Getting to the end of RNA: structural analysis of protein recognition of 5′ and 3′ termini

Stephen Curry^{1*}, Olga Kotik-Kogan¹, Maria R. Conte², Peter Brick¹

¹*Biophysics Section, Blackett Laboratory, Imperial College, Exhibition Road, London, SW7 2AZ, UK*

²*Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1U, UK*

doi:[10.1016/j.bbagr.2009.07.003](https://doi.org/10.1016/j.bbagr.2009.07.003)

ABSTRACT

*Corresponding author:

Stephen Curry
Biophysics Section
Blackett Laboratory
Imperial College
London SW7 2AZ, UK.

Email: s.curry@imperial.ac.uk
Tel: +44-20-7594-7632
Fax: +44-20-7589-0191

The specific recognition by proteins of the 5′ and 3′ ends of RNA molecules is an important facet of many cellular processes, including RNA maturation, regulation of translation initiation and control of gene expression by degradation and RNA interference. The aim of this review is to survey recent structural analyses of protein binding domains that specifically bind to the extreme 5′ or 3′ termini of RNA. For reasons of space and because their interactions are also governed by catalytic considerations, we have excluded enzymes that modify the 5′ and 3′ extremities of RNA. It is clear that there is enormous structural diversity among the proteins that have evolved to bind to the ends of RNA molecules. Moreover, they commonly exhibit conformational flexibility that appears to be important for binding and regulation of the interaction. This flexibility has sometimes complicated the interpretation of structural results and presents significant challenges for future investigations.

1. Introduction

Specific binding of proteins to the 5′ and 3′ ends of RNA is crucial for many aspects of the production and use of RNA molecules in living cells. Within the past several years, structural studies have greatly increased our understanding of the molecular details of these protein-RNA recognition events. They involve a wide variety of RNA molecules, in single and double-stranded RNA contexts, and play important roles in many different cellular functions, including RNA maturation, translational control, regulation of gene expression and viral defense mechanisms (and counter-measures).

This topic has already been touched upon in several excellent reviews, usually devoted to a specific functional context, for example, 5′ cap binding [1, 2], RNA interference [3-5] or ribonuclease function [6]. Although in many cases, the 5′ and 3′ termini of RNA undergo modification in the cell, requiring specific recognition of the ends to be processed by the modifying enzymes (*e.g.* polymerases, CCA-adding enzymes, aminoacyl tRNA synthetases, pyrophosphohydrolases, 5′-capping and decapping enzymes), these will not be reviewed here. In part this is for reasons of space but it also reflects the fact that recognition events involved in enzymatic processing are more dynamic—often coupled to the binding of other substrates of the modification reaction—and arguably reflect a distinct class of binding interaction. Here we will review structural analyses of binding domains that have evolved to specifically recognise the extreme 5′ or 3′ ends of RNA,

either separately or together, with a particular focus on more recent results.

2. 5′ cap recognition

2.1. The architecture of cap-binding proteins.

The 5′ cap structure (m⁷GpppN) is formed by the enzymatic addition of N7-methylated guanosine (G) to the first base (N) of eukaryotic RNA polymerase II transcripts such as mRNA and U snRNA. Capping is essential for mRNA transport, processing, translation and protection from degradation by 5′ exonucleases [1, 2]. These functions are mediated by several proteins that interact specifically with the 5′ cap structure. The molecular details of these binding interactions have been obtained from extensive structural and functional studies of cellular and viral proteins. Interestingly, these proteins have disparate origins and very different structures (Fig. 1); however, they share a similar aromatic cap-binding pocket, and represent an interesting case of convergent evolution [2].

One of the best studied 5′ cap-binding proteins is the eukaryotic initiation factor 4E (eIF4E), which is essential for the initiation of translation of capped mRNAs [7]. It associates with eIF4G, which facilitates cap binding, and eIF4A to form the eIF4F complex that helps to recruit the 40S ribosomal subunit to the mRNA to initiate translation [8]. In mammals the eIF4E family contains three members: eIF4E (eIF4E-1), 4EHP (eIF4E-2) and eIF4E-3 [9]. Human 4E

homology protein (h4EHP) is 30% identical in amino acid sequence to eIF4E and binds both the m⁷G cap on mRNA and the regulator protein, 4E binding protein 1 (4E-BP1), but — intriguingly — does not interact with eIF4G and its physiological role remains unclear [9–11]. The structures of eIF4E and h4EHP in the apo and m⁷GTP-bound forms are very similar, possessing a central curved β -sheet flanked by three helices on a convex ‘outer’ surface and an m⁷GTP binding pocket on the ‘inner’ concave side [12–15] (Fig. 1c). Most of the residues involved in cap binding are conserved between the two proteins [10, 14].

Prior to reaching eIF4E in the cytoplasm, the mRNA in the nucleus is partnered by the nuclear Cap Binding Complex, CBC. This heterodimeric complex consists of the 5′ cap binding protein, CBP20, in association with the larger ancillary protein, CBP80. CBP20 contains an RNA Recognition Motif (RRM) that binds to the m⁷GpppG cap structure using the central -sheet platform while engaging CBP80 via its dorsal helical surface and an N-terminal extension [16–18] (Fig. 1a), an interaction that is necessary for cap-binding [19]. Exemplifying the highly adaptable properties of the RRM platform, CBP20 binds capped RNA across its β -sheet RNA-binding surface in the opposite direction to that observed for RNA ligands of all other RRM domains that have been examined structurally [17, 20]. CBP80 has also been shown to associate with another cap-interacting protein, the poly-A-specific ribonuclease (PARN) [21]. PARN, a specific 3′ exoribonuclease involved in mRNA decay, is unique among the poly(A) nucleases since it interacts with both the 3′ en[22]d and the 5′ cap of the mRNA during deadenylation [23–25]. PARN contains an RRM domain that specifically recognises the 5′ cap and recent structural investigations have revealed significant differences with other cap binding proteins [26–28] (see below). In particular, the PARN RRM differs strikingly from the CBP20 RRM since it binds the cap not on the canonical β -sheet surface, but at the extreme edge of the sheet, in a cavity between β 2 and β 1 (see Fig. 6 in [27]).

The cap-binding complex is also involved in the export of the m⁷G-capped transcripts of U1, U2, U4 and U5 snRNAs. In the cytoplasm, the m⁷G cap of these U snRNAs are hypermethylated to a 2,2,7-trimethyl-guanosine cap form (m³G). This modification allows the specific binding of the protein, snurportin1, which is necessary for re-import of the U snRNPs to the nucleus where they are assembled into the spliceosome [22]. Curiously the cap-binding domain of snurportin1 resembles the mRNA guanylyl-transferase enzyme that uses GTP to add the guanosine base to mRNA at the outset of cap-formation [29] (Fig. 1d).

In addition to cellular factors, specific 5′ cap recognition is also achieved by viral proteins. In vaccinia virus, a double-stranded DNA poxvirus, the cap-dependent 2′-O-methyltransferase VP39 recognises its own capped mRNA [30]. In the later stages of viral mRNA cap synthesis, VP39 transfers a methyl group to the ribose 2′-OH of the first transcribed residue of the capped mRNA to produce a ‘cap 1′ type structure (m⁷Gppp(G^m/A^m)). VP39 has a compact globular α/β domain similar to the catalytic domains of other methyltransferases (see Fig. 1 in [31]). The core of the domain specifically binds the m⁷G cap of the mRNA to one side of the active site, thereby placing the first transcribed base in the enzyme active site [32, 33].

In contrast to vaccinia virus, influenza virus does not synthesise its own caps but steals them from cellular mRNA in a process known as cap-snatching that is performed by the heterotrimeric viral polymerase, formed from PA, PB1 and PB2 subunits. Immediately prior to transcription of viral mRNA, the 5′ cap of host mRNA is bound by PB2 [34], and then cleaved at nucleotide 10–13 by the endonuclease subunit, PA [35, 36]. The minimal cap-binding fragment of PB2 has a compact α/β architecture which does not resemble any known fold [34] (Fig. 1b). The m⁷G moiety binds in a pocket on PB2 at the interface of the concave side of the -sheet and the flanking -helices [34].

2.2. Structural features of protein 5′ cap recognition.

Despite the architectural differences between the 5′ cap-binding proteins, the specific mode of recognition of the cap follows a common theme, invariably involving base-stacking interactions with aromatic side chains [1, 2, 14, 34]. In most cases the m⁷G base is stacked between two aromatic amino acids in a ‘cation--sandwich’, where the delocalised positive charge of the base that arises from methylation of N7 of the base forms strong electrostatic interactions with the aromatic electrons (Fig. 1e). The pair of aromatic residues involved in this sandwich stacking arrangement have been identified as Trp 102/Trp 56 in eIF4E, Tyr 20/Tyr 43 in CBP20, Tyr 22/Phe 180 in VP39, Tyr 78/Trp 124 in h4EHP and Phe 404/His 357 in PB2. With one exception, the plane-parallel stacking of the methylated base between the two aromatic rings is very similar, suggesting that this sandwich arrangement exerts a powerful grip on the methylated base. Substitution of the aromatic groups with aliphatic residues decreases specificity and affinity for the cap structure, underscoring the importance of the aromatic stacking interaction for specific cap recognition [17, 33, 34, 37]. Curiously, the two-sided sandwich arrangement is rarely observed in proteins that specifically bind to unmodified RNA bases; the best known examples are the PUF proteins, though even here Tyr/Arg sandwich pairs are more common than pairs with two ring-containing amino-acid side chains [38].

The Phe 404/His 357 pair in the PB2 subunit of influenza virus polymerase is an interesting variant of this type of cap-binding pocket since the benzyl ring of the Phe is tilted 30° to the plane of the cap base and a histidine side chain acts as the second stacking group (Fig. 1b). This latter feature is unusual even among different strains of influenza, since the PB2 proteins of types B and C influenza have Tyr at this position [34].

PARN is the only m⁷G-binding protein identified thus far in which the canonical ‘cation--sandwich’ signature is absent, since the methylated m⁷G base stacks on a single tryptophan residue (Trp 475) [26]. Intriguingly, the solution structure of the PARN RRM complexed with m⁷GpppG shows that the second G can provide the top layer of the sandwich by stacking on the m⁷G [27, 28]. This variant of the cation--type binding interaction may still account for the discrimination between methylated and unmethylated ligands, as GTP was found to bind at least 100-fold more weakly than m⁷GTP [26].

Another common characteristic of cap recognition is the formation of hydrogen bonds between nitrogen atoms N1 and/or N2 of

Fig. 1: Cap-binding proteins are structurally diverse. Structures of (a) CBP20 [17] (note that, for simplicity, the CBP80 subunit is not shown), (b) Influenza virus PB2 [34], (c) eIF4E [45] and (d) snurportin1 [29] are depicted. In each case the cap bound to the protein is depicted in a stick representation with atoms colour-coded by type (C – grey; O – red; N – blue; P – orange; S – yellow). The cap-stacking residues are labelled. Close-up views showing the details of cap-binding are shown for (e) eIF4E and (f) snurportin1. Selected amino-acid side chains are shown as sticks with atoms coloured by type as above except that carbon atoms take on the colour of the protein.

the base and carboxylate groups of key acidic residues (Asp or Glu) in the cap-binding pocket (Fig. 1e). At least one essential acidic amino acid has been identified in all the cap binding proteins that exhibit the cap-sandwich mode of binding. The importance of this interaction is demonstrated by the finding that

methylation at N2 reduces the affinity of CBC for m⁷G by about 100-fold [39]. The PARN RRM also forms hydrogen bonds with the N1 and N2 atoms of the m⁷G but is again an outlier since, instead of acidic side chains, it deploys two backbone carbonyl groups to make the interactions [26].

