Review

Katherine E. Bohnsack, Ralf Ficner, Markus T. Bohnsack* and Stefanie Jonas*

Regulation of DEAH-box RNA helicases by G-patch proteins

<https://doi.org/10.1515/hsz-2020-0338>Received October 13, 2020; accepted December 9, 2020;
published online January 6, 2021

Abstract: RNA helicases of the DEAH/RHA family form a large and conserved class of enzymes that remodel RNA protein complexes (RNPs) by translocating along the RNA. Driven by ATP hydrolysis, they exert force to dissociate hybridized RNAs, dislocate bound proteins or unwind secondary structure elements in RNAs. The sub-cellular localization of DEAH-helicases and their concomitant association with different pathways in RNA metabolism, such as pre-mRNA splicing or ribosome biogenesis, can be guided by cofactor proteins that specifically recruit and simultaneously activate them. Here we review the mode of action of a large class of DEAH-specific adaptor proteins of the G-patch family. Defined only by their eponymous short glycine-rich motif, which is sufficient for helicase binding and stimulation, this family encompasses an immensely varied array of domain compositions and is linked to an equally diverse set of functions. G-patch proteins are conserved throughout eukaryotes and are even encoded within retroviruses. They are involved in mRNA, rRNA and snoRNA maturation, telomere maintenance and the innate

immune response. Only recently was the structural and mechanistic basis for their helicase enhancing activity determined. We summarize the molecular and functional details of G-patch-mediated helicase regulation in their associated pathways and their involvement in human diseases.

Keywords: cofactor; DEAH/RHA family; G-patch protein; pre-mRNA splicing; ribosome biogenesis; RNA helicase.

Introduction

RNA helicases of the helicase superfamily 1 and 2 (SF1 and SF2) contribute to diverse aspects of RNA metabolism through their functions in structurally remodelling RNAs and ribonucleoprotein complexes (RNPs) (reviewed in Jarmoskaite and Russell (2014)). These nucleotide triphosphate (NTP)-dependent enzymes are characterised by a common core composed of tandem RecA-like (RecA1 and RecA2) domains that harbour conserved sequence motifs involved in RNA substrate binding, and NTP binding and hydrolysis (Caruthers and McKay 2002). Within SF2, DEXD-box and DEAH/RHA proteins form two highly abundant families of RNA helicases that are closely related, but display differences in their respective mode of RNA remodelling. While enzymes of the DEXD-box family typically bind to double-stranded (ds) RNA substrates and induce local strand unwinding, DEAH/RHA helicases translocate along RNA strands with a 3′-5′ directionality, inducing duplex unwinding in a potentially processive manner (Hamann et al. 2019; Mallam et al. 2012; Sengoku et al. 2006; Tauchert et al. 2017; Yang and Jankowsky 2006; Yang et al. 2007). However, additional molecular functions, such as RNA strand annealing and RNA clamping have also been attributed to specific RNA helicases (e.g. Ballut et al. 2005; Fairman et al. 2004; Rossler et al. 2001; Yang and Jankowsky 2005). Mechanistically, the RecA domains of both types of RNA helicase exist in open (inactive) and closed (active) conformations. For DEXD-box RNA helicases, formation of a closed state upon ATP and RNA substrate binding forces the RNA into a sharp bend via

***Corresponding authors: Markus T. Bohnsack**, Department of Molecular Biology, University Medical Center Göttingen, Humboldtallee 23, D-37073 Göttingen, Germany; and Göttingen Centre for Molecular Biosciences, Georg-August University, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany, E-mail: markus.bohnsack@med.uni-goettingen.de. <https://orcid.org/0000-0001-7063-5456>; and **Stefanie Jonas**, Department of Biology, Institute of Molecular Biology and Biophysics, ETH Zurich, Otto-Stern-Weg 5, CH-8093 Zurich, Switzerland, E-mail: stefanie.jonas@mol.biol.ethz.ch. <https://orcid.org/0000-0002-8751-6741>

Katherine E. Bohnsack, Department of Molecular Biology, University Medical Center Göttingen, Humboldtallee 23, D-37073 Göttingen, Germany, E-mail: katherine.bohnsack@med.uni-goettingen.de
Ralf Ficner, Department of Molecular Structural Biology, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany; and Göttingen Centre for Molecular Biosciences, Georg-August University, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany, E-mail: rficner@uni-goettingen.de

the protrusion of an α -helix within RecA1. This kink is incompatible with a double-stranded RNA conformation, thereby leading to base-pair melting. Subsequent ATP hydrolysis leads, in turn, to disassembly of the helicase-RNA complex. In contrast, conserved “Hook-turn” and “Hook-loop” motifs in the RecA1 and RecA2 domains of DEAH/RHA helicases contact the RNA substrate such that the transition of the RecA2 domain to the open conformation upon NTP hydrolysis accommodates an additional nucleotide within the helicase RNA binding channel, meaning that cycles of NTP-dependent opening and closing are coupled to movement of the RNA substrate through the helicase core in one nucleotide steps.

As RNA helicases play central roles in most aspects of gene expression, dysregulation of RNA helicase activity is often associated with tumorigenesis and disease (Steimer and Klostermeier 2012), and careful regulation of RNA helicase activity in cells is essential. While DExD and DEAH/RHA helicases are inactive in their open conformations, their inherent capacity for transition to active closed conformations upon formation of non-specific interactions with the backbone of RNA substrates (Andersen et al. 2006; Sengoku et al. 2006), necessitates dedicated strategies for limiting the promiscuity of RNA helicases and ensuring their target specificity. The activity of RNA helicases can be regulated by a variety of different means (Sloan and Bohnsack 2018). Some RNA helicases possess auto-inhibitory domains that maintain low activity in the native state until binding to a correct substrate triggers a rearrangement compatible with catalysis (Absmeier et al. 2020; Chakrabarti et al. 2011; Gowravaram et al. 2018). Others possess additional domains that recognise specific RNA features (e.g. RNA modifications) thereby directing the helicases to appropriate RNA substrates (Kretschmer et al. 2018; Luo et al. 2011; Wojtas et al. 2017). Furthermore, the catalytic activity of some RNA helicases has been shown to be regulated by post-translation modifications, enabling them to be specifically activated in response to particular conditions or upon association with target complexes (Jacobs et al. 2007; Mathew et al. 2008; Song et al. 2017). However, the predominant mechanism of RNA helicase regulation is through interaction with cofactor proteins. Such proteins can bind and regulate the catalytic activity of their cognate helicases and/or facilitate their recruitment to target RNAs/RNPs either by creating an electrostatic environment that promotes interaction with negatively charged RNA (e.g. Btz complex-mediated regulation of eIF4AIII within the exon-junction complex (Bono et al. 2006)) or by mediating direct RNA or RNA-binding protein interactions. While many RNA helicases rely on dedicated cofactor proteins whose mode of interaction

with the helicase is protein-specific (for example, Dbp8 and Esf2, DHX9 and EWS-FLI, Upf1 and Upf2, and Rok1 and Rps5 (Chakrabarti et al. 2011; Erkizan et al. 2015; Granneman et al. 2006; Young et al. 2013)), other cofactor proteins contain characteristic domains through which they interact with and regulate the activity of specific classes of RNA helicases. Several proteins containing a middle of eIF4G (MIF4G) domains regulate eIF4A-like RNA helicases (see for example, Alexandrov et al. 2011; Davila Gallesio et al. 2020; Mugler et al. 2016; Schütz et al. 2008). However, in this review, we focus on the G-patch proteins as another family of RNA helicase cofactors. We will discuss their characteristics, their interactions with RNA helicases and mode of helicase regulation as well as describing the current knowledge on their cellular functions.

Overview of G-patch protein family

G-patch proteins are characterised by a glycine-rich motif of approximately 50 amino acids, which was identified in a large number of highly diverse proteins (Figure 1, Table 1) in a computational study purely based on sequence conservation (Aravind and Koonin 1999). The consensus motif $Gx_2hhx_3Gax_2GxGlGx_3pxux_3sx_{10-16}GhG$ (a – aromatic, h – hydrophobic, l – aliphatic, s – small, u – tiny, x – variable amino acid) contains seven highly conserved glycine residues, an invariable aromatic amino acid following the second glycine and three defined hydrophobic patches (Figure 2). The G-patch motif is widespread in proteins from eukaryotes but is absent from bacteria and archaea. It is also encoded in numerous betaretroviruses and several endogenous retroviral elements, amongst them the human endogenous retrovirus K. However, in most cases, it is likely that these endogenous retroviruses remain repressed (see below for further details (Garcia-Montojo et al. 2018; Hanke et al. 2016)). In the budding yeast *Saccharomyces cerevisiae* (Sc), five G-patch proteins are annotated (Table 1) and while these proteins are conserved in higher eukaryotes, the inventory of human (Hs) G-patch proteins far exceeds that of yeast with more than 20 members. The increased number of human G-patch proteins likely reflects the increased need for regulation of gene expression in higher eukaryotes. Several G-patch proteins have been shown to interact with a subset of helicases of the DEAH/RHA family (Table 1). In yeast, four of the five G-patch proteins (Cmg1, Pxr1, Sqs1, Spp382) bind to the multifunctional helicase Prp43, while only one (Spp2) associates with Prp2. In human cells, a similar picture is emerging, with most of the characterized

interactions (NKRF, PINX1, RBM5, RBM17, TFIP11, ZGPAT) focused on DHX15 (human Prp43 ortholog) and only one G-patch partner (GPKOW) for DHX16 (human Prp2 ortholog). Nevertheless, for 14 G-patch proteins, a corresponding helicase partner has not been characterized and there are in total 15 human DEAH/RHA helicases (Fairman-Williams et al. 2010), which could all in principle be interaction partners.

Proteins containing a G-patch motif vary greatly in size (21–264 kDa) and domain composition (Figure 1). However,

a commonality is that they all harbour only a single G-patch motif and that this is embedded in an intrinsically disordered region. Another signature of G-patch proteins is the high prevalence of RNA binding motifs (e.g. KOW (Kyprides, Ouzounis, Woese motif), RG/RGG (arginine-glycine) or SR (serine-arginine) repeats, SURP (suppressor-of-white-apricot and PRP21/SPP91) and RNA binding domains (R3H (arginine-3 histidine motif), dsRBD (double-stranded RNA binding domain), RRM (RNA recognition motif), Zinc fingers), in keeping with their general

Figure 1: Domain organisation of G-patch proteins.

Domains, motifs and sequence repeats present in the yeast and human G-patch proteins are shown schematically. Proteins and their components are drawn to scale except where indicated (// = 650 amino acids) and are annotated as: dsRBD – double-stranded RNA binding domain, GCFC – GC-rich sequence DNA-binding factor-like protein, HMG – high-mobility group box, KOW – Kyprides, Ouzounis, Woese motif, R3H – R3H motif, RG – arginine-glycine repeats, RRM – RNA recognition motif, SR – serine-arginine repeats, SURP – suppressor-of-white-apricot and PRP21/SPP91, Tudor – Tudor domain, ZF – Zinc finger, ANK – ankyrin repeat, CC – coiled coil, CID – RNA polymerase 2 C-terminal domain interacting domain, FHA – forkhead-associated domain, OCRE – octamer repeat domain, SOX – SOX17/18 central domain, TID – telomerase inhibitory domain, TIPN – Tuftelin interacting protein N-terminal domain, UHM – U2AF homology motif, WW – two tryptophan containing domain, XTBD – XRN2 binding domain, 2'-O-MT – ribose 2'-O-methyltransferase, DUF... – Domain of unknown function.

