, 2075–2079.
31. Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front Biosci.*, **12**, 4336–4351.
32. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.
33. Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **273**, 27587–27592.
34. Fry,M. and Loeb,L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **274**, 12797–12802.

35. Wu,Y., Shin-ya,K. and Brosh,R.M. Jr. (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.*, **28**, 4116–4128.
36. Sanders,C.M. (2010) Human Pif1 helicase is a G-quadruplex DNA binding protein with G-quadruplex DNA unwinding activity. *Biochem. J.*, **430**, 119–128.
37. Chu,W.K. and Hickson,I.D. (2009) RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer*, **9**, 644–654.
38. Wu,Y., Suhasini,A.N. and Brosh,R.M. Jr. (2009) Welcome the family of FANCJ-like helicases to the block of genome stability maintenance proteins. *Cell Mol. Life Sci.*, **66**, 1209–1222.
39. Wang,W. (2007) Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat. Rev. Genet.*, **8**, 735–748.
40. Bochman,M.L., Sabouri,N. and Zakian,V.A. (2010) Unwinding the functions of the Pif1 family helicases. *DNA Repair*, **9**, 237–249.
41. Vaughn,J.P., Creacy,S.D., Routh,E.D., Joyner-Butt,C., Jenkins,G.S., Pauli,S., Nagamine,Y. and Akman,S.A. (2005) The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **280**, 38117–38120.
42. Creacy,S.D., Routh,E.D., Iwamoto,F., Nagamine,Y., Akman,S.A. and Vaughn,J.P. (2008) G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.*, **283**, 34626–34634.
43. Chalupnikova,K., Lattmann,S., Selak,N., Iwamoto,F., Fujiki,Y. and Nagamine,Y. (2008) Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. *J. Biol. Chem.*, **283**, 35186–35198.
44. Tran,H., Schilling,M., Wirbelauer,C., Hess,D. and Nagamine,Y. (2004) Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol. Cell.*, **13**, 101–111.
45. Lattmann,S., Giri,B., Vaughn,J.P., Akman,S.A. and Nagamine,Y. (2010) Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.*, **38**, 6219–6233.
46. Zheng,L., Baumann,U. and Reymond,J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
47. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
48. Carvalho,B.S. and Irizarry,R.A. (2010) A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, **26**, 2363–2367.
49. Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.
50. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.
51. Altschul,S.F. and Erickson,B.W. (1985) Significance of nucleotide sequence alignments: a method for random sequence permutation that preserves dinucleotide and codon usage. *Mol. Biol. Evol.*, **2**, 526–538.
52. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
53. Herbert,B.S., Hochreiter,A.E., Wright,W.E. and Shay,J.W. (2006) Nonradioactive detection of telomerase activity using the telomeric repeat amplification protocol. *Nat. Protoc.*, **1**, 1583–1590.
54. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
55. Collins,K. (2008) Physiological assembly and activity of human telomerase complexes. *Mech. Ageing Dev.*, **129**, 91–98.
56. Mitchell,J.R., Wood,E. and Collins,K. (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*, **402**, 551–555.
57. Schwer,B. and Guthrie,C. (1991) PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, **349**, 494–499.
58. Teigelkamp,S., McGarvey,M., Plumpton,M. and Beggs,J.D. (1994) The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J.*, **13**, 888–897.
59. Schwer,B. and Gross,C.H. (1998) Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.*, **17**, 2086–2094.
60. Bohnsack,M.T., Martin,R., Granneman,S., Ruprecht,M., Schleiff,E. and Tollervey,D. (2009) Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol. Cell.*, **36**, 583–592.
61. Halder,K., Wieland,M. and Hartig,J.S. (2009) Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes. *Nucleic Acids Res.*, **37**, 6811–6817.
62. Bashkirov,V.I., Scherthan,H., Solinger,J.A., Buerstedde,J.M. and Heyer,W.D. (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell Biol.*, **136**, 761–773.
63. Darnell,J.C., Jensen,K.B., Jin,P., Brown,V., Warren,S.T. and Darnell,R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*, **107**, 489–499.
64. Khateb,S., Weisman-Shomer,P., Hershco,I., Loeb,L.A. and Fry,M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)n is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
65. Davidovic,L., Bechara,E., Gravel,M., Jaglin,X.H., Tremblay,S., Sik,A., Bardoni,B. and Khandjian,E.W. (2006) The nuclear microspherule protein 58 is a novel RNA-binding protein that interacts with fragile X mental retardation protein in polyribosomal mRNPs from neurons. *Hum. Mol. Genet.*, **15**, 1525–1538.
66. Bensaid,M., Melko,M., Bechara,E.G., Davidovic,L., Berretta,A., Catania,M.V., Gecz,J., Lalli,E. and Bardoni,B. (2009) FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res.*, **37**, 1269–1279.
67. Chen,J.L., Blasco,M.A. and Greider,C.W. (2000) Secondary structure of vertebrate telomerase RNA. *Cell*, **100**, 503–514.
68. Chen,J.L. and Greider,C.W. (2003) Template boundary definition in mammalian telomerase. *Genes Dev.*, **17**, 2747–2752.
69. Sexton,A.N. and Collins,K. (2011) The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. *Mol. Cell. Biol.*, **31**, 736–743.
70. Giri,B., Smaldino,P.J., Thys,R.G., Creacy,S.D., Routh,E.D., Hantgan,R.R., Lattmann,S., Nagamine,Y., Akman,S.A. and Vaughn,J.P. (2011) G4 Resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Res.*, doi:10.1093/nar/gkr234.
71. Iwamoto,F., Stadler,M., Chalupnikova,K., Oakeley,E. and Nagamine,Y. (2008) Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. *Exp. Cell Res.*, **314**, 1378–1391.