The binding of the methyl group of m⁷G generally appears to be stabilised by van der Waals interaction with surrounding side chains and/or backbone atoms within the binding pocket [1]. In the case of CBP20 [17], evidence suggests that direct interactions with the methyl moiety itself are not important for specificity [17, 34]. Rather, it is likely that the most significant impact of methylation is to enhance the cationic nature of the base, which strengthens the stacking interactions [40]. Nonetheless, an interesting recent study has shown that modification of the methyl group on N7 can enhance the binding affinity of m⁷G caps and may facilitate the development of eIF4E as a potential drug target in diseases such as cancer [41].

The structural data reveal a number of interactions with other features of the m⁷G cap, though their use varies between the proteins that have been investigated. For example ribose hydroxyl moieties of the m⁷G are hydrogen-bonded to the protein in CBP20 but not in eIF4E, VP39 or PB2 where they are solvent exposed [13, 34, 42]. Interactions with the triphosphate moiety of the cap are more common, mostly involving Arg and Lys side chains, though the geometry of the interactions exhibits considerable variation [1, 17, 28, 33, 34, 43].

The crystal structure of snurportin1 complexed with m³GpppG revealed a distinct mode of cap binding from the proteins discussed above which seems likely to be due to the need to discriminate m³G from m⁷G caps [29] (Fig. 1f). Like the cap-binding domain of PARN [27, 28], the cation- π sandwich is completed by stacking the downstream G of the transcript onto the cap [29].

The dimethylated N2 group of the m³G cap packs against a second Trp residue (Trp 107) which lies on the surface of snurportin1, perpendicular to Trp 276 (Fig. 1f). This interaction is clearly important for specific recognition of the m³G cap (which binds at least 14-fold tighter than an m⁷G cap) since the substitution of Trp 107 by Ala reduces the binding affinity and renders the protein unable to distinguish m³G from

m⁷G [29]. Although N2 is modified by methylation in m₃G, the protein maintains a specific hydrogen bond to the N1 group of the cap (Fig. 1f), an interaction that is observed in m⁷G cap-binding proteins; however, the general conservation of hydrogen-bonding groups in proteins that interact specifically with *both* the N1 and N2 groups of m⁷G caps (as shown on eIF4E in Fig. 1e) explains their ability to discriminate against m₃G-capped U snRNAs.

2.3. Affinity and specificity of the cap-binding proteins.

An intriguing and mysterious aspect of cap recognition is that, despite many evident similarities, the cap binding proteins discussed above have different affinities for m⁷G caps. In particular, whereas the CBC and eIF4E bind m⁷G caps with nanomolar affinity, VP39, PARN, h4EHP and PB2 have dissociation constants in the micromolar range. In the case of influenza virus PB2, it has been proposed that the imperfect aromatic sandwiching due to the smaller aromatic ring of His 357 may explain its lower binding affinity and weaker discrimination between methylated m⁷GTP and unmethylated GTP [34]. Qualified support for this notion comes from the finding that substitution of His 357 with Trp increases the PB2 affinity for the cap seven-fold [34]. Equally, the lower affinity of PARN for the cap (micromolar K_d) may be due to the fact that the 'cation- π -sandwich' interaction can only be completed by the RNA [27, 28].

It remains unclear why VP39 and h4EHP have much lower affinity for cap than eIF4E and CBC, even though they possess similar cap-sandwiching residues. The 200-fold difference in cap affinity between eIF4E and h4EHP is the most difficult to explain since their overall structures are extremely similar: the root-mean-square difference in C α positions is only 1 Å [14]. eIF4E uses a Trp/Trp aromatic pair to sandwich the m⁷G cap whereas h4EHP has a Trp/Tyr pair; however, this variation does not correlate with affinity and indeed mutation of eIF4E to a Trp/Tyr stacking pair had only a marginal effect on the cap-binding affinity, actually enhancing it about 1.5-fold [11]. Overall, the contacts that the methylated base makes with both proteins are remarkably similar, with only minor differences in the connections made by the ribose and the triphosphate moieties. The main differences between the two structures occur in the N and C-termini and in the 1-2 loop which forms one wall of the m⁷GTP binding site pocket; this loop is marginally longer in h4EHP and contains an additional short helix [14]. Whether these can account for the large difference in binding affinity remains to be determined.

Interaction of 5' capped mRNAs with the cap-binding proteins may be affected by portions of the molecule downstream of the cap, though the particular features of the binding sites responsible for these effects are not entirely clear. For example, whereas PB2 appears unable to discriminate between m⁷G and m⁷GpppG [34], the addition of the second guanosine impacts binding to eIF4E and the CBC, though in different ways. The CBC binds to m⁷GpppG with an affinity that is one to two orders of magnitude tighter than for m⁷GTP [39, 44]. The crystal structure suggests that the affinity enhancement may arise, at least in part, by the stacking of the second guanosine base on the conserved Tyr 138 (Fig. 1a). However, this stacking interaction is not consistently observed in the four independent structures [17, 18] and, fur-

thermore, substitution of Tyr 138 by Ala only modestly reduces binding [17, 39].

Nor can the finding that m⁷GpppG binds 10-20 times more weakly to eIF4E than m⁷GTP be definitively explained by crystallographic analysis since, although the co-crystal structure of eIF4E-m⁷GpppA shows that the downstream base (adenosine) interacts with a C-terminal loop on the protein (residues 203-210) [45] (Fig. 1c), in the crystal structure of the eIF4E-m⁷GpppG complex the second base is disordered [43]. It may be that variations in crystal contacts are responsible for the different structural results, a potential artefact of the method that can complicate the interpretation of crystal structures, particularly for parts of ligands that are situated on the periphery of the protein structure.

2.4. Regulation of cap binding proteins.

The difficulties discussed above highlight some of the limitations of our current understanding of the structural basis of cap recognition. But the picture is more complex still since additional protein co-factors that associate with cap-binding proteins *in vivo* may also impact on their binding affinity. At one end of the spectrum, CBP20 appears to bind the cap only in the presence of CBP80 [19], even though the larger subunit makes no direct interactions with the RNA [17, 18]. The effect appears to be cooperative: in the absence of cap, only the core RRM domain of CBP20 (residues 33-125) is ordered but, upon cap binding, Tyr20 within the N-terminal extension of CBP20 takes up a position as the top layer of the aromatic sandwich and secures the cap in place [17, 18] (Fig. 1a). This interaction appears to favour general stabilisation of the N-terminus of CBP20, which adopts an extended conformation lying across the upper surface of the CBP80 subunit. The binding of some capped RNAs to the CBC may also be regulated by other factors. For example, the export of certain capped U snRNAs by association with the CBC is dependent on the protein co-factor PHAX, which associates with CBC and has an RNA binding domain [46]. Conceivably PHAX may bind to the RNA downstream of the cap and thereby influence the binding of the cap to CBP20.

In contrast to CBP20, VP39 and PB2 can bind to the m⁷G cap as isolated polypeptides even though they are associated with other proteins *in vivo* that may regulate cap binding [1]. Recent studies on the polymerase from influenza virus suggest there may be allosteric communication between the PB2 and the PB1 and/or PA subunits, that could explain the functional cooperativity observed between binding of the 3' and 5' ends of viral RNA and cap-snatching from host mRNAs (see [47] and references therein).

The most intensively studied case of regulation of cap binding through interactions with other proteins centres on eIF4E [7]. Both eIF4G and regulators such as 4E-BP1 are reported to enhance the affinity of eIF4E for m⁷G caps, though the precise magnitude of the observed effect varies between different laboratories [48-51]. The strongest effects have been obtained with m⁷G-capped mRNA; for example, it has been shown that 4E-BP1 can retard the rate of dissociation of eIF4E from m⁷G-capped mRNA by about two orders of magnitude [50].

What is the structural basis for this apparent cooperativity? Solution and crystal structures of apo-eIF4E [15, 52, 53] revealed

considerable structural flexibility in the surface loops surrounding the cap-binding site and on the dorsal surface that binds eIF4G and 4E-BP1 [54, 55]. Most strikingly Trp 56 and Trp 102, which sandwich the m⁷G cap in eIF4E [13, 56], are both flipped outwards in the absence of ligand [13, 15, 52, 53]. This flexibility is likely to mediate allosteric communication between the cap-binding site and the back of eIF4E, allowing eIF4G and 4E-BP1 (and its homologues) to enhance cap-binding [48, 49, 51]. Both proteins interact with eIF4E via a conserved core motif (Tyr-X-X-X-Leu-Φ, where X is any amino acid and Φ is Leu, Met or Phe); however, binding of short peptides containing this motif is insufficient to enhance the affinity of eIF4E for cap [43, 51]. This observation is consistent with structural analyses that show that peptide fragments containing this motif bind to the dorsal surface of cap-bound eIF4E structures but cause no significant conformational changes [13, 54]. Strikingly, the solution structure of a yeast ternary complex formed from m⁷GDP, full-length eIF4E and a fragment of eIF4G (393-490) revealed highly cooperative folding of the N-terminus of eIF4E and the eIF4G domain. This resulted in a much larger binding interface between the two proteins which clearly offers scope for greater allosteric impact [57]. However, the precise structural differences at the cap-binding site are difficult to discern, in part because the high molecular weight and poor solubility of complex limited the accuracy of the structure determination, making it difficult to compare in meaningful detail with the eIF4E-m⁷GDP binary complexes [12, 13, 57]. The amino acids that form the eIF4E-eIF4G interface are quite well conserved between yeast and mammals: in eIF4E, 28/30 interface amino acids are similar, though for eIF4G this ratio is only 12/29 [57]. While it is possible that a similar eIF4E-eIF4G complex may form in mammals, this has not yet been demonstrated; initial hints suggest that the nature of the interaction may actually be different [58].

Moreover, it is not clear whether binding of the mammalian 4E-BPs induces a similar co-folding of the N-terminus of eIF4E. Although they are a similar size to the eIF4E-binding domain of eIF4G (~110 amino acids) and have the conserved heptameric binding motif, 4E-BPs are otherwise unrelated in sequence to eIF4G. So far structural work has been confined to peptides derived from 4E-BP1 that contain the core binding motif but are only 17 or 34 amino acids long and do not enhance cap binding [54, 55]. The structures formed in each case are very similar; use of the longer peptide revealed no significant additional structural details and gave no hint of a possible co-folding mechanism. It seems likely that the allosteric mechanism of 4E-BPs differs from that of eIF4G, but further work is necessary to support this notion.

More recent results have clouded the picture and suggest that additional factors may influence cap binding by eIF4E. A study using human proteins found that binding of eIF4G(557-646) (which is equivalent to the fragment used in the yeast complex), did not consistently enhance binding of m⁷GpppG to eIF4E [59]; an enhancement effect of the eIF4G domain was only observed with recombinant eIF4E expressed in *E. coli* that had been re-folded after purification from inclusion bodies. This echoes an earlier report that the purification protocol used for eIF4E can drastically affect binding measurements [43] and suggests that

preparation methods will need careful consideration in future work. A separate investigation that also used human proteins showed that eIF4G only stimulated cross-linking of cap to eIF4E if the RNA-binding domain of eIF4G was present [58]. Although this result contradicts earlier reports that yeast eIF4G fragments lacking any RNA binding activity could enhance cap-binding by eIF4E [51] and needs to be confirmed using techniques that probe binding affinity more directly, it reasserts the possibility that (as with CBC) association of the complex with the downstream RNA may assist with binding of cap to eIF4E.