Table 1: Inventory of yeast and human G-patch proteins.

Yeast G-patch protein	Common alias	Size (kDa)	Interacting RNA helicase	Cellular pathway	Reference
Sqs1	Pfa1	87	Prp43	Ribosome biogenesis	Lebaron et al. (2009); Pertschy et al. (2009)
Pxr1	Gno1	31	Prp43	Ribosome biogenesis/snoRNP biogenesis/telomerase inhibition	Guglielmi and Werner (2002); Lin and Blackburn (2004); Robert-Paganin et al. (2017)
Spp382	Ntr1	83	Prp43	Pre-mRNA splicing	Christian et al. (2014); Fourmann et al. (2016); Fourmann et al. (2017)
Cmg1	YLR271W	32	Prp43	?	Heininger et al. (2016)
Spp2		21	Prp2	Pre-mRNA splicing	Warkocki et al. (2015); Hamann et al. (2019)
Human G-patch protein	Yeast ortholog	Size (kDa)	Interacting RNA helicase	Cellular pathway	Reference
AGGF1		81	?	Angiogenesis/transcription regulation	Tian et al. (2004); Major et al. (2008)
CHERP		104	?	Pre-mRNA splicing	Agafonov et al. (2011); De Maio et al. (2018)
CMTR1		95	DHX15	Pre-mRNA capping	Inesta-Vaquera et al. (2018); Toczydlowska-Socha et al. (2018)
GPANK1		39	?	?	
GPATCH1		103	?	Pre-mRNA splicing?	Agafonov et al. (2011)
GPATCH2		59	DHX15	Ribosome biogenesis?	Lin et al. (2009)
GPATCH3		59	?	Innate immune response to viral infection	Nie et al. (2017)
GPATCH4		50	?	?	
GPATCH8		164	?	?	
GPATCH11	Cmg1	33	?	?	
GPLOW	Spp2	52	DHX16	Pre-mRNA splicing/Innate immune response?	Hegele et al. (2012); Zang et al. (2014); Zhang et al. (2005)
NKRF	Sqs1	78	DHX15	Ribosome biogenesis/transcription regulation	Nourbakhsh and Hauser (1999); Nourbakhsh et al. (2000, 2001); Feng et al. (2002); Memet et al. (2017); Coccia et al. (2017)
PINX1	Pxr1	37	DHX15	Telomerase inhibition/ ribosome biogenesis	Zhou and Lu (2001); Banik and Counter (2004); Zhou et al. (2011); Chen et al. (2014)
RBM5		92	DHX15	Alternative splicing	Fushimi et al. (2008); Bonnal et al. (2008); Agafonov et al. (2011); Niu et al. (2012); Bechara et al. (2013)
RBM6		129	?	Alternative splicing	Bechara et al. (2013)
RBM10		104	?	Alternative splicing	Agafonov et al. (2011); Bechara et al. (2013); Wang et al. (2013)
RBM17 (SPF45)		45	DHX15	Pre-mRNA splicing	Lallena et al. (2002); Corsini et al. (2007); Agafonov et al. (2011); De Maio et al. (2018)
SON		264	?	Pre-mRNA splicing/Alternative splicing/transcription regulation	Ahn et al. (2013); Kim et al. (2016); Tokita et al. (2016)
SOX7		49	?	Transcription	Peng et al. (2019)
SUGP1		72	?	Alternative splicing	Kim et al. (2016); Zhang et al. (2019); Liu et al. (2020)
SUGP2		120	?	Alternative splicing?	Agafonov et al. (2011)
TFIP11	Spp382	97	DHX15	Pre-mRNA splicing	Yoshimoto et al. (2009)
ZGPAT (ZIP)		57	DHX15	Pre-mRNA splicing/transcription regulation	Li et al. (2009); Yu et al. (2010); Chen et al. (2017)

Table 1: (continued)

Viral/parasitic G-patch proteins	Organism	Size (kDa)	Interacting RNA helicase	Cellular Pathway	Reference
G-patch between protease and reverse transcriptase	Human endogenous retrovirus K	?	?	?	Aravind and Koonin (1999); Gifford et al. (2005)
Retroviral protease and reverse transcriptase	Betaretroviruses (e.g. Mason- Pfizer monkey virus)	17 and 54	?	Reverse transcrip- tion of viral RNA	Křížová et al. (2012); Gifford et al. (2005); Švec et al. (2004)
DRE	<i>Toxoplasma gondii</i> , <i>Plasmodium falciparum</i> , <i>Plasmodium yoelii</i>	50	?	DNA repair?	Dendouga et al. (2002)

Figure 2: Structural view of G-patch-RNA helicase interaction. (A) Conserved domain architecture of DEAH/RHA helicases annotated as N-terminus (N-term), RecA1 and 2, winged helix (WH), helical-bundle (HB) and oligonucleotide/oligosaccharide-binding fold (OB) domains. (B) Structure of the G-patch of *HsNKRF* (ScSqs1, coloured red) bound to *HsDHX15* (ScPrp43, coloured blue/purple/grey) based on PDB-ID 6SH6 (Studer et al. 2020). RNA (black) was modelled using superposition with *CtPrp43* from PDB-ID 5LTA (Taichert et al. 2017) to indicate the RNA binding channel. ADP and Mg²⁺ (dark grey) localise the ATPase active site (C, D). The G-patch peptide has two major contact points on the helicase one formed by its N-terminal α -helix (brace-helix) on the WH domain (C) and one formed by its C-terminal loop (brace-loop) on the RecA2 domain (D). Side chains of all residues involved in the interface are shown and G-patch residues are labelled, highly conserved positions of the consensus G-patch motif are highlighted by red boxes. (E) Sequence alignment of the five yeast G-patch proteins and their human counterparts. Invariable residues and positions with 70% similarity are highlighted in dark and light pink, respectively. The consensus sequence is shown in red below the alignment with small letters denoting a – aromatic, h – hydrophobic, l – aliphatic, s – small, u – tiny, and x – variable amino acids.

implication in RNA metabolism. In addition, modules mediating protein-protein interactions, besides the G-patch itself, are found in almost half of all family members. The diversity of domains that accompany the motif is mirrored by the diversity of cellular functions and cellular localizations of G-patch proteins (Heininger et al. 2016).

Cellular functions of G-patch proteins and their helicase interaction partners

G-patch proteins have been identified in the nucleoplasm, cytoplasm, nucleoli and mitochondria, and the proteins characterised so far have been implicated in a range of different cellular processes. In some cases, the ascribed functions are directly linked to an identified helicase interaction partner, while others are currently attributed to only the G-patch protein. However, whether these truly reflect helicase-independent functions or whether the role of a helicase interaction partner has been overlooked, often remains to be determined.

In pre-mRNA splicing

Many precursor messenger RNA (pre-mRNA) contain non-coding intron sequences that need to be removed by the spliceosome in order to obtain a mature mRNA that can serve as a template for translation (Will and Lührmann 2011). At least eight different DExD/H-box ATPases or helicases are essential for pre-mRNA splicing. Two of them, the DEAH-box proteins *ScPrp2/HsDhx16* and *ScPrp43/HsDhx15*, act in complex with the G-patch proteins *ScSpp2/HsGPKOW* and *ScSpp382/HsTFIP11*, respectively.

The yeast Prp2-Spp2 complex is required for the activation of the spliceosome just prior to the first *trans*-esterification reaction. Spp2 stimulates the ATPase activity of RNA-loaded Prp2, but no helicase activity of Prp2 could be observed (Warkocki et al. 2015). Prp2 is thought to activate the spliceosome by translocation of a single-stranded RNA reminiscent of a winching mechanism as it is located at the periphery of the spliceosome (Hamann et al. 2019). The human counterparts of Prp2 and Spp2, DHX16 and GPKOW have been identified in analogous spliceosomal complexes and a direct interaction between these proteins involving the G-patch domain of GPKOW has been demonstrated (Hegele et al. 2012; Zang et al. 2014). Although DHX16 is still recruited to spliceosomes in cells lacking GPKOW, pre-mRNA splicing is impaired under this

condition. However, GPKOW carrying mutations within the G-patch domain is still functional in pre-mRNA splicing suggesting that presence of GPKOW in spliceosomal complexes is more relevant than its ability to bind DHX16 (Zang et al. 2014). In this context, it is suggested that the GPKOW-DHX16 complex may be required for recruitment of hPRP16, which can then partially outcompete DHX16 for GPKOW interaction (Hegele et al. 2012). Interestingly, GPKOW is a phosphorylation substrate of protein kinase A and phosphorylation of GPKOW impairs its interaction with RNA, suggesting that dynamic post-translational modification may regulate GPKOW recruitment to spliceosomes and/or its functions in pre-mRNA splicing (Aksaas et al. 2011).

In yeast, Prp43 is part of the spliceosomal NTR (Nine-Teen-Related) complex and dismantles post-catalytic intron-lariat spliceosomes in cooperation with Spp382 (Ntr1) and Ntr2, which is necessary for the recycling of the snRNPs and other splicing factors. The G-patch domain of Spp382 is sufficient to stimulate ATPase and helicase activities of Prp43 (Christian et al. 2014). Prp43 activated just by the G-patch domain disassembles all spliceosomal complexes, like the A complex, the B^{act} complex, the B complex, and the post-catalytic intron-lariat complex, while the Prp43-Spp382 complex exclusively acts on the correct substrate, the intron-lariat complex (Fourmann et al. 2016, Fourmann et al. 2017). Hence, full-length Spp382 ensures the correct spatial and temporal recruitment of Prp43 to the spliceosome, demonstrating the important safeguarding role of the domain adjacent to the G-patch domain. In human cells, DHX15 is also implicated in disassembly of intron-lariat complexes, where it has been shown to interact with the Spp382 ortholog TFIP11 in a G-patch dependent manner (Yoshimoto et al. 2009), implying that the function of this complex is conserved from yeast to humans.

Several human G-patch proteins without yeast orthologs have also been identified in purified spliceosomal complexes (CHERP, GPATCH1, RBM5, RBM10, RBM17, SON, SUGP1, SUGP2) (Agafonov et al. 2011). One of the best characterised human G-patch proteins is SON, which was first identified as a DNA-binding protein, but was then found to play prominent roles in pre-mRNA splicing (reviewed in Lu et al. 2014). SON contains an SR domain enriched in serine/arginine dipeptide repeats and localises to nuclear speckles, where it is involved in regulation of splicing factor organisation perhaps by functioning as a molecular scaffold. SON contributes to efficient splicing of a specific set of pre-mRNAs containing weak or dual-specificity splice sites that encode cell cycle regulators (Ahn et al. 2011). This function involves direct interaction of

SON with its substrate transcripts and requires both the SR and G-patch domains, while the multiple unique repetitive motifs serve to further enhance splicing activity (Ahn et al. 2011). Mechanistically, it has also been proposed that SON bridges interactions between RNA polymerase II and other components of the spliceosome to promote splicing at such sub-optimal sites. Beyond constitutive splicing, SON has also been implicated in regulating alternative splicing of transcripts encoding factors involved in cell cycle regulation, apoptosis, cell adhesion and cell signalling (Sharma et al. 2011). SON is highly expressed in human embryonic stem cells (hESC) and is implicated in acquisition and regulation of pluripotency. In hESC, depletion of SON leads to alternative splicing of cassette exons and increases intron retention, with short introns flanked by weak splice sites in GC-rich contexts preferentially included (Lu et al. 2013). For example, upon lack of SON, exon 2 of *PRDM14* is skipped leading to expression of a short isoform unable to promote pluripotency induction and hESC differentiation. Consistent with its importance for maintaining correct splicing patterns, mutations within *SON* are linked to a human disorder. *SON* mutations that are observed in patients with Intellectual Disability syndrome and failure to thrive, affect splicing of pre-mRNAs encoding proteins essential for brain development and metabolism (Kim et al. 2016; Tokita et al. 2016).