CHAPTER 9

Appendix

Simon Lattmann

Year of birth: 1982
Nationality: Swiss
Marital status: Single
Langage proficiency: French (native)
English (fluent)
German (good knowledge)

Rochettes 39
1454 La Vraconnaz
Switzerland

Simon.Lattmann@fmi.ch
+41 24 454 15 69
+41 77 437 15 05

Education

2006–2011 Ph.D. in Molecular Biology and Biochemistry
Friedrich Miescher Institute
University of Basel, Switzerland

2000–2005 B.Sc. and M.Sc. in Molecular and Cellular Biology
University of Lausanne, Switzerland

1997–2000 Type-C (scientific) Swiss Federal Maturity
High school Gymnase d'Yverdon-les-Bains, Switzerland

Research Experience

2006–2011 Ph.D. student
Laboratory of Dr. Yoshikuni Nagamine
Friedrich Miescher Institute
Basel, Switzerland

2005 (6 mo.) Research associate
Laboratory of Prof. Hans Acha Orbea
Department of Biochemistry
University of Lausanne, Switzerland

2004–2005 M.Sc. student
Laboratory of Prof. Andreas Mayer
Department of Biochemistry
University of Lausanne, Switzerland

2003 (4 mo.) Undergraduated trainee
Laboratory of Prof. Jean-Daniel Horisberger†
Department of Pharmacology and Toxycology
University of Lausanne, Switzerland

Publications

- Lattmann, S., Stadler, M.B., Vaughn, J.P., Akman, S.A. and Nagamine, Y. (2011) **The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme**. *Nucleic Acids Res.*, **39**, 9390–404. PubMed CrossRef
- Giri, B., Smaldino, P.J., Thys, R.G., Creacy, S.D., Routh E.D., Hantgan, R.R., Lattmann, S., Nagamine, Y., Akman, S.A. and Vaughn, J.P. (2011) **G4R1/RHAU tightly binds and unwinds unimolecular G4-DNA**. *Nucleic Acids Res.*, **39**, 7161–78. PubMed CrossRef
- Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A. and Nagamine, Y. (2010) **Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU**. *Nucleic Acids Res.*, **38**, 6219–33. PubMed CrossRef
- Chalupnikova, K., Lattmann, S., Selak, N., Iwamoto, F., Fujiki, Y. and Nagamine, Y. (2008) **Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain**. *J. Biol. Chem.*, **283**, 35186–98. PubMed CrossRef

Technical Skills

- *Model organisms and cell lines*
  Various types of mammalian and insect cell lines | *S. cerevisiae* | *Xenopus* oocytes | Mice (basic knowledge).
- *Technics*
  Molecular biology (cloning) | Standard biochemistry methods | Expression, purification and analysis of recombinant proteins | qPCR | RIP-chip | (R)EMSA | Radio-isotopic labelling of nucl. ac. and proteins | FACS.
- *Computing*
  Most common bioinformatics tools | PyMOL | Programming/scripting languages: basic knowledge in Java, Python, R and HTML | Productivity suites: Microsoft Office, Adobe Creative Suite.

International Conferences

- *Poster presentation at*
  the EMBO Conference Series 2010
  Telomeres and the DNA Damage Response
  Marseille, France
- *Poster presentation at*
  the Swiss RNA Workshop 2010
  Bern, Switzerland
- *Poster presentation at*
  the EMBO Conference Series 2009
  Helicase and NTP-Driven Nucleic Acid Motors: Structure, Function, Mechanism and Roles in Human Diseases
  Les Diablerets, Switzerland
- *Poster presentation at*
  the FASEB Summer Research Conference 2007
  Helicase and NTP-Driven Nucleic Acid Motors: Structure, Function, Mechanism and Roles in Human Diseases
  Indian Wells, CA
- *Poster presentation at*
  the Novartis Biomedical Research Conference 2006
  Venice, Italy

Extracurricular Activities

- 2003–2010: Member of the Lausanne section committee of the Swiss Academic Ski (SAS).
- Hobbies: sports (cross-country skiing, running, cycling) | Computing | Photography | Travelling.