Although our understanding of 5' cap recognition has advanced enormously as a result of detailed structural investigations and affinity measurements, there are still many interesting questions to answer, particularly with regard to dissecting the fine detail of co-factor and RNA modulation of cap binding. Tackling these problems will present stiff challenges but is likely to produce structural insights that will relate more closely to the *in vivo* function of these cap-binding proteins.

3. Specific recognition of RNA 3'-ends by La and NSP3

Whereas 5' cap recognition depends largely on interactions with a single modified nucleotide, the best known examples of specific recognition of the 3' ends of RNA involve protein interactions with a short stretch of nucleotides in a sequence-specific manner. Structural studies have revealed some interesting similarities and differences between the mode of binding of the 3'-termini of RNA by two unrelated proteins, the human La protein and rotavirus NSP3. A further interesting example, binding of PAZ domains to the 3' ends of interfering RNAs, will be considered in the next section.

3.1. Recognition of RNA 3' ends by La.

The eukaryotic La protein is predominantly localised in the nucleus of eukaryotic cells [60, 61] where it aids processing and maturation of non-coding RNA transcripts. In metazoans it binds specifically to the UUU-OH sequences found at the 3' termini of RNA polymerase III RNA transcripts, including pre-tRNA, 5S RNA and SRP RNA, [62-64]. In yeast cells, La binds to the 3' poly(U) ends of RNA polymerase II transcripts, such as snRNAs, that are generated by endonucleolytic processing [65, 66]. The primary function of the protein is to protect the 3'-termini of these transcripts from digestion by 3'-5' exonucleases [63, 67], though it has additional roles in regulating RNA maturation pathways and aiding assembly of the transcripts into functional RNA-protein complexes [64, 66-70].

The La protein is modular in construction. Human La contains a La motif and two RRM in its N-terminal half and an unstructured C-terminal region [60, 61, 71, 72]. Yeast La has a somewhat simpler form that lacks the second RRM [60]. The ability to bind 3'-UUU-OH sequences is nevertheless retained across these species since this function is performed by the common N-terminal domain (LaNTD; residues 1-194) that contains the La motif and RRM1 [71, 73].

Fig. 2: Contrasting binding sites in La and rotavirus NSP3 for RNA 3' ends. (a) Surface representation of LaNTD complexed with the RNA oligomer AUAUUU [74]. The protein is coloured by domain (Green – La motif; light-orange – RRM1). The RNA is shown as sticks, colour-coded by atom-type as in Figure 1; for clarity the first three bases are omitted. Arrows indicate the location of the canonical RNA binding surfaces on the La motif and RRM domains. (b) Close-up view of LaNTD-RNA interactions for the 3' terminal three bases of AUAUUU [74] in the same orientation as panel a. Hydrogen bonds are shown as grey dashed lines. (c) Surface representation of the rotavirus NSP3 dimer complexed with the RNA oligomer GUGACC [82]. (d) Close-up view of the NSP3-RNA interactions in the same orientation as panel c. The first base of the RNA was disordered and not included in the refined model [82].

The crystal structures of human LaNTD bound to a variety of RNAs (all terminating with at least three U's) show a common mode of binding that nevertheless displays an interesting degree of plasticity [74, 75]. The two domains, which are connected by a flexible linker in solution [71], clamp onto the 3'-end of the RNA, forming a sequence-specific pocket at the inter-domain cleft [74, 75] (Fig. 2a). The bound RNA 3'-end adopts a tight turn that is generally stabilised by the stacking of U₃ onto U₁ (nucleotides are numbered negatively starting with U₁ at the 3'-end), both bases only making contact with the La motif. U₂ is splayed out away from this stacked pair and contacts both the La motif and RRM1 domains.

This mode of binding was initially observed for duplex RNA with a three-uridylylate overhang at the 3'-end that formed unexpectedly during crystallisation [75]. Subsequent studies showed that the presence of the duplex prevented the two domains from clamping down completely on the RNA 3'-terminus; co-crystallisation with purely single-stranded RNA ligands allowed a closer approach of the La motif to RRM1 and the formation of fully intimate contacts with the RNA [74]. Although the overall U-turn conformation was retained, there were significant adjustments of the RNA ligand that resulted in full — rather than offset — stacking of U₃ on U₁ and formation of additional interactions with the protein. Perhaps surprisingly, given the high affinity of the interaction with RNAs ending in oligo(U) sequences ($K_D \sim 20$ nM for human LaNTD - [74, 75]), the binding cleft remains rather open. In the complex of LaNTD with a AUUUU oligomer this openness was observed to allow an alternative binding conformation of the 3'-end of a single-stranded RNA ligand in which U₄ rather than U₃ stacked onto the U₁ nucleotide at the 3'-end. This arrangement involves a wider U-turn at the 3'-end in which U₃ flips out towards RRM1, but does not make any contact with it [74].

Curiously, in view of the well-characterised specificity of La for the UUU-OH 3'-ends of RNA, the network of base-specific interactions with the protein is less extensive than one might expect (Fig. 2b). U₁ interacts exclusively with the La motif, which only makes specific polar contacts with the backbone of the nucleotide. The 2' and 3'-OH groups from the ribose ring of U₁ are specifically recognised by hydrogen bonds to the side-chain carboxylate of Asp 33, a key residue in the binding site [73, 75],

while the U₁ phosphate group makes three hydrogen bonds to the La motif. The U₁ base stacks directly onto the aromatic side chain of Phe 35. Although it packs closely enough to make contacts with the body of the La motif, there are no base-specific interactions with the protein. Consistent with this structural observation, variation of the terminal base only reduced the binding affinity very modestly [75].

The penultimate nucleotide, U₂, makes intimate contacts with both the La motif and RRM1 in the deepest recess of the binding cleft; in striking contrast to U₁, it is the nucleotide base of U₂ rather than the backbone that makes specific hydrogen bonds to the LaNTD (Fig. 2b). The O2 and O4 atoms of the U₂ pyrimidine ring hydrogen bond to the side-chain amide of Gln 20 (La motif)

and the main-chain amide of Ile 140 (RRM1). The pyrimidine ring of U₂ is pinned between Tyr 23, a La motif residue, and the hydrophobic side chain of Leu 124, an arrangement that partially recapitulates the sandwiching mode of base recognition seen in cap-binding proteins (see above). It is evident from the structure that the specific recognition of U₂ arises from a cooperative set of protein-RNA and protein-protein interactions involving residues from *both* domains to form a tightly defined pocket that is tailored in size, shape and hydrogen-bonding capacity for a uridylylated base (Fig. 2a,b). This induced fit of the binding pocket around U₂ accounts well for the cooperative nature of RNA binding by both domains of LaNTD [71, 73, 76, 77].

Either U₃ or U₄ appears able to stack on U₁ to stabilise the tight turn at the end of the bound RNA (see above). This stacked base is also held in place by base-specific hydrogen bonds involving the O2 and N3 atoms to the backbone of the La motif [74] (Fig. 2b).

The crystal structures of LaNTD complexes with RNA ligands having 3-8 uridylylates at the 3'-end are consistent with binding measurements from several laboratories. The particular contacts observed between the proteins and U₁-U₄ (depending on the structure) account readily for the observation that 3-4 uridylylates are needed for optimal binding affinity [75, 78, 79]. However, it is clear that these do not contribute equally and in fact, although modest reductions in binding affinity (1.5 to 3.3-fold) are observed following substitutions of U₁, U₃ or U₄, mutation of U₂ is much more deleterious, reducing the affinity by up to 14-fold [74, 75]. It is clear that U₂ is the primary source of binding energy; the structures reveal that this arises because it occupies the most enclosed portion of the binding cleft making numerous polar and non-polar contacts with the La motif and RRM1. Conceivably, the requirement for uridylylated as the penultimate base at the 3'-end may be sufficient to select from the pool of nuclear RNAs only that subfraction with oligo(U) sequences at the 3' end.

A particularly surprising aspect of the mode of binding of RNA 3' ends to LaNTD is that neither of the canonical nucleic acid binding surfaces of either the La motif or RRM1 (Fig. 2a) is involved in RNA contacts [74, 75]. This leaves these surfaces free to interact with other portions of the RNA ligand. Intriguingly, recent evidence suggests that the β -sheet of RRM1 may have an important role in the chaperone activity of La by stabilising the conformation of pre-tRNA molecules that bind to the protein via their 3' termini [80, 81].

3.2. Recognition of RNA 3' ends by rotavirus NSP3.

Rotavirus NSP3 (non-structural protein 3) also specifically recognises a short sequence at the 3' end of RNA. Although it operates in a different biological context to La and has a different structure, there are some interesting functional parallels between the two proteins. Rotaviruses, the major cause of gastroenteritis in children, have an 11-segment dsRNA genome contained within a layered protein capsid. The mRNA transcripts of the genomic segments have a 5' cap structure but otherwise differ significantly from normal eukaryotic mRNAs in that they lack a 3' poly(A) tail. Instead the mRNAs have a conserved tetranucleotide sequence

that is specifically bound by NSP3, a functional homolog of poly(A) binding protein (PABP). Like PABP, rotavirus NSP3 may enable tethering of the 3' and 5' ends of the mRNAs by binding to the eIF4G component of the cap-binding complex and thereby facilitating translation [82], although it is not yet clear whether the protein actually plays this role in infected cells [83]. Other functions may be more important: for example, by competing with PABP for binding to eIF4G, NSP3 may also help to down-regulate host cell protein synthesis [83]. It has also been suggested that, like La, NSP3 performs a protective role, shielding 3'-termini from exonucleases [82].

NSP3 from group A rotavirus interacts specifically with the GACC sequence found at the 3' end of the viral mRNAs. The N-terminal domain (residues 1-170) is responsible for RNA binding while the C-terminal domain binds to eIF4G [82]. Crystallographic analysis of the N-terminal fragment of NSP3 complexed with a GU-GACC hexanucleotide revealed that the protein forms a highly asymmetric homodimer containing a deep central binding cleft that accommodates the last four nucleotides at the 3' end of the RNA [82] (Fig. 2c). Differential scanning calorimetry (DSC) showed that RNA binding is associated with significant thermal stabilisation of the protein — the melting temperature is raised by 20°C. Since the structure shows that the RNA ligand could not dissociate without distortion of the binding pocket, the DSC result implies that conformational changes in the protein are involved in securing a tight interaction with the RNA ligand. This echoes the mode of binding of La to the 3' ends of its RNA ligands, though in that case rigid body motions of the two component domains are primarily responsible for locking onto the RNA.

In contrast to La the RNA-binding cleft in NSP3 is enclosed so that all four nucleotides of the binding sequence are largely shielded from solvent and the bases all make sequence-specific interactions with the protein (Fig. 2c,d). The 3' terminal nucleotide is accommodated in the deepest part of the pocket. As in La, both the 2' and 3'-OH groups are close-packed to the protein and involved in specific hydrogen bonds (in this case to the side chains of Asp 100 and Arg 105 – Fig. 2d) — explaining the specificity for ligands with a free 3' end. In addition to base-specific contacts, there are hydrogen bonds to all five of the phosphate groups present in the ligand and to each of the four 2'-OH groups of the tetranucleotide sequence. Moreover, the bases of the tetranucleotide fit snugly into the binding pocket, making many van der Waals and stacking interactions. The RNA backbone is relatively extended compared to the RNA ligands of La, though there is one kink due to base-stacking of the nucleotides in the -3 and -4 positions.