The G-patch protein RBM17 (alias SF45) regulates alternative splicing of the pre-mRNAs encoding sex-lethal (*SXL*) and the apoptosis regulatory factor *FAS* in *Drosophila melanogaster* and humans respectively (Corsini et al. 2007; Lallena et al. 2002). The 3' splice site of exon 3 of the *SXL* pre-mRNA is composed of tandem AG dinucleotides flanking a pyrimidine-rich (Py) tract; the Py tract and downstream AG are bound by the U2AF complex (U2AF65 and U2AF35) whereas the upstream AG is bound by RBM17 (Lallena et al. 2002). Binding of RBM17 and *SXL* inhibits the second catalytic step of the splicing reaction, preventing inclusion of exon 3, which contains in-frame termination codons, thereby promoting expression of the functional *SXL* proteins (Lallena et al. 2002). As well as the G-patch domain, RBM17 contains a U2AF-homology motif (UHM; Figure 1) via which it interacts with the UHM-ligand motifs (ULMs) of U2AF65, SF1 and SF3b155 that all associate with the 3' splice site (Corsini et al. 2007). The UHM of RBM17 is required for its role in promoting skipping of exon 6 in the *FAS* pre-mRNA leading to production of a soluble, anti-apoptotic version of *FAS* rather than the proapoptotic, transmembrane isoform known as the “death receptor” (Corsini et al. 2007). More recently, RBM17 was also shown to form a sub-complex with U2SURP and another G-patch protein CHERP, in which the proteins reciprocally regulate

each other's stability. Depletion of any of these factors modulates the alternative splicing of a common subset of pre-mRNAs encoding RNA processing factors and cell cycle regulators, and leads to the use of thousands of cryptic junctions (De Maio et al. 2018), suggesting that these proteins cooperate to repress cryptic events and influence RNA metabolism by altering expression of RNA-binding proteins.

While the closely-related SURP and G-patch domain-containing proteins SUGP1 and SUGP2 were identified in spliceosomal complexes several decades ago (Jurica et al. 2002; Neubauer et al. 1998; Rappsilber et al. 2002), they remained poorly characterised. Recently, however, SUGP1 was found to interact with the essential splicing factor SF3B1, and it was revealed that disruption of this interaction by disease-associated mutations in SF3B1 induces splicing errors by promoting recognition of upstream branchpoints during the splicing reaction that result in use of cryptic upstream 3' splice sites (Zhang et al. 2019). Subsequently, cancer-associated mutations in SUGP1 that also disrupt SF3B1 interaction and induce analogous splicing defects were also identified (Liu et al. 2020). Interestingly, overexpression of SUGP1 carrying an amino acid exchange in the G-patch domain causes similar splicing defects to lack of SUGP1 or mutation of SF3B1, suggesting that this important function may also involve interaction with a DEAH-box RNA helicase. SUGP1 has also recently been linked to cholesterol metabolism through regulation of alternative splicing of *HMGCR*. Depletion of SUGP1 promotes skipping of several *HMGCR* exons, leading to increased expression of normally rare isoforms lacking a portion of the catalytic domain. Consistent with diminished *HMGCR* activity, lack of SUGP1 reduces cholesterol synthesis and drives LDL uptake (Kim et al. 2016).

RBM5 is a known cofactor of the RNA helicase DHX15 (Niu et al. 2012), which acts as an important regulator of cell proliferation and survival through its roles in regulating alternative splicing. RBM5 is involved in 3' splice site recognition and regulates expression of alternative isoforms of various apoptosis-related genes including caspase-2 and the Fas receptor (Bonnal et al. 2008; Fushimi et al. 2008). More specifically, in the case of the Fas receptor exon 6, early events of splice site recognition via stable association of the U1 and U2 snRNPs occur independent of RBM5. However, RBM5 stalls the downstream spliceosome assembly steps thus impeding splice site pairing (Bonnal et al. 2008). RBM5 and its close homologs RBM6 and RBM10 have also been shown to antagonistically regulate cancer cell proliferation by modulating alternative splicing of *NUMB*, a key regulator of the Notch signalling

pathway (Bechara et al. 2013). Exon 9 skipping, promoted by RBM5 or RBM6 depletion, leads to expression of a NUMB isoform that acts as NOTCH pathway repressor driving cellular proliferation whereas inclusion of exon 9, observed when RBM10 is lacking, leads to reduced NUMB levels and activation of the NOTCH pathway (Bechara et al. 2013). Genetic mutations within *RBM10* underlie the X-linked disorder TARP syndrome and it has been shown that splicing defects in patient-derived cells arise due to lack of functional RBM10 (Wang et al. 2013). RBM5, RBM6 and RBM10 typically bind to exons in proximity to weak 5' (or 3') splice sites where RBM5 recognises a UC-rich motif and RBM6 and RBM10 preferentially bind a CUCUGAA motif reminiscent of PTB binding sites. Mechanistically, RBM6 is proposed to promote exon skipping by enhancing the function of the distal splice sites whereas the binding pattern of RBM10 suggests a role in 5' splice site repression (Bechara et al. 2013).

In contrast to the G-patch proteins identified in core spliceosomal complexes, ZGPAT likely plays an indirect role in splicing regulation as it has been found in an ~35S assembly intermediate of the U4/U6.U5 tri-snRNP present in Cajal bodies (Chen et al. 2017). ZGPAT binds DHX15 to stimulate its ATPase and unwinding activities, and as DHX15 is also present in the 35S tri-snRNP, it is possible that ZGPAT activates DHX15 during U4/U6.U5 maturation.

In ribosome biogenesis

Production of the small and large ribosomal subunits (40 and 60S, respectively) requires the action of numerous *trans*-acting ribosome biogenesis factors that fulfil diverse structural and catalytic function during the assembly process (Bohnsack and Bohnsack 2019). Among these are numerous RNA helicases, including the DEAH/RHA proteins Dhr1/DHX37, Dhr2 and Prp43/DHX15 (Martin et al. 2013). While no function or mechanism of regulation has yet been described for Dhr2, the activity of Dhr1/DHX37 in release of the U3 snoRNA from pre-ribosomes is regulated by the non-G-patch protein cofactor Utp14/UTP14A (Boneberg et al. 2019; Choudhury et al. 2019; Sardana et al. 2015). In yeast, the G-patch protein Sqs1 (alias Pfa1) binds to and stimulates the ATPase and unwinding activities of Prp43, and both proteins are physically associated with pre-40S particles (Bohnsack et al. 2009; Lebaron et al. 2009; Robert-Paganin et al. 2017). Cells lacking either Prp43 or Pfa1, or where the interaction between these proteins is perturbed, show a strong accumulation of the 20S pre-rRNA species, indicating a defect in 18S rRNA 3' end processing (Lebaron et al. 2009; Pertschy

et al. 2009). It is proposed that Sqs1 stimulates the remodelling activity of Prp43 to facilitate access of the endonuclease Nob1 to its cleavage site at the 3' end of the 18S rRNA sequence. Interestingly, while the proposed human ortholog of Sqs1, GPATCH2, has been shown to associate with and stimulate the catalytic activity of DHX15 (Lin et al. 2009), in contrast to Prp43, DHX15 does not bind late pre-40S particles (Sloan et al. 2019), suggesting that the role of the DHX15-GPATCH2 complex may differ from its yeast counterpart. Prp43 is also implicated in pre-60S maturation by releasing a subset of small nucleolar RNAs (snoRNAs) from their pre-rRNA binding sites within the 25S rRNA sequence (Bohnsack et al. 2009). The yeast G-patch protein Pxr1 (alias Gno1) and its human orthologue PINX1 associate with Prp43/DHX15 via their G-patch domains and stimulate their activity in a G-patch-dependent manner (Chen et al. 2014). Pxr1 physically associates with 90S and pre-60S particles where it likely stimulates the activity of Prp43 for snoRNA release (Robert-Paganin et al. 2017). A comparable role for DHX15 and PINX1 in release of snoRNAs from human pre-60S complexes is yet to be demonstrated. Interestingly, in yeast, Pxr1 has also been observed to be required for maturation of the intron-encoded small nucleolar RNAs (snoRNAs) U18 and U24 that guide rRNA 2'-O-methylation (Guglielmi and Werner 2002). Notably, this function is independent of the G-patch domain of Pxr1, and therefore likely also independent of Prp43, and instead requires a KK(E/D) motif commonly found in nucleolar proteins including the core snoRNP proteins Nop56 and Nop58.

In human cells, DHX15 is required for an early, metazoan-specific pre-rRNA cleavage event at the A' site within the 5' external transcribed spacer of the initial pre-rRNA transcript (Memet et al. 2017). Stimulation of the catalytic activity of DHX15 by its G-patch protein cofactor NKRF likely promotes pre-rRNA remodelling to facilitate cleavage of this site by an unknown endonuclease. Alongside its G-patch domain, NKRF contains an XTBD domain through which it associates with the 5'-3' exoribonuclease XR2 (Miki et al. 2014). NKRF works antagonistically with another XTBD-containing protein CARF to maintain the nucleoplasmic-nucleolar distribution of XR2 with NKRF being responsible for recruitment of the exonuclease into the nucleolus for its functions in pre-rRNA processing and degradation of excised pre-rRNA spacer fragments (Miki et al. 2014; Memet et al. 2017). Notably, upon heat stress, NKRF re-localises from the nucleolus to the nucleoplasm and the corresponding lack of nucleolar XR2 triggers defects in pre-rRNA processing leading to the description of NKRF as a stress-regulated switch for nucleolar homeostasis surveillance (Coccia et al. 2017).

In pre-mRNA capping

CMTR1 is the only human G-patch protein possessing a catalytic domain and is responsible for the 2'-*O*-methylation of the first transcribed nucleotide of mRNAs, thus contributing to cap formation. Binding of DHX15 to CMTR1 has been suggested to impede the methyltransferase activity of CMTR1 (Inesta-Vaquera et al. 2018), while this interaction also promotes RNA duplex unwinding by DHX15 (Inesta-Vaquera et al. 2018; Toczydłowska-Socha et al. 2018). The precise role of this complex remains unclear as on the one hand, CMTR1-DHX15 has been implicated in regulating translation of a subset of mRNAs linked to cell growth while on the other hand, the interaction of CMTR1 with DHX15 has been shown to promote first nucleotide 2'-*O*-methylation of mRNAs containing highly structured 5' ends by facilitating resolution of secondary structures thereby enhancing access of the methyltransferase to its target sites (Toczydłowska-Socha et al. 2018).

In transcription regulation

Alongside their other functions in ribosome biogenesis, and pre-mRNA splicing, several G-patch proteins act as transcriptional regulators. NKRF binds an 11 nt negative regulatory element (5'-AATTCCTCTGA-3') within the promoters of specific NF- κ B regulated genes, thereby suppressing NF- κ B activity. NKRF is involved in the constitutive silencing of interferon- β and iNOS as well as regulation of IL-8 expression (Feng et al. 2002; Nourbakhsh and Hauser 1999; Nourbakhsh et al. 2000, 2001). In contrast, ZGPAT binds a 5'-GGAG[GA]A[GA]A-3' motif and represses transcription of target genes including *EGFR* by recruitment of the nucleosome remodelling and deacetylase complex NuRD (Li et al. 2009). Via transcriptional regulation of the EGFR signalling pathway, ZGPAT impacts cell growth. Consequently its deletion leads to tumour growth *in vivo* and its downregulation was demonstrated in breast carcinomas (Li et al. 2009). Interestingly, expression of an alternatively spliced version of ZGPAT lacking the ZNF domain and unable to bind DNA antagonises this function by competing for interaction with the NuRD complex (Yu et al. 2010). This alternative isoform of ZGPAT still contains the G-patch domain and, as a core component of the NuRD complex is the helicase-domain-containing ATPase CHD4(Mi-2 β), it is tempting to speculate whether the G-patch domain of ZGPAT may mediate this interaction. SON also associates with DNA near transcription start sites, where it interacts with the menin component of the MLL histone methylation complex, preventing complex

assembly (Kim et al. 2016). This in turn leads to decreased methylation of H3K4 and consequently, transcriptional repression. Similar to ZGPAT, expression of alternative SON isoforms able to bind chromatin, but impaired in menin interaction, counteract transcriptional repression by the full-length protein. Interestingly, SON also acts as a transcriptional repressor of the *miR32a-27a-24a* cluster, and SON-dependent changes in miRNA expression lead to post-transcriptional regulation of *miR24a/27a* target genes, such as *GATA2* (Ahn et al. 2013). In contrast to the transcription repression functions of NKRF, ZGPAT and SON, AGGF1 contributes to the activation of β -catenin target genes (e.g. *LEF1* and *AXIN2*) via a physical interaction with the SWI/SNF complex that physically moves nucleosomes along DNA to modulate chromatin structure promoting or inhibiting gene transcription (Major et al. 2008). Potentially via Wnt/ β -catenin signalling AGGF1 plays a central role in angiogenesis (Tian et al. 2004) and mutation of *AGGF1* is observed in a congenital disorder with vascular and tissue malformations called Klippel-Trénaunay syndrome (Hu et al. 2008; Tian et al. 2004).

SOX7 is a transcription factor that binds 5'-(A/T)(A/T)CAA(A/T)G-3' motifs in DNA directly via its HMG box (Harley et al. 1994). Like other SOX proteins, SOX7 is required for cell fate decision during development (Julian et al. 2017), involved in human disorders (Angelozzi and Lefebvre 2019) and implicated in tumorigenesis (Grimm et al. 2020). It inhibits Wnt signalling by directly binding β -catenin via its central domain and blocking β -catenin-mediated transcription activation (Takash et al. 2001; Zorn et al. 1999). Interestingly, only SOX7 isoform 2 encodes a canonical G-patch motif, which replaces the N-terminal nuclear localization signal of isoform 1 (Peng et al. 2019). It has not been addressed whether isoform 2 of SOX7 is imported into the nucleus or whether it has a cytoplasmic function. However, in a prostate cancer cell line, which mainly expresses isoform 2, transcriptional activation of enhancers targeted by SOX7 was not dependent on isoform 2 and could only be rescued by ectopic expression of isoform 1, suggesting that the G-patch-containing SOX7 might have additional roles in cells where it is expressed (Peng et al. 2019).

In telomerase regulation

An additional function for PINX1 beyond ribosome biogenesis has also been described; via a 74 amino acid sequence within the protein C-terminal region, PINX1 binds to the catalytic telomerase subunit TERT, as well as the telomerase RNA component (TR), strongly inhibiting

telomerase activity and leading to telomere attrition (Banik and Counter 2004; Zhou and Lu 2001; Zhou et al. 2011). The action of PINX1 as a telomerase inhibitor is attenuated by a physical interaction with the ribosome assembly factor nucleophosmin, which could perhaps reflect withdrawal of PINX1 towards its other function in pre-ribosomal complexes (Cheung et al. 2017). Interestingly, in mitosis, PINX1 re-localises from nucleoli and telomeres to kinetochores and chromosome peripheries where a role in ensuring accurate chromosome segregation has been described (Yuan et al. 2009). Since telomerase activity is a hallmark of many immortalised cells, the telomerase-inhibitor PINX1 is a major tumour suppressor whose down-regulation has been reported in a number of human breast cancers (Zhou et al. 2011).

In innate immunity

Several G-patch proteins are implicated in regulating the immune response to viral infection. In *Arabidopsis thaliana*, the G-patch protein MOS2 (GPKOW homolog) is essential for innate immunity (Zhang et al. 2005) whereas in humans, GPATCH3 is reported to be a negative regulator of the innate antiviral response (Nie et al. 2017). GPATCH3 disrupts assembly of the VISA signalosome, which is a critical adaptor in RIG-I-like receptor (RLR)-mediated induction of the innate immunity, leading to a reduced antiviral response. However, binding of GPATCH3 to VISA and its role in the immune response are both independent of its G-patch domain, suggesting that this function does not involve an interaction with an RNA helicase. Interestingly, the G-patch protein-interacting RNA helicase DHX15 is a known RLR binding partner and is required for virus-induced RLR-signalling of innate immune gene expression (Pattabhi et al. 2019), but whether its activity in this pathway is regulated by any G-patch protein cofactor remains to be explored.

In retroviruses

G-patch proteins have been implicated in various aspects of viral infection. Several beta-retroviral proteases, including that from the Mason-Pfizer monkey virus (M-PMV), carry G-patch domains at their C-terminal ends. Interestingly, the G-patch domain is not required for the processing of viral polyproteins but is important for M-PMV infectivity through its roles in stimulating the activity of the reverse transcriptase and facilitating capsid assembly (Bauerová-Zábranska et al. 2005). The M-PMV G-patch

domain, which binds single stranded nucleic acids, physically associates with viral reverse transcriptase where it is suggested to promote interaction with the substrate RNA (Křizová et al. 2012; Švec et al. 2004). Members of a group of betaretrovirus-like endogenous human retroviral elements termed HERV-Ks also encode G-patch domains. These endogenous retroviral elements are generally suppressed by epigenetic and anti-viral mechanisms and are only expressed at very low levels in healthy human tissues (Garcia-Montojo et al. 2018; Hanke et al. 2016). Nevertheless, they are highly expressed in the early stages of embryonic development and in tumours including teratocarcinomas and melanomas. Nothing is currently known about the functions of such endogenous retroviral G-patch proteins and it is unclear whether they interact with endogenous DEAH-box RNA helicases.

Interactions of G-patch proteins with RNA helicases

Initially, the G-patch motif was characterized solely as a recruitment platform for DEAH helicases. Only later did it become apparent that it also regulates the RNP remodeling activity of helicases. Even before the G-patch motif had been categorized, yeast Spp2 was identified as the first protein from the G-patch family to bind a DEAH helicase, namely Prp2 (Roy et al. 1995). In this context, it was also first shown that a G-patch protein is responsible for targeting a helicase to an RNP, i.e. the B^{act} spliceosome (Silverman et al. 2004). Soon after, the minimal G-patch motif itself was established as the required feature for G-patch protein-helicase interaction within the Spp382-Prp43 heterodimer, raising the possibility that G-patch proteins could be general cofactors of DEAH helicases (Tsai et al. 2005). The ability of G-patch proteins to directly enhance the catalytic activity of DEAH helicases was first shown for this complex (Tanaka et al. 2007) and subsequently demonstrated for several different yeast and human family members and their partner helicases acting in different pathways (Chen et al. 2014; Heininger et al. 2016; He et al. 2017; Lebaron et al. 2009; Memet et al. 2017; Niu et al. 2012; Robert-Paganin et al. 2017; Warkocki et al. 2015). *In vitro* experiments could finally demonstrate that the G-patch motif is fully sufficient both for association with DEAH-box proteins as well as enhanced catalysis (Christian et al. 2014; Tauchert et al. 2017), which is also highlighted by the fact that a transplanted G-patch motif from a ribosome biogenesis factor could functionally replace the G-patch of a splicing factor (Fourmann et al. 2017). The importance of

the eponymous glycines and conserved hydrophobic residues within the motif for helicase binding and activation was underlined by negative effects of point mutations in several G-patch proteins (Chen et al. 2014; Guglielmi and Werner 2002; Memet et al. 2017; Niu et al. 2012; Pandit et al. 2006; Studer et al. 2020; Tanaka et al. 2007; Zang et al. 2014).

While the minimal region of G-patch proteins required for helicase interaction was readily identified, the opposite contact surface on DEAH proteins remained challenging to pinpoint in the absence of structural information. Like all members of the large superfamily 2 of helicases, DEAH/RHA helicases contain two RecA domains at their core (Figure 2A, B) that form a split active site for ATP hydrolysis (Fairman-Williams et al. 2010). This RecA core provides an RNA-binding channel together with the three C-terminal domains, termed winged-helix (WH), helical-bundle (HB) and oligonucleotide/oligosaccharide-binding fold (OB) (Prabu et al. 2015; Walbott et al. 2010). The channel has been suggested to allow DEAH/RHA enzymes to translocate along single stranded RNA during cycles of ATP hydrolysis without intermittent RNA dissociation (Boneberg et al. 2019; Hamann et al. 2019). The conserved domains are flanked by varying N- and C-terminal extensions that are mostly predicted to be poorly structured but in some family members also contain additional domains. Initially, deletion constructs and point mutations identified the C-terminal OB domain in Prp2 and Prp43/DHX15 as being required for binding to several G-patch partners (Memet et al. 2017; Mouffok et al. 2020; Silverman et al. 2004). Consistent with this, crosslinking mass spectrometry (XL-MS) indicated that the N-terminus of the G-patch from Spp382 binds close to the C-terminal OB domain of Prp43 both in the isolated complex (Christian et al. 2014) and when incorporated into the spliceosome (Wan et al. 2017). However, random mutations within the RecA2 domain were also observed to impair G-patch binding (Tanaka et al. 2007). Furthermore, systematic deletion of individual domains in ScPrp43 and testing for interactions with three of its G-patch partners (Pxr1, Spp382, Sqs1) indicated the strongest global effects for removal of the RecA2 and WH domains in yeast two hybrid experiments (Banerjee et al. 2015). This confusing picture was only recently resolved by the congruent determination of crystal structures of two different G-patch-helicase complexes, the Spp2-Prp2 complex from the thermophilic fungus *Chaetomium thermophilum* (Ct) (Hamann et al. 2020) and the HsNKRF-DHX15 complex (Studer et al. 2020) (Figures 2B–D and 3A). In both complexes, the G-patch stretches across the surface of the helicase at the back side of the RNA binding channel, making major contacts to the WH domain

with its N-terminus and to the RecA2 domain with its C-terminus. The middle section of the G-patch motif has barely any interaction with the helicase, consistent with its poor sequence conservation (Figure 2E). Thus, the binding mode can be compared to a brace that tethers two mobile sections of the helicase together. The N-terminus of the G-patch folds into an α -helix (brace-helix) upon helicase binding, while the rest of the peptide remains extended and devoid of secondary structure. A loop at the C-terminus (brace-loop) inserts into a deep cavity on the RecA2 domain (Figure 2B). The interface established by the brace-helix provides the main contribution to binding affinity while the brace-loop is the weaker affinity interaction site. Amino acid substitutions in the brace-helix contact region efficiently disrupt NKRF-DHX15 complex formation (Studer et al. 2020) and the Spp2 helix alone is sufficient to bind, although with a strongly reduced affinity compared to the full G-patch motif ($K_d = 40.3 \mu\text{M}$ vs. $0.72 \mu\text{M}$; Hamann et al. 2020). Corresponding substitutions within, or deletions of, the brace-loop interface in either complex have weaker effects on affinity (Hamann et al. 2020; Studer et al. 2020).