It is curious that despite binding a similar number of nucleotides in a much more enclosed pocket than found in LaNTD, the affinity of NSP3 for its cognate RNA is, if anything, slightly weaker ($K_D \sim 70$ nM) [82]. This may reflect the energetic cost of the conformational rearrangements of the protein that are involved in stabilising the bound conformation.

4. RNAi – Argonaute proteins.

Novel modes of recognition of RNA termini have been revealed in structural studies of Argonaute (Ago) proteins which bind both

the 5' and 3' ends of single-stranded RNA molecules to form the functional core of macromolecular complexes that are ultimately responsible for RNA interference (RNAi), a multi-faceted mode of gene silencing found in all eukaryotes. RNAi can suppress gene expression directly, either by cleaving or 'slicing' mRNA molecules or by repressing their translation into protein. Other modes of interference can induce transcriptional silencing or repress transposons. This topic has been covered in great detail in a number of recent reviews [3, 7, 84–87] and the reader is referred to these for a more detailed consideration of the functional aspects of RNAi; we will focus here on the structural details and mechanistic implications of the tethering of 5' and 3' ends of RNA (and DNA) that have emerged from structural studies of Argonaute-nucleic acid complexes.

Despite the diversity of interference mechanisms, silencing is mediated by a common structural unit formed by loading an Argonaute protein with a short RNA co-factor of ~20–30 nucleotides in length to form the functional core of a protein-RNA complex known as an RNA-induced silencing complex (RISC). The inserted RNA guide strand is tethered to Ago at its 5' and 3' extremities and determines, via base-pairing interactions, the specificity of RISC for RNA targets in the cell.

The slicing and translation repression modes of RNAi, which have received the most attention to date, have similar pathways of RISC formation [3, 85]. This involves the initial loading of Ago with a small RNA duplex that is pre-cut by Dicer, an RNase III-type endonuclease that is specific for dsRNA and uses specific 3' end recognition to generate duplexes of the appropriate length. The small RNA duplexes produced by Dicer have a characteristic monophosphate at the 5' end and a dinucleotide overhang at the 3' end, features that are specifically recognised by Ago. After loading onto Ago one strand of the duplex is retained as the guide while the complementary passenger strand is cleaved and/or ejected.

Although there are notable exceptions to the general rule, the particular mode of silencing enforced by RISC depends on the extent of base-pairing with the RNA target. Incorporation of short interfering RNAs (siRNA) normally leads to cleavage or slicing of the target RNA because extensive complementary base-pairing activates the RNase activity of the Ago subunit. In contrast, when a microRNA (miRNA) guide strand within Ago recognises target sequences in mRNA 3'-untranslated regions, the interaction usually involves mis-matches and results in imperfect base-pairing. This prevents cleavage of the mRNA, even if the Ago protein has a functional active site, but leads to translational repression, by mechanisms that are still unclear [3, 7].

Argonaute proteins are also found in archaea and bacteria, although their roles in these domains of life have not yet been clarified: small RNAs have so far only been found in eukaryotes [84]. Nevertheless, archaeal and bacterial Argonautes have significant sequence homology with their eukaryotic counterparts and have been very useful for investigating the structures of Ago proteins [88–93], particularly since it has proved difficult to produce samples of eukaryotic Agos that are suitable for structural analysis. The use of archaeal and bacterial Ago proteins in structural studies adds a potentially complicating factor in terms of

understanding eukaryotic RNAi, since these proteins appear to use DNA guide strands to target RNA for cleavage.

Argonaute proteins (~100 kDa) have four domains that fold into a flexible bi-lobed structure that forms an extended binding platform for the guide strand RNA [89, 92, 93] (Fig. 3a). The N-terminal and PAZ domains form one lobe, while the MID and PIWI form the other. Initial structural work on domain fragments revealed that the 3' end of the guide strand is bound by the PAZ domain [94, 95]; the 5' end is anchored by the MID domain within the MID/PIWI lobe at the other end of the binding platform [88, 91]. More recent work has begun to elucidate how an intact Ago protein fixes the guide strand in place so that it is poised to recognise its RNA target [89, 90] (Fig. 3a).

4.1. 3' end recognition by PAZ domains.

The structures of Ago PAZ domains have been solved for proteins taken from all 3 domains of life in the free state and complexed with a variety of nucleic acid ligands: ssRNA [94], a duplex siRNA mimic [95] and—most recently—a DNA guide strand (with and without an RNA target present) [89, 90]. These structures reveal the general features of a conserved mode of 3' end binding despite small structural variations between PAZ domains.

The PAZ domain is composed of a compact oligonucleotide-binding (OB) fold elaborated by an inserted mini-domain that adopts a $\beta\beta\alpha$ [96, 97] or $\alpha\alpha$ topology [92, 93]. Initial solution and crystal structures of RNA complexes with PAZ domain fragments from two metazoan species revealed that the two nucleotides at the 3' end insert into a pre-formed cleft between the two domains, creating a distinctive mode of oligonucleotide binding that is not observed in other OB proteins [94, 95] (see Fig 2 in [5]).

The solution structure of the PAZ domain from *Drosophila melanogaster* Ago2 (*DmAgo2*) complexed with a short single stranded RNA oligomer (5'-CUCAC-3') resolved only the two terminal bases; the rest of the RNA was disordered, apparently not contacting the protein [94]. Intriguingly the co-crystallisation of the PAZ domain from human Ago1 (*HsAgo1*) complexed with a 9 nt ssRNA inadvertently generated a short siRNA-like duplex RNA with the characteristic two-base 3'-overhang [95]. This interesting structure may represent a loading stage of RISC formation, prior to removal of the passenger strand.

Comparison of the PAZ-RNA complexes from *HsAgo1* and *DmAgo2* shows that the ribose phosphate backbone of the terminal pair of nucleotides is inserted into the inter-domain cleft and makes extensive interactions with the protein that are focused particularly on the terminal nucleotide and the backbone connecting it to the penultimate base. The ribose of the terminal nucleotide occupies the deepest part of the pocket. In the PAZ domain from *HsAgo1* the 2' and 3'-OH groups make hydrogen bonds with the amide and carbonyl groups of Tyr 336 respectively [95] (Fig. 3b). The solution structure of the PAZ-RNA complex from *DmAgo2* suggests that a similar pair of H-bonds is made with the amide and carbonyl groups of the spatially equivalent residue, L110, although the orientation of the ribose moiety is slightly variable among members of the ensemble of structures [94]. This interaction contrasts with 3' recognition by La and NSP3, where charged side chains make

Fig. 3: Structures of RNAi effectors and repressors. (a) Structure of TtAgo complexed with a DNA guide strand (dark-yellow) and RNA target strand (dark-grey) [89]. The domains of the Ago protein are colour differently and labelled. The disordered segment of the guide strand is indicated by a yellow dotted line. (b) Close-up view of the binding of the 3' end of an RNA guide strand to HsAgo1 [95]. Selected side chains are shown as sticks and waters molecules as small spheres. (c) Close-up view of the binding of the 5' end of an RNA guide strand to AfPiwi [91]. The divalent metal cation that is coordinated by the carboxy-terminus of the protein (Leu 427) and the phosphate groups of the first and third nucleotides of the guide strand is shown as a purple sphere. Structures of (d) the CIRV p19 homodimer [109] and (e) the TAV2b homodimer [108] in each case complexed with an siRNA duplex. The proteins are shown in a space-filling representation with the monomers coloured light-blue and pink. Close-up view of the recognition of the 5' and 3' ends of the siRNA duplex by (f) CIRV p19 and (g) TAV2b. For the TAV2b structure the disordered 3' overhang is depicted as an orange dotted line in panels e and g.

hydrogen bonds to the ribose hydroxyls [74, 82] (Fig. 2b,d). The non-bridging oxygens of the phosphodiester backbone between the terminal pair of nucleotides make a further four hydrogen bonds with adjacent side chains in *HsAgo1* (Fig. 3b). A similar array of side chains appears able to coordinate the bridging phosphate in *DmAgo* [94]. The terminal ribose and base are packed snugly into one end of the pocket making van der Waals and hydrophobic contacts with the protein (a close-packed mode of binding similar to that observed in rotavirus NSP3 [82] – see above). The terminal base packs against the aromatic side chain of a conserved Phe in both structures (Phe 292 in Fig. 3b), although this is not a normal stacking interaction since the benzyl ring is oblique to the plane of the base. The two terminal nucleotides of the guide strand are, however, stacked plane parallel on one another in such a way that their base-pairing edges face outward, thus avoiding any base-specific contacts with the protein, in keeping with the non-sequence-specific nature of the PAZ-RNA interaction (Fig. 3b).

More recent work has revealed that the 3' end of the DNA guide strand in *Thermus thermophilus* Ago (*TtAgo*) binds to the PAZ domain in a manner that closely mimics RNA binding to eukaryotic PAZ domains [89, 90]. Strikingly, although the secondary structure of the binding pocket is preserved, the detailed architecture is quite distinct and appears to reflect the preference of the bacterial protein for DNA. In the PAZ domain from *TtAgo* the peptide backbone that runs underneath the ribose of the 3'-terminal nucleotide of the guide strand incorporates a proline residue (Pro 225) that distorts the main chain from the conformation needed to make a pair of hydrogen bonds to the 2' and 3'-OH groups of the terminal RNA base in eukaryotic PAZ domains [94, 95]. Instead recognition of the 3'-OH of the terminal DNA base is via hydrogen bonds to the side chain of His 227 and the carbonyl oxygen of Pro 225 [89, 90]. The dimensions of the pocket appear to leave no room for the 2'-OH that would be present in an RNA ligand.

There are other interesting structural differences between the prokaryotic and eukaryotic PAZ domains that underscore how similar modes of binding can be achieved in different ways. Although the non-bridging oxygens of the terminal phosphate are also bound by hydrogen bonds to four side chains from *TtAgo* PAZ, only one of these (Tyr 226) aligns with an equivalent residue in *HsAgo1* (Tyr 309 - Fig. 3b) [89, 90]. Moreover, there is no aromatic side chain to pack against the terminal base, closing off one end of the binding pocket; instead, this function is fulfilled by the hydrophobic side chains of Leu 217 and Pro 218 in the prokaryotic PAZ.

Although there is clearly a common mode of binding of the terminal dinucleotides of RNA and DNA guide strands to different PAZ domains, the path of the oligonucleotide upstream of the 3'-dinucleotide exhibits considerable variation in the known structures. In part this may arise from species differences or variations in the particular RNA and DNA ligands used in the structural studies reported to date, but it nevertheless seems likely that this is mostly due to the mobility of this portion of the tethered guide strand. Strong support for this notion comes from structural analysis of *TtAgo* bound to DNA in three different liganded states. In the binary *TtAgo*-DNA complexes using truncated or full-length

guide strands [90] and a ternary *TtAgo* -guide DNA-target RNA complex [89], the terminal two nucleotides are held in the same stacked conformation but the structures show clear divergence in the paths of the upstream sequences (see Fig. 3b in [90]). Although the phosphate backbone of the guide strand can interact with residues lining the basic channel proceeding from the PAZ domain through to the MID/PIWI lobe that binds the 5' end, there is evidently a significant degree of residual mobility, which may be important for target recognition (see below).