The structures also revealed that none of the highly conserved glycine residues in the motif apart from one make direct contacts with the helicase (Figure 2C, D). They rather confer the necessary conformational freedom to the backbone of the peptide so that the conserved hydrophobic/aromatic sidechains can stack into the two hydrophobic pockets on the helicase surface. These strong backbone torsions also allow the fully conserved aromatic residue (W564 in NKRF, Figure 2C, E) to insert between the G-patch peptide backbone on the one side and a composite aliphatic surface provided by brace-helix and helicase on the other side.

Both structures rationalize the previous mutational and crosslinking data listed above. They suggest that full or partial deletion of the OB domain likely affects the fold of the neighbouring WH domain such that the crucial binding site for the G-patch brace-helix is disrupted. Furthermore, given that the G-patch N-terminus contacts the helicase close to the OB domain, all crosslinked lysine residues are within the maximal crosslinking distance (Christian et al. 2014; Hamann et al. 2020; Wan et al. 2017). Finally, sidechains that were previously indicated as necessary for G-patch binding or the functionality of the helicases in their respective pathways map to the interaction surfaces (Hamann et al. 2020; Studer et al. 2020). The crystal structures also allowed for the first time to locate the G-patch peptides on the helicases in the cryo-electron microscopy structures of the spliceosome (Studer et al. 2020). This placement was previously impeded by the low local resolution of these structures at the

Figure 3: Mechanism of regulation of DEAH-box RNA helicase action by G-patch proteins.

(A) Superposition of the two G-patch conformations of fungal Spp2 (*HsGPKOW*, gold/yellow) from PDB-IDs 6RM8 and 6RM9 (Hamann et al. 2020) with the G-patch of *HsNKRf* (*ScSqs1*, red) from PDB-ID 6SH6 (Studer et al. 2020). (B) Schematic model for the mechanism of G-patch regulated DEAH-helicase action on RNPs. The helicase is recruited to an RNP via the intrinsically disordered motif in a G-patch protein, which is specifically embedded into the RNP. G-patch binding stabilizes the enzyme in a conformation with high RNA affinity and activates ATPase and helicase activity. By translocating along the RNA in one-nucleotide steps the helicase pulls on the RNA from the periphery of the RNP and thereby elicits conformational changes or disruption of RNA base-pairs at distant locations within the RNP.

periphery of the spliceosomes where the DEAH helicases are bound. Docking of the G-patch structure into the low resolution electron density maps indicated that their steric organisation on the DEAH surfaces is maintained within the spliceosome.

The high similarity of G-patch binding to two different helicases (i.e. Prp2/DHX16 and Prp43/DHX15) raises the question of how specific recruitment of only one desired helicase is achieved during different steps of RNP rearrangements, such as splicing (Cordin and Beggs 2013). Available structural and XL-MS data suggest that the unique N- and C-terminal extensions of the DEAH helicases could add a layer of specificity by providing additional contact points on RNPs. In addition, other distinct surface patches on the central domains of the helicase could be sources for differentiation. For example, the unique N-terminus of Prp2 crosslinks with the spliceosome

component Prp45 in yeast (Rauhut et al. 2016). Prp43 meanwhile binds to the spliceosome component Syf1 via surfaces on the OB and HB domains that are not well conserved in Prp2, such as the last C-terminal α -helix which is completely absent in Prp2 (Schmitt et al. 2018; Wan et al. 2017).

Conservation of the G-patch motif between many different family members also means that multiple G-patch proteins can compete for binding to the same helicase, as has been observed for ScPrp43 and its four G-patch cofactors (Heininger et al. 2016). Exploiting these mutually exclusive interactions, the cofactors effectively partition the helicase between its different target pathways in distinct cellular compartments. Potentially to increase the efficiency of this competition, the G-patch cofactor Sqs1 forms additional contacts with Prp43. The N-terminus of Sqs1 was demonstrated to bind Prp43 independently of the

G-patch containing C-terminus (Lebaron et al. 2009). This association does not influence the helicase activity of Prp43 and depends on the presence of the N-terminal extension of Prp43 (Mouffok et al. 2020). As Sqs1 and Pxr1 are both part of the ribosome biogenesis pathway, with Pxr1 acting in early (Robert-Paganin et al. 2017) and Sqs1 acting in later assembly steps (Lebaron et al. 2009; Pertschy et al. 2009), the additional anchor point on Prp43 might ensure an efficient handover of the helicase between the two G-patch partners at different maturation states of the ribosomal precursor.

The available structural information also allows predicting selectivity of G-patch proteins for a subset of DEAH/RHA helicases (Studer et al. 2020). Unlike DHX15 and DHX16, most other helicase family members in human cells have substitutions in the two G-patch contact sites that are incompatible with conserved G-patch binding. This observation explains why for example the other two splicing DEAH helicases (*ScPrp16/HsDHX38*, *ScPrp22/HsDHX8*) are not bound by G-patch activators (Cordin and Beggs 2013), or why *ScDhr1/HsDHX37* is not targeted by G-patch proteins but has another type of cofactor (Boneberg et al. 2019; Choudhury et al. 2019; Sardana et al. 2015). However, it still remains to be seen experimentally whether any of the other DEAH helicases in human cells can associate with G-patch partners.

Regulation of RNA helicase activity by G-patch proteins

Many DEAH helicases are poor enzymes on their own and also have no intrinsic specificity for selecting RNA targets. Potentially to prevent spurious dissolution of RNA secondary structures in cells, many RNA helicases require partners to endow substrate specificity and trigger activation on appropriate targets. Consistent with roles as cofactors, all G-patch proteins characterized thus far have stimulatory effects on the enzymatic activities of their cognate helicase partners (Chen et al. 2014; Christian et al. 2014; Fourmann et al. 2017; He et al. 2017; Heininger et al. 2016; Lebaron et al. 2009; Lin et al. 2009; Memet et al. 2017; Niu et al. 2012; Tauchert et al. 2017; Warkocki et al. 2015). These G-patch proteins increase the ATPase activity of their associated helicase and enhance its affinity for single stranded RNA thereby promoting dsRNA unwinding. The available structural and biochemical data suggest that the molecular basis of this activation is tethering together of the C-terminal WH and the RecA2 domain. An accumulation of DEAH/RHA helicase structures in different states

has visualized the conformational plasticity of these enzymes in isolation (Boneberg et al. 2019; Chen et al. 2018a; Chen et al. 2018b; Christian et al. 2014; Hamann et al. 2019; He et al. 2017; Schmitt et al. 2018). The C-terminal domains are able to rotate out by up to 40° thereby completely disrupting the enclosed nature of the RNA binding channel (Chen et al. 2018a; Tauchert et al. 2017). Similarly, the two RecA domains can shift relative to each other by 6 Å and rotate 20°, depending on the ATP hydrolysis state (Boneberg et al. 2019; Chen et al. 2018b; Hamann et al. 2019). The conformation in which the G-patch is bound is only consistent with the conformation occupied upon RNA binding where the RNA channel is closed. Biochemical data indicates that disruption of this tether without loss of overall G-patch binding not only decreases RNA affinity but also alleviates the stimulatory effect on ATPase and RNA unwinding activity (Studer et al. 2020).

Another feature of G-patch proteins that is of great significance for their mode of activation is that the G-patch motifs are embedded in intrinsically disordered regions and are themselves devoid of a stable fold (Christian et al. 2014). In the absence of a binding partner, they have only a weak tendency to form a short α -helix at the N-terminus (Hamann et al. 2020). This lack of secondary and tertiary structure offers several advantages for efficient association with binding partners, due to what has been termed a “fly-casting mechanism” (Shoemaker et al. 2000). Firstly, a disordered region can sample a larger radius for presence of a binding partner than a folded domain. Thus, G-patch motifs can efficiently “fish” for a helicase while being attached to a large, immobile RNP. Consistent with this, in all spliceosome structures analysed so far, the G-patch and the DEAH helicases sit on the solvent-accessible periphery of the particles (Rauhut et al. 2016; Wan et al. 2017). Secondly, disordered peptides often fold on the surface of their interaction partners and have several lower affinity binding sites. This means that they can first associate with one of these sites, which induces formation of secondary structure elements (e.g. α -helices). The resulting spatial compression of the formerly extended peptide shortens the distance between the two proteins, effectively “reeling in” the binding partner. Subsequent attachment to the other contact sites can then stabilise the complex. This observation also holds for the G-patch, as its N-terminal α -helix is just transiently sampled in isolation, folding stably only after association with the WH domain (Christian et al. 2014; Hamann et al. 2020). Furthermore, the inherent flexibility that is a characteristic of intrinsically disordered motifs, such as the G-patch, provides an ideal binding platform for enzymes that have to undergo larger conformational changes during catalysis. In the case of RNA helicases,

efficient RNA translocation coupled to ATP hydrolysis requires that opening-closing motions of the two RecA domains are not hindered by the activator (Hamann et al. 2019). This requirement is perfectly fulfilled by the extended conformation of the G-patch peptide that stretches along the helicase surface and can follow domain motions dynamically like a spring (Studer et al. 2020). Further supporting the flexible attachment of the G-patch, Spp2 has been crystallized in two conformations on the Prp2 surface (Hamann et al. 2020) with the main difference in the orientation of the C-terminal brace-loop (Figure 3A). In both arrangements, the loop stacks into the same pocket on the RecA2 surface, however it can use two different neighbouring leucine residues for this connection, which could further increase the conformational freedom of the helicase. Finally, multiple surface contacts also allow for regulatory interactions by restricting the available conformational space of an enzyme to states that are either inhibitory (Bason et al. 2014) or on-path of the catalytic trajectory (Wurm and Sprangers 2019). G-patch proteins exploit this principle by stabilizing the closed RNA channel of DEAH helicases.

The observation that DEAH helicases are mostly placed on the RNP exterior, both in the case of spliceosome intermediates (Kastner et al. 2019; Yan et al. 2019) and ribosome precursors (Cheng et al. 2020), seems puzzling given that their catalytic action has wide-ranging effects that extend to conformational changes at the centre of the RNP. Consistent with the structural snapshots of spliceosomal complexes, it has been shown for Prp16 and Prp22 that they do not unwind double-stranded sections of the spliceosome by translocating directly through them (Semlow et al. 2016). Instead, it has been suggested that these peripheral helicases work by pulling the RNA through the RNP. In agreement with this model, Prp2 cannot unwind dsRNA both in the presence and absence of Spp2 or the spliceosome but rather seems to remodel the B^{act} core remotely (Bao et al. 2017; Kim et al. 1992; Warkocki et al. 2015).

In summary, the available data is consistent with a model in which DEAH/RHA helicases are in a highly flexible autoinhibited state before being recruited by the unstructured G-patch motif to the free 3'-end of an RNA on a target RNP (Figure 3B). Upon contact, the G-patch folds onto the surface of the helicase forming a flexible brace that stabilises the RNA-bound conformation. This binding mode still allows for productive RecA domain motions that lead to efficient ATP hydrolysis and RNA ratcheting through the channel, thereby exerting force on the RNA which can lead to distant conformational changes within the RNP.

The activation mechanism of DEAH helicases by G-patch proteins shares several parallels with the

mechanism that is used by MIF4G domains to stimulate the related DEAD-box helicases (Bourgeois et al. 2016; Ozgur et al. 2015; Schütz et al. 2008; Sloan and Bohnsack 2018). The two RecA domains of DEAD-box proteins can also exhibit a multitude of respective orientations but need to align in order to bind their RNA targets on a composite surface. This aligned conformation with high RNA affinity is stabilized analogously by the MIF4G domain-containing cofactors by tethering the two RecA domains together. Similar to G-patch proteins, MIF4G domains provide a low (RecA1) and high affinity patch (RecA2) for the helicase. ATP hydrolysis, RNA backbone distortion and product release also require conformational changes within the helicase core. Given that the MIF4G adaptor is a folded domain in contrast to the G-patch, it can only enable such movements by detaching from the low affinity site. However, by remaining bound to RecA2, the MIF4G cofactor can quickly re-establish the second domain contact and therefore, like a G-patch, also helps to minimize spurious domain movements of the helicase.

Conclusions and outlook

The G-patch proteins represent a heterogeneous family of proteins involved in diverse aspects of RNA metabolism. Biochemical analyses have confirmed the roles of several members of this protein family as *bona fide* cofactors of DEAH-box RNA helicases, but for many human G-patch proteins, it remains unclear if they interact directly with an RNA helicase *in vivo*. Recent structural analyses of yeast and human G-patch-helicase complexes have provided the first mechanistic insights into how these cofactors can stimulate the ATPase and unwinding activities of their cognate DEAH-box RNA helicases, but it remains to be seen whether all G-patch proteins that associate with an RNA helicase interact and stimulate catalytic activity in a similar manner. Despite a wealth of information on cellular processes involving G-patch proteins, beyond a handful of well-characterised G-patch protein-helicase complexes, it often remains unclear if these functions are fulfilled by the G-patch protein alone or whether the contribution of an RNA helicase interaction partner has thus far been overlooked. The regulation of multifunctional RNA helicases, such as Prp43/DHX15, through interactions with related cofactor proteins represents an efficient mechanism by which to fine-tune RNA helicase function within the complex cellular environment and opens the possibility for cross-talk between different aspects of RNA metabolism. However, obtaining a comprehensive inventory of interactions between G-patch proteins and DEAH-box RNA

helicases, together with greater knowledge on the cellular functions of the individual complexes will be necessary before the dynamics and interplay of the G-patch protein network can be fully understood.

Acknowledgements: This work was funded by the Deutsche Forschungsgemeinschaft (SFB860 to R.F., M.T.B. and K.E.B.), the Swiss National Science Foundation (SNSF) through the National Center for Competence in Research “RNA & Disease” (to S.J.) and SNSF Grant 31003A_179498 (to S.J.).

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- Absmeier, E., Santos, K.F., and Wahl, M. C. (2020). Molecular mechanism underlying inhibition of intrinsic ATPase activity in a Ski2-like RNA helicase. *Structure* 28: 236–243.
- Agafonov, D.E., Deckert, J., Wolf, E., Odenwälder, P., Bessonov, S., Will, C.L., Urlaub, H., and Lührmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell Biol.* 31: 2667–2682.
- Ahn, E.Y., DeKolver, R.C., Lo, M.C., Nguyen, T.A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X.D., and Zhang, D.E. (2011). SON controls cell-cycle progression by coordinated regulation of RNA splicing. *Mol. Cell.* 42: 185–198.
- Ahn, E.E., Higashi, T., Yan, M., Matsuura, S., Hickey, C.J., Lo, M.C., Shia, W.J., DeKolver, R.C., and Zhang, D.E. (2013). SON protein regulates GATA-2 through transcriptional control of the microRNA 23a~27a~24-2 cluster. *J. Biol. Chem.* 288: 5381–5388.
- Aksaas, A.K., Larsen, A.C., Rogne, M., Rosendal, K., Kvissel, A.K., and Skålhegg, B.S. (2011). G-patch domain and KOW motifs-containing protein, GPKOW; a nuclear RNA-binding protein regulated by protein kinase A. *J. Mol. Signal.* 6: 10.
- Alexandrov, A., Colognori, D., and Steitz, J. A. (2011). Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. *Genes Dev.* 25: 1078–1090.
- Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313: 1968–1972.
- Angelozzi, M. and Lefebvre, V. (2019). SOXopathies: growing family of developmental disorders due to SOX mutations. *Trends Genet.* 35: 658–671.
- Aravind, L. and Koonin, E.V. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem. Sci.* 24: 342–344.
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12: 861–869.
- Banerjee, D., McDaniel, P.M., and Rymond, B.C. (2015). Limited portability of G-patch domains in regulators of the Prp43 RNA helicase required for pre-mRNA splicing and ribosomal RNA maturation in *Saccharomyces cerevisiae*. *Genetics* 200: 135–147.
- Banik, S.S. and Counter, C.M. (2004). Characterization of interactions between PinX1 and human telomerase subunits hTERT and hTR. *J. Biol. Chem.* 279: 51745–51748.
- Bao, P., Höbartner, C., Hartmuth, K., and Lührmann, R. (2017). Yeast Prp2 liberates the 5′ splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA* 23: 1770–1779.
- Bason, J.V., Montgomery, M.G., Leslie, A.G.W., and Walker, J.E. (2014). Pathway of binding of the intrinsically disordered mitochondrial inhibitor protein to F1-ATPase. *Proc. Natl. Acad. Sci. U.S.A.* 111: 11305–11310.
- Bauerová-Zábranská, H., Štokrová, J., Stříšovský, K., Hunter, E., Ruml, T., and Pichová, I. (2005). The RNA binding G-patch domain in retroviral protease is important for infectivity and D-type morphogenesis of Mason-Pfizer monkey virus. *J. Biol. Chem.* 280: 42106–42112.
- Bechara, E.G., Sebestyén, E., Bernardis, I., Eyra, E., and Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell.* 52: 720–733.
- 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.
- Bohnsack, K.E. and Bohnsack, M.T. (2019). Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *EMBO J.* 38: e100278.
- Boneberg, F.M., Brandmann, T., Kobel, L., van den Heuvel, J., Bargsten, K., Bammert, L., Kutay, U., and Jinek, M. (2019). Molecular mechanism of the RNA helicase DHX37 and its activation by UTP14A in ribosome biogenesis. *RNA* 25: 685–701.
- Bonnal, S., Martínez, C., Förch, P., Bachi, A., Wilm, M., and Valcárcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol. Cell.* 32: 81–95.
- 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.
- Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* 17: 426–438.
- Caruthers, J.M. and McKay, D.B. (2002). Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* 12: 123–133.
- Chakrabarti, S., Jayachandran, U., Bonneau, F., Fiorini, F., Basquin, C., Domcke, S., Le Hir, H., and Conti, E. (2011). Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell.* 41: 693–703.
- Chen, Y.L., Capeyrou, R., Humbert, O., Mouffok, S., Kadri, Y.A., Lebaron, S., Henras, A.K., and Henry, Y. (2014). The telomerase inhibitor Gno1p/PINX1 activates the helicase Prp43p during ribosome biogenesis. *Nucleic Acids Res.* 42: 7330–7345.
- Chen, Z., Gui, B., Zhang, Y., Xie, G., Li, W., Liu, S., Xu, B., Wu, C., He, L., Yang, J., et al. (2017). Identification of a 35S U4/U6.U5 tri-small

- nuclear ribonucleoprotein (tri-snRNP) complex intermediate in spliceosome assembly. *J. Biol. Chem.* 292: 18113–18128.
- Chen, M.C., Tippana, R., Demeshkina, N.A., Murat, P., Balasubramanian, S., Myong, S., and Ferré-D'Amaré, A.R. (2018a). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* 558: 465–469.
- Chen, W.-F., Rety, S., Guo, H.-L., Dai, Y.-X., Wu, W.-Q., Liu, N.-N., Auguin, D., Liu, Q.-W., Hou, X.-M., Dou, S.-X., et al. (2018b). Molecular mechanistic insights into *Drosophila* DHX36-mediated G-quadruplex unfolding: a structure-based model. *Structure* 26: 403–415.e4.
- Cheng, J., Lau, B., La Venuta, G., Ameismeier, M., Berninghausen, O., Hurt, E., and Beckmann, R. (2020). 90S pre-ribosome transformation into the primordial 40S subunit. *Science* 369: 1470–1476.
- Cheung, D.H., Ho, S.T., Lau, K.F., Jin, R., Wang, Y.N., Kung, H.F., Huang, J.J., and Shaw, P.C. (2017). Nucleophosmin interacts with PIN2/TERF1-interacting telomerase inhibitor 1 (PinX1) and attenuates the PinX1 inhibition on telomerase activity. *Sci. Rep.* 7: 43650.
- Choudhury, P., Hackert, P., Memet, I., Sloan, K.E., and Bohnsack, M.T. (2019). The human RNA helicase DHX37 is required for release of the U3 snoRNP from pre-ribosomal particles. *RNA Biol.* 16: 54–68.
- Christian, H., Hofele, R.V., Urlaub, H., and Ficner, R. (2014). Insights into the activation of the helicase Prp43 by biochemical studies and structural mass spectrometry. *Nucleic Acids Res.* 42: 1162–1179.
- Coccia, M., Rossi, A., Riccio, A., Trotta, E., and Santoro, M.G. (2017). Human NF- κ B repressing factor acts as a stress-regulated switch for ribosomal RNA processing and nucleolar homeostasis surveillance. *Proc. Natl. Acad. Sci. U.S.A.* 114: 1045–1050.
- Cordin, O. and Beggs, J.D. (2013). RNA helicases in splicing. *RNA Biol.* 10: 83–95.
- Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcárcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat. Struct. Mol. Biol.* 14: 620–629.
- Davila Galesio, J., Hackert, P., Bohnsack, K.E., and Bohnsack, M.T. (2020). Sgd1 is an MIF4G domain-containing cofactor of the RNA helicase Fal1 and associates with the 5' domain of the 18S rRNA sequence. *RNA Biol.* 17: 539–553.
- De Maio, A., Yalamanchili, H.K., Adamski, C.J., Gennarino, V.A., Liu, Z., Qin, J., Jung, S.Y., Richman, R., Orr, H., and Zoghbi, H.Y. (2018). RBM17 interacts with U2SURP and CHERP to regulate expression and splicing of RNA-processing proteins. *Cell Rep.* 25: 726–736, e7.
- Dendouga, N., Callebaut, I., and Tomavo, S. (2002). A novel DNA repair enzyme containing RNA recognition, G-patch and specific splicing factor 45-like motifs in the protozoan parasite *Toxoplasma gondii*. *Eur. J. Biochem.* 269: 3393–3401.
- Erkizan, H.V., Schneider, J.A., Sajwan, K., Graham, G.T., Griffin, B., Chasovskikh, S., Youbi, S.E., Kallarakal, A., Chruszcz, M., Padmanabhan, R., et al. (2015). RNA helicase A activity is inhibited by oncogenic transcription factor EWS-FLI1. *Nucleic Acids Res.* 43: 1069–1080.
- Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., and Jankowsky, E. (2004). Protein displacement by DEXH/D RNA helicases without duplex unwinding. *Science* 304: 730–734.
- Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* 20: 313–324.
- Feng, X., Guo, Z., Nourbakhsh, M., Hauser, H., Ganster, R., Shao, L., and Geller, D. A. (2002). Identification of a negative response element in the human inducible nitric-oxide synthase (hNOS) promoter: the role of NF- κ B-repressing factor (NRF) in basal repression of the hNOS gene. *Proc. Natl. Acad. Sci. U.S.A.* 99: 14212–14217.
- Fourmann, J.B., Dybkov, O., Agafonov, D.E., Tauchert, M.J., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2016). The target of the DEAH-box NTP triphosphatase Prp43 in *Saccharomyces cerevisiae* spliceosomes is the U2 snRNP-intron interaction. *eLife* 5: e15564.
- Fourmann, J.B., Tauchert, M.J., Ficner, R., Fabrizio, P., and Lührmann, R. (2017). Regulation of Prp43-mediated disassembly of spliceosomes by its cofactors Ntr1 and Ntr2. *Nucleic Acids Res.* 45: 4068–4080.
- Fushimi, K., Ray, P., Kar, A., Wang, L., Sutherland, L.C., and Wu, J.Y. (2008). Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5. *Proc. Natl. Acad. Sci. U.S.A.* 105: 15708–15713.
- Garcia-Montojo, M., Doucet-O'Hare, T., Henderson, L., and Nath, A. (2018). Human endogenous retrovirus-K (HML-2): a comprehensive review. *Crit. Rev. Microbiol.* 44: 715–738.
- Gifford, R., Kabat, P., Martin, J., Lynch, C., and Tristem, M. (2005). Evolution and distribution of class II-related endogenous retroviruses. *J. Virol.* 79: 6478–6486.
- Gowravaram, M., Bonneau, F., Kanaan, J., Maciej, V.D., Fiorini, F., Raj, S., Croquette, V., Le Hir, H., and Chakrabarti, S. (2018). A conserved structural element in the RNA helicase UPF1 regulates its catalytic activity in an isoform-specific manner. *Nucleic Acids Res.* 46: 2648–2659.
- Granneman, S., Lin, C., Champion, E.A., Nandineni, M.R., Zorca, C., and Baserga, S.J. (2006). The nucleolar protein Esf2 interacts directly with the DEXH/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. *Nucleic Acids Res.* 34: 3189–3199.
- Grimm, D., Bauer, J., Wise, P., Krüger, M., Simonsen, U., Wehland, M., Infanger, M., and Corydon, T.J. (2020). The role of SOX family members in solid tumours and metastasis. *Semin. Canc. Biol.* 67: 122–153.
- Guglielmi, B. and Werner, M. (2002). The yeast homolog of human PinX1 is involved in rRNA and small nucleolar RNA maturation, not in telomere elongation inhibition. *J. Biol. Chem.* 277: 35712–35719.
- Hamann, F., Enders, M., and Ficner, R. (2019). Structural basis for RNA translocation by DEAH-box ATPases. *Nucleic Acids Res.* 47: 4349–4362.
- Hamann, F., Schmitt, A., Favretto, F., Hofele, R., Neumann, P., Xiang, S., Urlaub, H., Zweckstetter, M., and Ficner, R. (2020). Structural analysis of the intrinsically disordered splicing factor Spp2 and its binding to the DEAH-box ATPase Prp2. *Proc. Natl. Acad. Sci. U.S.A.* 117: 2948–2956.
- Hanke, K., Hohn, O., and Bannert, N. (2016). HERV-K(HML-2), a seemingly silent subtenant - but still waters run deep. *APMIS* 124: 67–87.
- Harley, V.R., Lovell-Badge, R., and Goodfellow, P.N. (1994). Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res.* 22: 1500–1501.