The mode of recognition of a dinucleotide 3' end overhang by PAZ domains that has emerged from these structural studies is consistent with binding studies. The snugness of fit of the 3'-terminus in the binding pocket is confirmed by the deleterious effect of covalent modifications of the 2' and 3' hydroxyls of RNA ligands [95]. Moreover, it is clear that for double-stranded RNA ligands a 2-nucleotide 3'-overhang provides for optimum binding affinity [95-97], a finding that is consistent with a possible role for PAZ in aiding the initial loading of dsRNA onto Ago during RISC assembly.

Interestingly, Patel and colleagues, using surface plasmon resonance to measure PAZ-RNA interactions, found that extension of the 3'-overhang greatly reduced binding: a 10 nt overhang bound 50-fold less well than the optimal 2 nt overhang [95]. Although this might be taken to suggest that RNA ligands with a more single-stranded character may bind less well to the protein, the finding is not consistent with earlier measurements, made using competition cross-linking assays, which showed that PAZ bound single-stranded RNA and siRNA-like duplexes with equal affinity [96, 97]. The reason for this apparent discrepancy is not yet clear but is worthy of further investigation since differentiation between double and single-stranded ligands may play a role in the formation of active RISC. However, it may be more rewarding — though clearly more difficult — to probe the contribution of PAZ to RNA binding within the context of intact Ago proteins.

4.2. 5' end recognition by MID/PIWI.

Anchoring and presentation of the 5' end of guide strand by Ago proteins are critical for two reasons. Firstly, cleavage of target RNAs during slicing is measured specifically from the 5' end of the guide strand and occurs between the nucleotides that are base-paired to nucleotides 10 and 11 of the guide [3]. Secondly, nucleotides 2-7 of the guide strand (the seed sequence) have been found to be the most important determinants of specific target recognition, for both slicing and translation repression modes of RNAi [98, 99].

As with 3' end recognition, structural analyses have detected common modes of binding of the 5' end of both RNA and DNA guide strands by the MID/PIWI lobe of Ago and homologous Piwi proteins. The structural work has so far been confined to archaeal and bacterial proteins but sequence homology and structure-based mutagenesis strongly suggest that many of the key features of 5' end recognition will also be found in eukaryotic orthologues.

The crystal structures of archaeal Ago proteins from *Pyrococcus furiosus* (*PfAgo*) and *Archaeoglobus fulgidus* (*AfPiwi*) established that the MID domain resembles the sugar-binding domain of the

lac repressor, whereas the PIWI domain has a similar fold to RNase H and contains the catalytic centre of the protein [92, 100]. Shortly thereafter, co-crystal structures of *A*Piwi with two different siRNA-like duplexes [88, 91] revealed the key features of how the 5′ end of the guide strand is anchored to the protein and may interact with passenger or guide strands. Although *A*Piwi is a small Ago protein lacking the N-terminal and PAZ domains (which probably made it more amenable to co-crystallisation studies), its observed mode of RNA binding has been generalised to DNA ligands and to full-sized Ago proteins [89, 90, 101].

At the 5′ end of the guide strand, interaction with the *A*Piwi Ago protein forces the first nucleotide to flip out of its base-pair interaction with the complementary strand into a highly-conserved pocket that is formed almost exclusively from MID domain residues but is positioned close to the interface of the MID and PIWI domains [88, 91] (Fig. 3c). This allows the terminal monophosphate to bind at the deepest part of the cleft where it makes numerous hydrogen bond or salt-bridge interactions with four amino-acid side chains, the amide nitrogen of Phe 138 and a divalent metal cation that is co-ordinated by Gln 159 and — unusually — the C-terminal carboxylate moiety of the protein (Leu 427). The bound cation also interacts with the phosphate of the third nucleotide of the guide strand so that the backbone of the first three residues is pinched into a tight U-bend (Fig. 3c). This mode of binding is likely to be replicated in eukaryotic Ago proteins since the side-chain ligands for the monophosphate are all strongly conserved, a contention supported by the finding that mutation of the equivalent residues in *HsAgo2* impaired its slicing activity, presumably due to mis-placement of the 5′ end of the guide strand [88].

The pocket that accommodates the 5′-monophosphate is less snug than the cleft in the PAZ domain that binds the 3′ end of the guide strand and there is room for several bound water molecules in the vicinity of the 5′-terminus (Fig. 3c). This suggests that the pocket may be able to accommodate secondary siRNA guide strands that are generated by RNA-dependent polymerases to amplify the silencing signal in some siRNA pathways and therefore have di- and triphosphate groups at their 5′ ends [84, 86].

The base of the first nucleotide is accommodated within a binding slot adjacent to the phosphate pocket where it is stabilised by stacking underneath Tyr 123 (Fig. 3c). A second tyrosine (Tyr 118) sits below the base, though its position is somewhat variable between the three co-crystal structures that have been determined. Only in the structure with a DNA guide strand does Tyr 118 form the bottom layer of a sandwich that resembles the paired arrangement of aromatic side chains that is common in cap-binding proteins [101]. The edge of the base (U or A in the two structures with RNA [88, 91] and T in a structure with a DNA duplex [101]) makes one or two hydrogen bonds on one side of the binding cleft but there appears to be no selectivity for the identity of the 5′-terminal base in *A*Piwi. In contrast, some eukaryotic Ago proteins have a preference for particular 5′-nucleotides, presumably due to modifications of the binding pocket that confer base-specificity [3].

The interactions of the 5′ terminal base with the MID domain inhibit it from base-pairing with target RNA and provide an ex-

planation for the observation that a mis-match at this position is not important for target recognition [3, 84]. Intriguingly, it has also been found that this mismatch can, in some circumstances, enhance slicing activity [99]. It may be that a target capable of base-pairing with the guide strand at this first position destabilises its interaction with the protein, weakening the 5′-anchoring by Ago and perhaps leading to mis-positioning of the scissile bond of the target with respect to the active site. A small dislocation of the substrate would be sufficient to significantly impair slicing activity.

Immediately downstream of the 5′-terminal nucleotide, the backbone makes a number of direct and water-mediated hydrogen-bonds with the protein, so that the bases in the 5′ section of the guide are exposed and available for base-pairing with the target. However beyond the third nucleotide the paths of the different RNA and DNA guide strands in the reported structures begin to diverge [88, 89, 101]. Moreover, the duplexes are disordered at their 3′ ends such that only four or five base-pairs are observable, fewer than half the number expected from the dsRNA and dsDNA ligands used in co-crystallisation [88, 89, 101]. These observations are rather surprising given that fully complementary oligonucleotides were used to prepare the duplexes in each case. It might have been thought that the disorder in the duplexes was due to the absence of a PAZ domain in *A*Piwi to tether the 3′ end of the guide. However, the structure of *TtAgo* in a complex with a DNA guide and an RNA target strand shows that, despite anchoring of the 3′ end of the guide in its PAZ domain, the 3′ segment of the duplex is also disordered [89] (Fig. 3a). The evident flexibility of this portion of the guide-target duplex, in contrast to the clear stabilisation of the 5′ end of the guide strand may therefore be a general feature and is consistent with some of the known features of target selection.

It is well established that base-pairing between the target RNA and nucleotides 2-7 at the 5′ end of the guide strand (the so-called seed region) are major determinants of RISC specificity [98, 99]. Although base-pairing of the target to the 3′ end of the guide is less important for specificity in both slicing and translation repression modes of silencing [98, 99], it can still contribute to efficacy. A case in point is the observation that base-mismatches in the 3′ end of the guide reduce the catalytic efficiency of slicing by *DmAgo* [99]. The current hypothesis of target recognition is that initial contact may be established by base-pairing of the target to the seed region, followed by propagation of base-pairing towards the 3′ end of the guide strand [98, 99]. This mode of action is supported by recent binding analyses of target recognition by *A*Piwi-guide complexes which showed that stabilisation of the 5′ end of the guide strand by interaction with the protein enhances the binding energy derived from base-pairing of the guide with the target by paying for the entropic cost of the interaction [101]. It is still unclear why stable base-pairing of the target to the 3′ segment of the guide strand downstream of the scissile bond was not observed for *TtAgo* [89]. Conceivably the duplex is formed but, because there are no stabilising interactions with the protein, it remains flexible and therefore invisible to crystallographic analysis. This order-disorder transition along the length of the guide strand may be necessary for key functions of Ago: stabilisation of the 5′ end of the guide to allow accurate measurement of the position of cleavage of the target but loose attachment of the 3′ segment of the guide to permit

accommodation of guide strands of variable lengths. The evident flexibility of Ago may also contribute to the ability of the protein to use guide strands of different sizes [89, 90].

Although the extreme 3' end of the guide is observed to be bound to the PAZ domain in *TtAgo* in the absence and presence of a target RNA, it is still not clear if this end remains attached to the protein during 'complete' target recognition. The formation of base-pairs along the whole length of the target strand would generate two turns of A-form double helix. Topologically, it is difficult to see how this can be achieved if the 3' end of the guide remains tethered to the protein [3, 85].

A new and interesting dimension to the ability of Ago proteins to recognise the ends of RNA molecules has emerged recently with reports that they may also be able to bind the 5' m⁷G cap of mRNA targets that are the subject of RISC-mediated translational repression [102, 103]. The working hypothesis is that miRNA-loaded RISCs tethered to near-complementary sequences in the 3'-UTRs of target mRNA may inhibit translation initiation by sequestering the 5' cap of the message from eIF4E. However, this model has been challenged in later work [104, 105]. In particular, an initial report [103] that the MID domain contains an eIF4E-like motif complete with a conserved pair of aromatic residues to sandwich the m⁷G cap has proved — on the basis of a detailed sequence and structural comparison — to be unfounded [106]. Further binding and structural analyses are needed to put the putative cap-binding properties of Ago on a secure footing.

5. Binding of siRNAs by viral suppressor proteins.

Plants and invertebrates use RNAi as a defence mechanism against viral infection and, as a counter-strategy, viruses express suppressor proteins to inhibit the RNAi response of the host cells [107]. Although several different viral anti-silencing strategies have been identified, the best characterised suppressors are those that function by binding to siRNA duplexes, thus inhibiting RISC assembly. Structures have been reported for the *Tomato aspermy virus* protein 2b (TAV2b) [108] and two related p19 proteins from the *tombusvirus* genus [109, 110]. These suppressors bind specifically to siRNA duplexes of defined length using a "molecular calliper" arrangement.

The crystal structure of the p19 suppressor bound to siRNA has been determined both for *Tomato bushy stunt virus* (TBSV) [110] and the closely related *Carnation Italian ringspot virus* (CIRV) [109] (Fig. 3d). As expected, the two proteins, which are both homodimeric, have nearly identical structures and bind siRNA in the same manner. The dimer has a central 8-stranded β -sheet that is formed by the association of the two C-terminal domains and capped at each end by the N-terminal domains. P19 binds the siRNA duplex symmetrically; the ends of the siRNA are bound to the N-terminal domains while the body of the duplex lies across the β -sheet surface, interacting directly or via water molecules with the phosphates of the backbone and the 2' OH groups of the sugars.

The 19bp RNA in the complex adopts the standard A form conformation but the two unpaired bases at each of the 3' ends are unstacked and twist away from the helix (Fig. 3d). Two tryptophan

residues in each monomer of the suppressor stack onto the exposed bases at the 5' and 3' ends of the double stranded RNA (Fig. 3f). The highly conserved Trp 42 stacks against the 5' base while Trp 39 stacks against the last paired 3' base. The imino nitrogen of Trp 42 also makes a hydrogen bond with the 5' phosphate in the CIRV p19 structure (the RNA used to co-crystallised TBSV p19 was unphosphorylated [110]). The unpaired 3' bases are disordered in the TBSV p19 crystal structure and make only weak interactions with the protein in the CIRV structure. As expected from the sequence-independent mode of binding, there are no direct interactions of the protein with the RNA bases.