- He, Y., Staley, J.P., Andersen, G.R., and Nielsen, K.H. (2017). Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA* 23: 1110–1124.
- Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell.* 45: 567–580.
- Heininger, A.U., Hackert, P., Andreou, A.Z., Boon, K.-L., Memet, I., Prior, M., Clancy, A., Schmidt, B., Urlaub, H., Schleiff, E., Sloan, K.E., et al. (2016). Protein cofactor competition regulates the action of a multifunctional RNA helicase in different pathways. *RNA Biol.* 13: 320–330.
- Hu, Y., Li, L., Seidelmann, S.B., Timur, A.A., Shen, P.H., Driscoll, D.J., and Wang, Q. K. (2008). Identification of association of common AGGF1 variants with susceptibility for Klippel–Trenaunay syndrome using the structure association program. *Ann. Hum. Genet.* 72: 636–643.
- Inesta-Vaquera, F., Chaugule, V.K., Galloway, A., Chandler, L., Rojas-Fernandez, A., Weidlich, S., Pegg, M., and Cowling, V.H. (2018). DHX15 regulates CMTR1-dependent gene expression and cell proliferation. *Life Sci. Alliance* 1: e201800092.
- Jacobs, A.M., Nicol, S.M., Hislop, R.G., Jaffray, E.G., Hay, R.T., and Fuller-Pace, F.V. (2007). SUMO modification of the DEAD box protein p68 modulates its transcriptional activity and promotes its interaction with HDAC1. *Oncogene* 26: 5866–5876.
- Jarmoskaite, I. and Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.* 83: 697–725.
- Julian, L.M., McDonald, A.C., and Stanford, W.L. (2017). Direct reprogramming with SOX factors: masters of cell fate. *Curr. Opin. Genet. Dev.* 46: 24–36.
- Jurica, M.S., Licklider, L.J., Gygi, S.P., Griorieff, N., and Moore, M.J. (2002). Purification and characterisation of native spliceosomes suitable for three-dimensional structural analysis. *RNA* 8: 426–439.
- Kastner, B., Will, C.L., Stark, H., and Lührmann, R. (2019). Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* 11: a032417.
- 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.
- Kim, J.H., Baddoo, M.C., Park, E.Y., Stone, J.K., Park, H., Butler, T.W., Huang, G., Yan, X., Pauli-Behn, F., Myers, R.M., et al. (2016). SON and its alternatively spliced isoforms control MLL complex-mediated H3K4me3 and transcription of leukemia-associated genes. *Mol. Cell.* 61: 859–873.
- Kretschmer, J., Rao, H., Hackert, P., Sloan, K.E., Höbartner, C., and Bohnsack, M.T. (2018). The m⁶A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5′-3′ exoribonuclease XRN1. *RNA* 24: 1339–1350.
- Křizová, I., Hadravová, R., Štokrová, J., Günterová, J., Doležal, M., Ruml, T., Rumlová, M., and Pichová, I. (2012). The G-patch domain of Mason-Pfizer monkey virus is a part of reverse transcriptase. *J. Virol.* 86: 1988–1998.
- Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., and Valcárcel, J. (2002). Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. *Cell* 109: 285–296.
- Lebaron, S., Papin, C., Capeyrou, R., Chen, Y.L., Froment, C., Monsarrat, B., Caizergues-Ferrer, M., Grigoriev, M., and Henry, Y. (2009). The ATPase and helicase activities of Prp43p are stimulated by the G-patch protein Pfa1p during yeast ribosome biogenesis. *EMBO J.* 28: 3808–3819.
- Li, R., Zhang, H., Yu, W., Chen, Y., Gui, B., Liang, J., Wang, Y., Sun, L., Yang, X., Zhang, Y., et al. (2009). ZIP: a novel transcription repressor, represses EGFR oncogene and suppresses breast carcinogenesis. *EMBO J.* 28: 2763–2776.
- Lin, J. and Blackburn, E.H. (2004). Nucleolar protein PinX1p regulates telomerase by sequestering its protein catalytic subunit in an inactive complex lacking telomerase RNA. *Genes Dev.* 18: 387–396.
- Lin, M.L., Fukukawa, C., Park, J.H., Naito, K., Kijima, K., Shimo, A., Ajiro, M., Nishidate, T., Nakamura, Y., and Katagiri, T. (2009). Involvement of G-patch domain containing 2 overexpression in breast carcinogenesis. *Canc. Sci.* 100: 1443–1450.
- Liu, Z., Zhang, J., Sun, Y., Perea-Chamblee, T.E., Manley, J.L., and Rabadan, R. (2020). Pan-cancer analysis identifies mutations in SUGP1 that recapitulate mutant SF3B1 splicing dysregulation. *Proc. Natl. Acad. Sci. U.S.A.* 117: 10305–10312.
- Lu, X., Göke, J., Sachs, F., Jacques, P.É., Liang, H., Feng, B., Bourque, G., Bubulya, P.A., and Ng, H.H. (2013). SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. *Nat. Cell Biol.* 15: 1141–1152.
- Lu, X., Ng, H.H., and Bubulya, P.A. (2014). The role of SON in splicing, development, and disease. *Wiley Interdiscip. Rev. RNA* 5: 637–646.
- Luo, D., Ding, S.C., Vela, A., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2011). Structural insights into RNA recognition by RIG-I. *Cell* 147: 409–422.
- Major, M.B., Roberts, B.S., Berndt, J.D., Marine, S., Anastas, J., Chung, N., Ferrer, M., Yi, X., Stoick-Cooper, C.L., von Haller, P.D., et al. (2008). New regulators of Wnt/beta-catenin signaling revealed by integrative molecular screening. *Sci. Signal.* 1: ra12.
- Mallam, A.L., Del Campo, M., Gilman, B., Sidote, D.J., and Lambowitz, A.M. (2012). Structural basis for RNA duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* 490: 121–125.
- Martin, R., Straub, A.U., Doebele, C., and Bohnsack, M.T. (2013). DEX/H-box RNA helicases in ribosome biogenesis. *RNA Biol.* 10: 4–18.
- Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nat. Struct. Mol. Biol.* 15: 435–443.
- Memet, I., Doebele, C., Sloan, K.E., and Bohnsack, M.T. (2017). The G-patch protein NF-κB-repressing factor mediates the recruitment of the exonuclease XRN2 and activation of the RNA helicase DHX15 in human ribosome biogenesis. *Nucleic Acids Res.* 45: 5359–5374.
- Miki, T.S., Richter, H., Rüegger, S., and Großhans, H. (2014). PAXT-1 promotes XRN2 activity by stabilizing it through a conserved domain. *Mol. Cell.* 53: 351–360.
- Mouffok, S., Capeyrou, R., Belhabich-Baumas, K., Joret, C., Henras, A.K., Humbert, O., and Henry, Y. (2020). The G-patch activators Pfa1 and PINX1 exhibit different modes of interaction with the Prp43 RNA helicase. *RNA Biol.*, (Epub ahead of print), <https://doi.org/10.1080/15476286.2020.1818458>.
- Mugler, C.F., Hondele, M., Heinrich, S., Sachdev, R., Vallotton, P., Koek, A.Y., Chan, L.Y., and Weis, K. (2016). ATPase activity of the