Electrophoretic gel mobility shift assays showed that removal of the 5' phosphate significantly reduced the affinity for CIRV p19 for a 21 nt siRNA, consistent with its close interaction with the protein. Elimination of the unpaired 3' nucleotides actually marginally increased the affinity [109], a result that is in marked contrast to the effect that this change has on the binding of siRNA to PAZ [95, 96]. The difference is evidently due to the fact, unlike p19, PAZ accommodates the overhang in a specific binding pocket.

Although the p19 suppressor binds most tightly to 21-nt siRNA as used in the crystal structures, it can also accommodate slightly longer siRNA duplexes albeit with reduced affinity [109]. This may be a necessary strategy for the virus as not all siRNAs are the same size. For example the different Dicer-like proteins in *Arabidopsis thaliana* produce siRNA duplexes with lengths ranging from 21 nt to 24 nt [111].

The RNA interference suppressor *Tomato aspermy virus* protein 2b is also dimeric and binds siRNA with a symmetric "molecular calliper" arrangement using tryptophan residues, but has a strikingly different architecture from the p19 suppressor [112] (Fig. 3e). Each monomer is 95 amino acids long but so far it has only been possible to obtain crystals of the fragment 1-69 bound to siRNA. These fold into a pair of hinged helices that insert into the major groove of the A-form dsRNA. The dimer interface is surprisingly small, comprising only a handful of residues at the beginning of the N-terminal helix.

Trp 50 in the C-terminal helix of each monomer stacks onto the exposed 5' base of the dsRNA (Fig. 3g). This tryptophan is not conserved in the homologous 2b protein from *Cucumber mosaic virus*, which has an arginine residue at this position, a substitution that is also seen in the binding pocket for the 5' base of the guide strand between different Ago proteins (see above). Unsurprisingly therefore, mutation of Trp 50 to Arg reduces the RNA binding affinity of TAV2b only by a factor of two, though mutation to Ala knocks binding down five-fold. The 5' phosphate forms a salt bridge with Lys 39 and also interacts with His 51. The unpaired 3' bases are disordered in the crystal structure suggesting that — as in p19 — they do not form strong interactions with the suppressor molecule, but this observation awaits confirmation since the last 21 amino acids of the monomer were not present in the protein used for crystallisation. It is perhaps notable that RNA binding measurements made with the same protein fragment that was used to grow crystals yielded a K_D of 75 nM for a 21-nt siRNA duplex [108], which indicates an affinity that is about 400 times lower than p19 [109]. Given that TAV2b and p19 have similar functions and binding targets within plant cells, one might have expected them to

Fig. 4: Structures of RNaseE complexed with mRNA substrates containing a 5'-monophosphate group. (a) Overview of a catalytic dimer, formed from two identical subunits. The dimeric two-fold symmetry axis is perpendicular to the plane of the page. Subunit 1 is depicted in a yellow surface representation with the bound RNA shown as an orange ribbon. The left-hand subunit (green) is drawn schematically with bound RNA represented by sticks coloured by atom type. The Mg^{2+} ions associated with the two active sites of the dimer are shown as blue-green spheres. (b) Close-up of subunit 1 with the structures of two different bound RNAs superposed – 5'-ACAGUAUUUG (orange) and 5'-UUUACAGUAUUUG (blue). With reference to panel a, the view is tilted around the horizon axis by about 20°. (c) Close-up of the binding pocket for the 5'-monophosphate of 5'-UUUACAGUAUUUG shown in sticks coloured by atom-type. Selected protein side chains are also shown as sticks. The orientation is similar to panel b.

bind with similar affinities. It is possible that the missing C-terminal region of TAV2b accounts for the lower affinity of the protein observed in these experiments.

Surprisingly, the TAV2b fragment binds to a 32 nt siRNA duplex with a dissociation constant only a factor of two greater than that for 21 nt siRNA. This would suggest that most of the binding energy derives from the interaction of the protein α -helices with the RNA major groove. It is not clear from the structure, particularly since it is incomplete, whether TAV2b could bind both ends of such a long dsRNA target but perhaps binding of the dimer to one 5' end and most of the body of the RNA duplex is sufficient to confer specificity.

6. 5'-monophosphate recognition by RNaseE

While the m^7G cap added to the 5'-end of eukaryotic mRNAs serves to protect them from degradation pathways, prokaryotic mRNAs are uncapped [113]. Nevertheless, in bacteria retention of the 5'-triphosphate moiety is sufficient to inhibit mRNA turnover and diversion to degradation pathways is only initiated once the RNA pyrophosphohydrolase RppH converts the 5' end to a mono-

phosphate [114, 115]. Following this modification the mRNA becomes a specific substrate for the RNaseE endonuclease.

Crystallographic analysis of RNaseE revealed that the enzyme is a homotetramer composed of multi-domain subunits. The core catalytic units of the enzyme are composed a pair of dimers; within each dimer a sensor domain from one monomer binds the 5'-monophosphate of the RNA substrate, but cleavage is performed several nucleotides downstream by the active site of the other monomer (Fig. 4a).

In broad terms there are similarities with Ago in the way that binding of the 5'-end of the substrate is coordinated with endonucleolytic cleavage downstream. However, there are two key differences: firstly, RNaseE is a non-specific endonuclease — there is no guide strand — and, secondly, it does not measure the distance between the 5'-end of the substrate and the cut site. Crystallographic analysis of RNaseE in complex with three RNA substrates of different sequences and lengths (10, 13 and 15 nucleotides) readily explains the structural basis of these distinguishing features [6, 116].

The RNA substrate is bound with its 5'-monophosphate tucked into a pocket formed by the sensor domain (Fig. 4b). This pocket is connected to the active site of a neighbouring monomer by a sur-

face channel that binds the downstream portion of the RNA; the mode of binding allows some of the intervening sequence to loop out, away from the enzyme surface.

Within the pocket that binds the 5' end of the RNA there are no hydrogen bonds to the terminal base, although the side chain of Val 128 provide important non-specific contacts to the substrate [116]. Specific recognition of the 5'-monophosphate is achieved through hydrogen bonds from Arg 169 and Thr 170 deep within the sensor pocket (Fig. 4c). The side chain of Arg 371 may also contribute to binding though its position varies with different RNA substrates. Arg 373 appears to contribute to stabilisation of the 5'-end since it can form salt bridges with the backbone phosphates with the nucleotides at the beginning and end of the portion of the RNA that is looped out (positions 2 and 5 for the RNA in Fig. 4c). However, again, this interaction is not consistently observed since in one complex (PDB 2c4r) the side chain is disordered. Therefore the pocket that binds the 5' end of the RNA appears to exhibit a certain degree of flexibility. In RNaseE this flexibility, along with the presence of a cavity beneath the position of the bound 5'-terminus may account for the observation that RNaseE binds RNAs with 5'-monophosphate or 5'-triphosphate with similar affinities [117]. The enhanced rates of cleavage observed for 5'-monophosphate RNA substrates [117] have been attributed to the conformational changes that coordinate 5' end recognition with the accurate positioning of the scissile bond in the active site [116], but the structural details of this change have yet to be elucidated.

The protrusion of a looped region of the RNA substrate from the surface of the enzyme makes it possible for RNaseE to cleave at different distances from the 5' end. In the two structures that exhibit looping, the RNA is positioned so that cleavage would occur after the eighth nucleotide (Fig. 4b,c). Curiously, RNaseE is capable of cleaving after the fifth nucleotide from the 5' end [118], suggesting that a more linear conformation of the substrate can be adopted. Examination of the structure suggests that this may be the minimum distance from the 5' end at which cleavage can occur. Conversely, the looping facility also allows cleavage much further downstream in the RNA sequence [116]; in mRNA targets with a stem-loop inserted just downstream of the terminal nucleotide, the position of cleavage would depend on the length of the insert.

7. Concluding remarks

The enormous amount of structural information accumulated over the past ten years or so has transformed our understanding of the protein recognition of RNA termini. One of the most striking aspects of the results that have arisen from this work is the tremendous diversity of structural solutions to the problem of specific binding of the 5' and 3' ends of RNA molecules.

The structural studies have already started to pave the way for more incisive functional investigations but are still some way from completion. In part this is due to the flexible nature of the binding proteins involved, an important facet of RNA terminus recognition that has emerged from the structural work, but one that presents technical difficulties, particularly for crystallo-

graphic approaches to structure determination. As we have seen, the best strategy for overcoming this difficulty is to compare multiple structures of orthologues of the same protein with different RNA ligands. This approach, since it invariably involves production of different crystal forms, also helps in the detection of artefacts that may arise from the impact of crystal contacts on portions of the ligand or binding pocket. A second problem is that many proteins that recognise RNA termini frequently do so as part of larger multimeric complexes. These complexes, which may exhibit additional flexibility, are also more difficult to prepare in sufficient quantities and at the level of purity needed for structural analysis. But the field is already moving to overcome such problems, on NMR and crystallographic fronts [57, 89], and the next decade surely promises further fascinating revelations.

Acknowledgements

SC and MRC are grateful to the Wellcome Trust and the BBSRC for funding of work on protein-RNA interactions.

References

- [1] P. Fechter, G.G. Brownlee, Recognition of mRNA cap structures by viral and cellular proteins, *J. Gen. Virol.* 86 (2005) 1239-1249.
- [2] F.A. Quijcho, G. Hu, P.D. Gershon, Structural basis of mRNA cap recognition by proteins, *Curr. Opin. Struct. Biol.* 10 (2000) 78-86.
- [3] M. Jinek, J.A. Doudna, A three-dimensional view of the molecular machinery of RNA interference, *Nature* 457 (2009) 405-412.
- [4] T.M. Hall, Structure and function of argonaute proteins, *Structure* 13 (2005) 1403-1408.
- [5] A. Lingel, M. Sattler, Novel modes of protein-RNA recognition in the RNAi pathway, *Curr. Opin. Struct. Biol.* 15 (2005) 107-115.
- [6] J.A. Worrall, B. Luisi, Information available at cut rates: structure and mechanism of ribonucleases, *Curr. Opin. Struct. Biol.* 17 (2007) 128-137.
- [7] N. Sonenberg, A. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets, *Cell* 136 (2009) 731-745.
- [8] T. Von Der Haar, J. Gross, G. Wagner, J. McCarthy, The mRNA cap-binding protein eIF4E in post-transcriptional gene expression, *Nat. Struct. Mol. Biol.* 11 (2004) 503-511.
- [9] B. Joshi, A. Cameron, R. Jagus, Characterization of mammalian eIF4E-family members, *Eur. J. Biochem.* 271 (2004) 2189-2203.
- [10] E. Rom, H.C. Kim, A.C. Gingras, J. Marcotrigiano, D. Favre, H. Olsen, S.K. Burley, N. Sonenberg, Cloning and characterization of 4EHP, a novel mammalian eIF4E-related cap-binding protein, *J. Biol. Chem.* 273 (1998) 13104-13109.
- [11] J. Zuberek, D. Kubacka, A. Jablonowska, J. Jemielity, J. Stepinski, N. Sonenberg, E. Darzynkiewicz, Weak binding affinity of human 4EHP for mRNA cap analogs, *RNA* 13 (2007) 691-697.
- [12] H. Matsuo, H. Li, A.M. McGuire, C.M. Fletcher, A.C. Gingras, N. Sonenberg, G. Wagner, Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein, *Nat. Struct. Biol.* 4 (1997) 717-724.
- [13] J. Marcotrigiano, A.C. Gingras, N. Sonenberg, S.K. Burley, Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP, *Cell* 89 (1997) 951-961.
- [14] P. Rosettani, S. Knapp, M.G. Vismara, L. Rusconi, A.D. Cameron, Structures of the human eIF4E homologous protein, h4EHP, in its m7GTP-bound and unliganded forms, *J. Mol. Biol.* 368 (2007) 691-705.
- [15] L. Volpon, M.J. Osborne, I. Topisirovic, N. Siddiqui, K.L. Borden, Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands, *EMBO J.* 25 (2006) 5138-5149.
- [16] C. Mazza, M. Ohno, A. Segref, I.W. Mattaj, S. Cusack, Crystal structure of the human nuclear cap binding complex, *Mol. Cell* 8 (2001) 383-396.
- [17] C. Mazza, A. Segref, I.W. Mattaj, S. Cusack, Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex, *EMBO J.* 21 (2002) 5548-5557.
- [18] G. Calero, K.F. Wilson, T. Ly, J.L. Rios-Steiner, J.C. Clardy, R.A. Cerione, Structural basis of m7GpppG binding to the nuclear cap-binding protein complex, *Nat. Struct. Biol.* 9 (2002) 912-917.
- [19] E. Izaurralde, J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, I.W. Mattaj, A nuclear cap binding protein complex involved in pre-mRNA splicing, *Cell* 78 (1994) 657-668.
- [20] C. Maris, C. Dominguez, F.H. Allain, The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, *FEBS J.* 272 (2005) 2118-2131.
- [21] N.A. Balatsos, P. Nilsson, C. Mazza, S. Cusack, A. Virtanen, Inhibition of mRNA deadenylation by the nuclear cap binding complex (CBC), *J. Biol. Chem.* 281 (2006) 4517-4522.
- [22] J. Huber, U. Cronshagen, M. Kadokura, C. Marshallsay, T. Wada, M. Sekine, R. Lührmann, Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure, *EMBO J.* 17 (1998) 4114-4126.
- [23] J. Martinez, Y.G. Ren, P. Nilsson, M. Ehrenberg, A. Virtanen, The mRNA cap structure stimulates rate of poly(A) removal and amplifies processivity of degradation, *J. Biol. Chem.* 276 (2001) 27923-27929.
- [24] E. Dehlin, M. Wormington, C.G. Korner, E. Wahle, Cap-dependent deadenylation of mRNA, *EMBO J.* 19 (2000) 1079-1086.
- [25] M. Gao, C.J. Wilusz, S.W. Peltz, J. Wilusz, A novel mRNA-decapping activity in HeLa cytoplasmic extracts is regulated by AU-rich elements, *EMBO J.* 20 (2001) 1134-1143.
- [26] T. Monecke, S. Schell, A. Dickmanns, R. Ficner, Crystal structure of the RRM domain of poly(A)-specific ribonuclease reveals a novel m(7)G-cap-binding mode, *J. Mol. Biol.* 382 (2008) 827-834.
- [27] T. Nagata, S. Suzuki, R. Endo, M. Shirouzu, T. Terada, M. Inoue, T. Kigawa, N. Kobayashi, P. Güntert, A. Tanaka, Y. Hayashizaki, Y. Muto, S. Yokoyama, The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition, *Nucleic Acids Res.* 36 (2008) 4754-4767.
- [28] M. Wu, P. Nilsson, N. Henriksson, A. Niedzwiecka, M. Lim, Z. Cheng, K. Kokkoris, A. Virtanen, H. Song, Structural basis of m(7)GpppG binding to poly(A)-specific ribonuclease, *Structure* 17 (2009) 276-286.
- [29] A. Strasser, A. Dickmanns, R. Lührmann, R. Ficner, Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1, *EMBO J.* 24 (2005) 2235-2243.
- [30] B. Moss, B.Y. Ahn, B. Amegadzie, P.D. Gershon, J.G. Keck, Cytoplasmic transcription system encoded by vaccinia virus, *J. Biol. Chem.* 266 (1991) 1355-1358.
- [31] F.A. Quijcho, G. Hu, P.D. Gershon, Structural basis of mRNA cap recognition by proteins, *Curr. Opin. Struct. Biol.* 10 (2000) 78-86.
- [32] A.E. Hodel, P.D. Gershon, X. Shi, S.M. Wang, F.A. Quijcho, Specific protein recognition of an mRNA cap through its alkylated base, *Nat. Struct. Biol.* 4 (1997) 350-354.

- [33] A.E. Hodel, P.D. Gershon, F.A. Quioco, Structural basis for sequence-nonspecific recognition of 5'-capped mRNA by a cap-modifying enzyme, *Mol. Cell* 1 (1998) 443-447.
- [34] D. Guilligay, F. Tarendeau, P. Resa-Infante, R. Coloma, T. Crepin, P. Sehr, J. Lewis, R.W. Ruigrok, J. Ortin, D.J. Hart, S. Cusack, The structural basis for cap binding by influenza virus polymerase subunit PB2, *Nat. Struct. Mol. Biol.* 15 (2008) 500-506.
- [35] A. Dias, D. Bouvier, T. Crepin, A.A. McCarthy, D.J. Hart, F. Baudin, S. Cusack, R.W. Ruigrok, The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit, *Nature* (2009).
- [36] P. Yuan, M. Bartlam, Z. Lou, S. Chen, J. Zhou, X. He, Z. Lv, R. Ge, X. Li, T. Deng, E. Fodor, Z. Rao, Y. Liu, Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site, *Nature* (2009).
- [37] S. Morino, H. Hazama, M. Ozaki, Y. Teraoka, S. Shibata, M. Doi, H. Ueda, T. Ishida, S. Uesugi, Analysis of the mRNA cap-binding ability of human eukaryotic initiation factor-4E by use of recombinant wild-type and mutant forms, *Eur. J. Biochem.* 239 (1996) 597-601.
- [38] M.T. Miller, J.J. Higgin, T.M. Hall, Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p, *Nat. Struct. Mol. Biol.* 15 (2008) 397-402.
- [39] R. Worch, A. Niedzwiecka, J. Stepinski, C. Mazza, M. Jankowska-Anyszka, E. Darzynkiewicz, S. Cusack, R. Stolarski, Specificity of recognition of mRNA 5' cap by human nuclear cap-binding complex, *RNA* 11 (2005) 1355-1363.
- [40] G. Hu, P.D. Gershon, A.E. Hodel, F.A. Quioco, mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 7149-7154.
- [41] C.J. Brown, I. McNae, P.M. Fischer, M.D. Walkinshaw, Crystallographic and mass spectrometric characterisation of eIF4E with N7-alkylated cap derivatives, *J. Mol. Biol.* 372 (2007) 7-15.
- [42] A.E. Hodel, P.D. Gershon, X. Shi, F.A. Quioco, The 1.85 Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends, *Cell* 85 (1996) 247-256.
- [43] A. Niedzwiecka, J. Marcotrigiano, J. Stepinski, M. Jankowska-Anyszka, A. Wyslouch-Cieszyńska, M. Dadlez, A.C. Gingras, P. Mak, E. Darzynkiewicz, N. Sonenberg, S.K. Burley, R. Stolarski, Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins, *J. Mol. Biol.* 319 (2002) 615-635.
- [44] E. Izaurralde, J. Stepinski, E. Darzynkiewicz, I.W. Mattaj, A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs, *J. Cell Biol.* 118 (1992) 1287-1295.
- [45] K. Tomoo, X. Shen, K. Okabe, Y. Nozoe, S. Fukuhara, S. Morino, T. Ishida, T. Taniguchi, H. Hasegawa, A. Terashima, M. Sasaki, Y. Katsuya, K. Kitamura, H. Miyoshi, M. Ishikawa, K. Miura, Crystal structures of 7-methylguanosine 5'-triphosphate (m⁷GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m⁷GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region, *Biochem. J.* 362 (2002) 539-544.
- [46] A. Segref, I.W. Mattaj, M. Ohno, The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export, *RNA* 7 (2001) 351-360.
- [47] R.J. Maraia, M.A. Bayfield, The La protein-RNA complex surfaces, *Mol. Cell* 21 (2006) 149-152.
- [48] A. Haghighat, N. Sonenberg, eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure, *J. Biol. Chem.* 272 (1997) 21677-21680.
- [49] M. Ptushkina, T. von Der Haar, M.M. Karim, J.M. Hughes, J.E. McCarthy, Repressor binding to a dorsal regulatory site traps human eIF4E in a high cap-affinity state, *EMBO J.* 18 (1999) 4068-4075.
- [50] G.C. Scheper, B. van Kollenburg, J. Hu, Y. Luo, D.J. Goss, C.G. Proud, Phosphorylation of eukaryotic initiation factor 4E markedly reduces its affinity for capped mRNA, *J. Biol. Chem.* 277 (2002) 3303-3309.
- [51] T. von der Haar, P.D. Ball, J.E. McCarthy, Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5'-Cap by domains of eIF4G, *J. Biol. Chem.* 275 (2000) 30551-30555.
- [52] C.J. Brown, C.S. Verma, M.D. Walkinshaw, D.P. Lane, Crystallization of eIF4E complexed with eIF4GI peptide and glycerol reveals distinct structural differences around the cap-binding site, *Cell Cycle* 8 (2009) 1905-1911.
- [53] A.F. Monzingo, S. Dhaliwal, A. Dutt-Chaudhuri, A. Lyon, J.H. Sadow, D.W. Hoffman, J.D. Robertus, K.S. Browning, The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond, *Plant Physiol.* 143 (2007) 1504-1518.
- [54] K. Tomoo, Y. Matsushita, H. Fujisaki, F. Abiko, X. Shen, T. Taniguchi, H. Miyagawa, K. Kitamura, K. Miura, T. Ishida, Structural basis for mRNA Cap-Binding regulation of eukaryotic initiation factor 4E by 4E-binding protein, studied by spectroscopic, X-ray crystal structural, and molecular dynamics simulation methods, *Biochim. Biophys. Acta* 1753 (2005) 191-208.
- [55] J. Marcotrigiano, A.C. Gingras, N. Sonenberg, S.K. Burley, Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G, *Mol. Cell* 3 (1999) 707-716.
- [56] H. Matsuo, H. Li, A.M. McGuire, C.M. Fletcher, A.C. Gingras, N. Sonenberg, G. Wagner, Structure of translation factor eIF4E bound to m⁷GDP and interaction with 4E-binding protein, *Nat. Struct. Mol. Biol.* 4 (1997) 717-724.
- [57] J. Gross, N.J. Moerke, T. Von Der Haar, A.A. Lugovskoy, A.B. Sachs, J. McCarthy, G. Wagner, Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E, *Cell* 115 (2003) 739-750.
- [58] A. Yanagiya, Y.V. Svitkin, S. Shibata, S. Mikami, H. Imataka, N. Sonenberg, Requirement of RNA binding of mammalian eukaryotic translation initiation factor 4GI (eIF4GI) for efficient interaction of eIF4E with the mRNA cap, *Mol. Cell. Biol.* 29 (2009) 1661-1669.