- DEAD-box protein Dhh1 controls processing body formation. *eLife* 5: e18746.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterisation of the multi-protein spliceosome complex. *Nat. Genet.* 20: 46–50.
- Nie, Y., Ran, Y., Zhang, H.Y., Huang, Z.F., Pan, Z.Y., Wang, S.Y., and Wang, Y.Y. (2017). GPATCH3 negatively regulates RLR-mediated innate antiviral responses by disrupting the assembly of VISA signalosome. *PLoS Pathog.* 13: e1006328.
- Niu, Z., Jin, W., Zhang, L., and Li, X. (2012). Tumor suppressor RBM5 directly interacts with the DExD/H-box protein DHX15 and stimulates its helicase activity. *FEBS Lett.* 586: 977–983.
- Nourbakhsh, M. and Hauser, H. (1999). Constitutive silencing of IFN- β promoter is mediated by NRF (NF-kappaB-repressing factor), a nuclear inhibitor of NF-kB. *EMBO J.* 18: 6415–6425.
- Nourbakhsh, M., Oumard, A., Schwarzer, M., and Hauser, H. (2000). NRF, a nuclear inhibitor of NF-kB proteins silencing interferon-beta promoter. *Eur. Cytokine Netw.* 11: 500–501.
- Nourbakhsh, M., Kalble, S., Dorrie, A., Hauser, H., Resch, K., and Kracht, M. (2001). The NF-kB repressing factor is involved in basal repression and interleukin (IL)-1-induced activation of IL-8 transcription by binding to a conserved NF-kB-flanking sequence element. *J. Biol. Chem.* 276: 4501–4508.
- Ozgun, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J.R., and Conti, E. (2015). The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J.* 282: 850–863.
- Pandit, S., Lynn, B., and Rymond, B.C. (2006). Inhibition of a spliceosome turnover pathway suppresses splicing defects. *Proc. Natl. Acad. Sci. U.S.A.* 103: 13700–13705.
- Pattabhi, S., Knoll, M.L., Gale, M., Jr., and Loo, Y.M. (2019). DHX15 is a coreceptor for RLR signaling that promotes antiviral defense against RNA virus infection. *J. Interferon Cytokine Res.* 39: 331–346.
- Peng, W., Guo, L., Tang, R., Liu, X., Jin, R., Dong, J.-T., Xing, C., and Zhou, W. (2019). Sox7 negatively regulates prostate-specific membrane antigen (PSMA) expression through PSMA-enhancer. *Prostate* 79: 370–378.
- Pertsch, B., Schneider, C., Gnädig, M., Schäfer, 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.
- Prabu, J.R., Müller, M., Thomae, A.W., Schüssler, S., Bonneau, F., Becker, P.B., and Conti, E. (2015). Structure of the RNA helicase MLE reveals the molecular mechanisms for uridine specificity and RNA-ATP coupling. *Mol. Cell.* 60: 487–499.
- Rappsilber, J., Ryder, U., Lamond, A., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12: 1231–1245.
- Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.-T., Urlaub, H., Kastner, B., et al. (2016). Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* 353: 1399–1405.
- Robert-Paganin, J., Halladjian, M., Bland, M., Lebaron, S., Delbos, L., Chardon, F., Capeyrou, R., Humbert, O., Henry, Y., Henras, A.K., et al. (2017). Functional link between DEAH/RHA helicase Prp43 activation and ATP base binding. *Nucleic Acids Res.* 45: 1539–1552.
- Rossler, O.G., Straka, A., and Stahl, H. (2001). Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res.* 29: 2088–2096.
- Roy, J., Kim, K., Maddock, J.R., Anthony, J.G., and Woolford, J.L. (1995). The final stages of spliceosome maturation require Spp2p that can interact with the DEAH box protein Prp2p and promote step 1 of splicing. *RNA* 1: 375–390.
- Sardana, R., Liu, X., Granneman, S., Zhu, J., Gill, M., Papoulas, O., Marcotte, E. M., Tollervey, D., Correll, C.C., and Johnson, A.W. (2015). The DEAH-box helicase Dhr1 dissociates U3 from the pre-rRNA to promote formation of the central pseudoknot. *PLoS Biol.* 13: e1002083.
- Schmitt, A., Hamann, F., Neumann, P., and Ficner, R. (2018). Crystal structure of the spliceosomal DEAH-box ATPase Prp2. *Acta Crystallogr. D Struct. Biol.* 74: 643–654.
- Schütz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 105: 9564–9569.
- Semlow, D.R., Blanco, M.R., Walter, N.G., and Staley, J.P. (2016). Spliceosomal DEAH-Box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell* 164: 985–998.
- 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.
- Sharma, A., Markey, M., Torres-Muñoz, K., Varia, S., Kadakia, M., Bubulya, A., and Bubulya, P.A. (2011). Son maintains accurate splicing for a subset of human pre-mRNAs. *J. Cell Sci.* 124: 4286–4298.
- Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 97: 8868–8873.
- Silverman, E.J., Maeda, A., Wei, J., Smith, P., Beggs, J.D., and Lin, R.-J. (2004). Interaction between a G-Patch protein and a spliceosomal DEXD/H-Box ATPase that is critical for splicing. *Mol. Cell Biol.* 24: 10101–10110.
- Sloan, K.E. and Bohnsack, M.T. (2018). Unravelling the mechanisms of RNA helicase regulation. *Trends Biochem. Sci.* 43: 237–250.
- Sloan, K.E., Knox, A.A., Wells, G.R., Schneider, C., and Watkins, N.J. (2019). Interactions and activities of factors involved in the late stages of human 18S rRNA maturation. *RNA Biol.* 16: 196–210.
- Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev.* 31: 1370–1381.
- Steimer, L. and Klostermeier, D. (2012). RNA helicases in infection and disease. *RNA Biol.* 9: 751–771.
- Studer, M.K., Ivanović, L., Weber, M.E., Marti, S., and Jonas, S. (2020). Structural basis for DEAH-helicase activation by G-patch proteins. *Proc. Natl. Acad. Sci. U.S.A.* 117: 7159–7170.
- Švec, M., Bauerová, H., Pichová, I., Konvalinka, J., and Střišovský, K. (2004). Proteinases of betaretroviruses bind single-stranded nucleic acids through a novel interaction module, the G-patch. *FEBS Lett.* 576: 271–276.
- Takash, W., Cañizares, J., Bonneaud, N., Poulat, F., Mattéi, M.G., Jay, P., and Berta, P. (2001). SOX7 transcription factor: sequence, chromosomal localisation, expression, transactivation and

- interference with Wnt signalling. *Nucleic Acids Res.* 29: 4274–4283.
- 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.
- Tauchert, M.J., Fourmann, J.B., Lührmann, R., and Ficner, R. (2017). Structural insights into the mechanism of the DEAH-box RNA helicase Prp43. *eLife* 6: e21510.
- Tian, X.L., Kadaba, R., You, S.A., Liu, M., Timur, A.A., Yang, L., Chen, Q., Szafranski, P., Rao, S., Wu, L., et al. (2004). Identification of an angiogenic factor that when mutated causes susceptibility to Klippel-Trenaunay syndrome. *Nature* 427: 640–645.
- Toczyłowska-Socha, D., Zielinska, M.M., Kurkowska, M., Astha-Almeida, C.F., Stefaniak, F., Purta, E., and Bujnicki, J.M. (2018). Human RNA cap1 methyltransferase CMTr1 cooperates with RNA helicase DHX15 to modify RNAs with highly structured 5' termini. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373: 20180161.
- Tokita, M.J., Braxton, A.A., Shao, Y., Lewis, A.M., Vincent, M., Küry, S., Besnard, T., Isidor, B., Latypova, X., Béziau, S., et al. (2016). *De novo* truncating variants in SON cause intellectual disability, congenital malformations, and failure to thrive. *Am. J. Hum. Genet.* 99: 720–727.
- Tsai, R.-T., Fu, R.-H., Yeh, F.-L., Tseng, C.-K., Lin, Y.-C., Huang, Y.-H., and Cheng, S.-C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes Dev.* 19: 2991–3003.
- Walbot, 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.* 29: 2194–2204.
- Wan, R., Yan, C., Bai, R., Lei, J., and Shi, Y. (2017). Structure of an intron lariat spliceosome from *Saccharomyces cerevisiae*. *Cell* 171: 120–132, e12.
- Wang, Y., Gogol-Döring, A., Hu, H., Fröhler, S., Ma, Y., Jens, M., Maaskola, J., Murakawa, Y., Quedenau, C., Landthaler, M., et al. (2013). Integrative analysis revealed the molecular mechanism underlying RBM10-mediated splicing regulation. *EMBO Mol. Med.* 5: 1431–1442.
- Warkocki, Z., Schneider, C., Mozaffari-Jovin, S., Schmitzová, J., Höbartner, C., Fabrizio, P., and Lührmann, R. (2015). The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes Dev.* 29: 94–107.
- Will, C.L. and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3: a003707.
- Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R., and Pillai, R.S. (2017). Regulation of m6A transcripts by the 3'→5' RNA helicase YTHDC2 is essential for a successful meiotic program in the mammalian germline. *Mol. Cell.* 68: 374–387.
- Wurm, J.P. and Sprangers, R. (2019). Dcp2: an mRNA decapping enzyme that adopts many different shapes and forms. *Curr. Opin. Struct. Biol.* 59: 115–123.
- Yan, C., Wan, R., and Shi, Y. (2019). Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harb. Perspect. Biol.* 11: a032409.
- Yang, Q. and Jankowsky, E. (2005). ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* 44: 13591–601.
- Yang, Q. and Jankowsky, E. (2006). The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol.* 13: 981–986.
- Yang, Q., Del Campo, M., Lambowitz, A.M., and Jankowsky, E. (2007). DEAD-box proteins unwind duplexes by local strand separation. *Mol. Cell.* 28: 253–263.
- Yoshimoto, R., Kataoka, N., Okawa, K., and Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res.* 37: 891–902.
- Young, C.L., Khoshnevis, S., and Karbstein, K. (2013). Cofactor-dependent specificity of a DEAD-box protein. *Proc. Natl. Acad. Sci. U.S.A.* 110: E2668–E2676.
- Yu, W., Li, R., Gui, B., and Shang, Y. (2010). sZIP, an alternative splice variant of ZIP, antagonizes transcription repression and growth inhibition by ZIP. *J. Biol. Chem.* 285: 14301–14307.
- Yuan, K., Li, N., Jiang, K., Zhu, T., Huo, Y., Wang, C., Lu, J., Shaw, A., Thomas, K., Zhang, D., et al. (2009). PinX1 is a novel microtubule-binding protein essential for accurate chromosome segregation. *J. Biol. Chem.* 284: 23072–23082.
- Zang, S., Lin, T.Y., Chen, X., Gencheva, M., Newo, A.N., Yang, L., Rossi, D., Hu, J., Lin, S.B., Huang, A., et al. (2014). GPKOW is essential for pre-mRNA splicing *in vitro* and suppresses splicing defect caused by dominant-negative DHX16 mutation *in vivo*. *Biosci. Rep.* 34: e00163.
- Zhang, Y., Cheng, Y.T., Bi, D., Palma, K., and Li, X. (2005). MOS2, a protein containing G-patch and KOW motifs, is essential for innate immunity in *Arabidopsis thaliana*. *Curr. Biol.* 15: 1936–1942.
- Zhang, J., Ali, A.M., Lieu, Y.K., Liu, Z., Gao, J., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2019). Disease-causing mutations in SF3B1 alter splicing by disrupting interaction with SUGP1. *Mol. Cell.* 76: 82–95, e7.
- Zhou, X.Z. and Lu, K.P. (2001). The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor. *Cell* 107: 347–359.
- Zhou, X.Z., Huang, P., Shi, R., Lee, T.H., Lu, G., Zhang, Z., Bronson, R., and Lu, K.P. (2011). The telomerase inhibitor PinX1 is a major haploinsufficient tumor suppressor essential for chromosome stability in mice. *J. Clin. Invest.* 121: 1266–1282.
- Zorn, A.M., Barish, G.D., Williams, B.O., Lavender, P., Klymkowsky, M.W., and Varmus, H.E. (1999). Regulation of Wnt signaling by Sox proteins: XSox17 alpha/beta and XSox3 physically interact with β -catenin. *Mol. Cell.* 4: 487–498.