- [59] S. Slepnev, N. Korneeva, R. Rhoads, Kinetic Mechanism for Assembly of the m7GpppG-eIF4E-eIF4G Complex, *J. Biol. Chem.* 283 (2008) 25227-25237.
- [60] S.L. Wolin, T. Cedervall, The La protein, *Annu. Rev. Biochem.* 71 (2002) 375-403.
- [61] R.J. Maraia, R.V. Intine, Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function, *Mol. Cell. Biol.* 21 (2001) 367-379.
- [62] R.V. Intine, A.L. Sakulich, S.B. Koduru, Y. Huang, E. Pierstorff, J.L. Goodier, L. Phan, R.J. Maraia, Control of transfer RNA maturation by phosphorylation of the human La antigen on serine 366, *Mol. Cell* 6 (2000) 339-348.
- [63] C.J. Yoo, S.L. Wolin, The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors, *Cell* 89 (1997) 393-402.
- [64] D.J. Van Horn, C.J. Yoo, D. Xue, H. Shi, S.L. Wolin, The La protein in *Schizosaccharomyces pombe*: a conserved yet dispensable phosphoprotein that functions in tRNA maturation, *RNA* 3 (1997) 1434-1443.
- [65] J. Kufel, C. Allmang, G. Chanfreau, E. Petfalski, D.L. Lafontaine, D. Tollervey, Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding, *Mol. Cell. Biol.* 20 (2000) 5415-5424.
- [66] D. Xue, D.A. Robinson, B.K. Pannone, C.J. Yoo, S.L. Wolin, U snRNP assembly in yeast involves the La protein, *EMBO J.* 19 (2000) 1650-1660.
- [67] H. Fan, J.L. Goodier, J.R. Chamberlain, D.R. Engelke, R.J. Maraia, 5' processing of tRNA precursors can be modulated by the human La antigen phosphoprotein, *Mol. Cell. Biol.* 18 (1998) 3201-3211.
- [68] M.A. Bayfield, T.E. Kaiser, R.V. Intine, R.J. Maraia, Conservation of a masked nuclear export activity of La proteins and its effects on tRNA maturation, *Mol. Cell. Biol.* 27 (2007) 3303-3312.
- [69] R.V. Intine, M. Dundr, T. Misteli, R.J. Maraia, Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs, *Mol. Cell* 9 (2002) 1113-1123.
- [70] G. Chakshusmathi, S.D. Kim, D.A. Robinson, S.L. Wolin, A La protein requirement for efficient pre-tRNA folding, *EMBO J.* 22 (2003) 6562-6572.
- [71] C. Alfano, D. Sanfelice, J. Babon, G. Kelly, A. Jacks, S. Curry, M.R. Conte, Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein, *Nat. Struct. Mol. Biol.* 11 (2004) 323-329.
- [72] A. Jacks, J. Babon, G. Kelly, I. Manolaridis, P.D. Cary, S. Curry, M.R. Conte, Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element, *Structure* 11 (2003) 833-843.
- [73] G. Dong, G. Chakshusmathi, S.L. Wolin, K.M. Reinisch, Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch, *EMBO J.* 23 (2004) 1000-1007.
- [74] O. Kotik-Kogan, E.R. Valentine, D. Sanfelice, M.R. Conte, S. Curry, Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein, *Structure* 16 (2008) 852-862.
- [75] M. Teplova, Y.R. Yuan, A.T. Phan, L. Malinina, S. Ilin, A. Teplov, D.J. Patel, Structural basis for recognition and sequestration of UUU(OH) 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen, *Mol. Cell* 21 (2006) 75-85.
- [76] J.L. Goodier, H. Fan, R.J. Maraia, A carboxy-terminal basic region controls RNA polymerase III transcription factor activity of human La protein, *Mol. Cell. Biol.* 17 (1997) 5823-5832.
- [77] U.M. Ohndorf, C. Steegborn, R. Knijff, P. Sondermann, Contributions of the individual domains in human La protein to its RNA 3'-end binding activity, *J. Biol. Chem.* 276 (2001) 27188-27196.
- [78] Y. Huang, R.V. Intine, A. Mozlin, S. Hasson, R.J. Maraia, Mutations in the RNA polymerase III subunit Rpc11p that decrease RNA 3' cleavage activity increase 3'-terminal oligo(U) length and La-dependent tRNA processing, *Mol. Cell Biol.* 25 (2005) 621-636.
- [79] J.E. Stefano, Purified Lupus antigen La recognizes an oligouridyate stretch common to the 3' termini of RNA polymerase III transcripts, *Cell* 36 (1984) 145-154.
- [80] M.A. Bayfield, R.J. Maraia, Precursor-product discrimination by La protein during tRNA metabolism, *Nat. Struct. Mol. Biol.* 16 (2009) 430-437.
- [81] Y. Huang, M.A. Bayfield, R.V. Intine, R.J. Maraia, Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation, *Nat. Struct. Mol. Biol.* 13 (2006) 611-618.
- [82] R.C. Deo, C.M. Groft, K.R. Rajashankar, S.K. Burley, Recognition of the rotavirus mRNA 3' consensus by an asymmetric NSP3 homodimer, *Cell* 108 (2002) 71-81.
- [83] H. Montero, C.F. Arias, S. Lopez, Rotavirus Nonstructural Protein NSP3 is not required for viral protein synthesis, *J. Virol.* 80 (2006) 9031-9038.
- [84] R. Carthew, E. Sontheimer, Origins and mechanisms of miRNAs and siRNAs, *Cell* 136 (2009) 642-655.
- [85] Y. Tomari, P.D. Zamore, Perspective: machines for RNAi, *Genes Dev.* 19 (2005) 517-529.
- [86] G. Hutvagner, M.J. Simard, Argonaute proteins: key players in RNA silencing, *Nat. Rev. Mol. Cell. Biol.* 9 (2008) 22-32.
- [87] A. Eulalio, E. Huntzinger, E. Izaurralde, Getting to the root of miRNA-mediated gene silencing, *Cell* 132 (2008) 9-14.
- [88] J. Ma, Y.R. Yuan, G. Meister, Y. Pei, T. Tuschl, D. Patel, Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein, *Nature* 434 (2005) 666-670.
- [89] Y. Wang, S. Juraneck, H. Li, G. Sheng, T. Tuschl, D.J. Patel, Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex, *Nature* 456 (2008) 921-926.
- [90] Y. Wang, G. Sheng, S. Juraneck, T. Tuschl, D.J. Patel, Structure of the guide-strand-containing argonaute silencing complex, *Nature* 456 (2008) 209-213.
- [91] J. Parker, S.M. Roe, D. Barford, Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex, *Nature* 434 (2005) 663-666.
- [92] J. Song, S. Smith, G. Hannon, L. Joshua-Tor, Crystal structure of Argonaute and its implications for RISC slicer activity, *Science* 305 (2004) 1434-1437.

- [93] Y.R. Yuan, Y. Pei, J.B. Ma, V. Kuryavyy, M. Zhadina, G. Meister, H.Y. Chen, Z. Dauter, T. Tuschl, D.J. Patel, Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage, *Mol. Cell* 19 (2005) 405-419.
- [94] A. Lingel, B. Simon, E. Izaurralde, M. Sattler, Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain, *Nat. Struct. Mol. Biol.* 11 (2004) 576-577.
- [95] J. Ma, K. Ye, D. Patel, Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain, *Nature* 429 (2004) 318-322.
- [96] J. Song, J. Liu, N. Tolia, J. Schneiderman, S. Smith, R. Martienssen, G. Hannon, L. Joshua-Tor, The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes, *Nat. Struct. Biol.* 10 (2003) 1026-1032.
- [97] A. Lingel, B. Simon, E. Izaurralde, M. Sattler, Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain, *Nature* 426 (2003) 465-469.
- [98] J.G. Doench, P.A. Sharp, Specificity of microRNA target selection in translational repression, *Genes Dev.* 18 (2004) 504-511.
- [99] B. Haley, P.D. Zamore, Kinetic analysis of the RNAi enzyme complex, *Nat. Struct. Mol. Biol.* 11 (2004) 599-606.
- [100] J. Parker, S.M. Roe, D. Barford, Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity, *EMBO J.* 23 (2004) 4727-4737.
- [101] J.S. Parker, E.A. Parizotto, M. Wang, S.M. Roe, D. Barford, Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein, *Mol. Cell* 33 (2009) 204-214.
- [102] G. Mathonnet, M.R. Fabian, Y.V. Svitkin, A. Parsyan, L. Huck, T. Murata, S. Biffo, W.C. Merrick, E. Darzynkiewicz, R.S. Pillai, W. Filipowicz, T.F. Duchaine, N. Sonenberg, MicroRNA inhibition of translation initiation *in vitro* by targeting the cap-binding complex eIF4E, *Science* 317 (2007) 1764-1767.
- [103] M. Kiriakidou, G.S. Tan, S. Lamprinak, M. De Planell-Saguer, P.T. Nelson, Z. Mourelatos, An mRNA m7G cap binding-like motif within human Ago2 represses translation, *Cell* 129 (2007) 1141-1151.
- [104] S. Iwasaki, T. Kawamata, Y. Tomari, *Drosophila* Argonaute1 and Argonaute2 employ distinct mechanisms for translational repression, *Mol. Cell* (2009).
- [105] A. Eulalio, E. Huntzinger, E. Izaurralde, GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay, *Nat. Struct. Mol. Biol.* 15 (2008) 346-353.
- [106] L. Kinch, N. Grishin, The human Ago2 MC region does not contain an eIF4E-like mRNA cap binding motif, *Biol. Direct* 4 (2009) 2.
- [107] S.W. Ding, O. Voinnet, Antiviral immunity directed by small RNAs, *Cell* 130 (2007) 413-426.
- [108] H.Y. Chen, J. Yang, C. Lin, Y.A. Yuan, Structural basis for RNA-silencing suppression by *Tomato aspermy virus* protein 2b, *EMBO Rep.* 9 (2008) 754-760.
- [109] J.M. Vargason, G. Szittyá, J. Burgyán, T.M. Hall, Size selective recognition of siRNA by an RNA silencing suppressor, *Cell* 115 (2003) 799-811.
- [110] K. Ye, L. Malinina, D.J. Patel, Recognition of small interfering RNA by a viral suppressor of RNA silencing, *Nature* 426 (2003) 874-878.
- [111] P. Brodersen, O. Voinnet, The diversity of RNA silencing pathways in plants, *Trends Genet.* 22 (2006) 268-280.
- [112] M. She, C.J. Decker, D.I. Svergun, A. Round, N. Chen, D. Muhlrad, R. Parker, H. Song, Structural basis of Dcp2 recognition and activation by Dcp1, *Mol. Cell* 29 (2008) 337-349.
- [113] S. Bail, M. Kiledjian, Tri- to be mono- for bacterial mRNA decay, *Structure* 17 (2009) 317-319.
- [114] H. Celesnik, A. Deana, J. Belasco, Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal, *Mol. Cell* 27 (2007) 79-90.
- [115] A. Deana, H. Celesnik, J. Belasco, The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal, *Nature* 451 (2008) 355-358.
- [116] A. Callaghan, M. Marcaida, J. Stead, K. McDowall, W. Scott, B. Luisi, Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover, *Nature* 437 (2005) 1187-1191.
- [117] X. Jiang, J. Belasco, Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA, *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004) 9211-9216.
- [118] T. Tomcsányi, D. Apirion, Processing enzyme ribonuclease E specifically cleaves RNA I. An inhibitor of primer formation in plasmid DNA synthesis, *J. Mol. Biol.* 185 (1985) 713-720.