RNA-based recognition and targeting: sowing the seeds of specificity

Stanislaw A. Gorski[1], Jörg Vogel[1,2] and Jennifer A. Doudna[3–8]

Abstract | RNA is involved in the regulation of multiple cellular processes, often by forming sequence-specific base pairs with cellular RNA or DNA targets that must be identified among the large number of nucleic acids in a cell. Several RNA-based regulatory systems in eukaryotes, bacteria and archaea, including microRNAs (miRNAs), small interfering RNAs (siRNAs), CRISPR RNAs (crRNAs) and small RNAs (sRNAs) that are dependent on the RNA chaperone protein Hfq, achieve specificity using similar strategies. Central to their function is the presentation of short 'seed sequences' within a ribonucleoprotein complex to facilitate the search for and recognition of targets.

[1]*Institute of Molecular Infection Biology, University of Würzburg, Josef-Schneider-Strasse 2/D15, D-97080 Würzburg, Germany.*
[2]*Helmholtz Institute for RNA based Infection Research (HIRI), University of Würzburg, D-97080 Würzburg, Germany.*
[3]*Department of Molecular and Cell Biology, University of California, Berkeley.*
[4]*Howard Hughes Medical Institute, University of California, Berkeley.*
[5]*California Institute for Quantitative Biosciences, University of California, Berkeley.*
[6]*Department of Chemistry, University of California, Berkeley.*
[7]*Lawrence Berkeley National Laboratory, Berkeley.*
[8]*Innovative Genomics Initiative, University of California, Berkeley, California 94720, USA.*

Correspondence to J.V. and J.A.D.
joerg.vogel@uni-wuerzburg.de; doudna@berkeley.edu

doi:10.1038/nrm.2016.174
Published online 15 Feb 2017

RNA has a central role in the cell because it directs multiple mechanistically distinct processes[1]. For example, tRNAs play a fundamental part in protein synthesis, small nuclear RNAs (snRNAs) have key roles in mRNA splicing, and small nucleolar RNAs (snoRNA) and small Cajal body-specific RNAs (scaRNAs) guide RNA modifications. Moreover, there are numerous and diverse classes of non-coding RNA that direct gene regulation or defend the genome to ensure its integrity. Several properties make RNA particularly well suited for controlling these processes[1,2]. Most importantly, RNA can regulate nucleic acids in a sequence-specific manner by base-pairing with defined RNA or DNA targets. It can also be rapidly synthesized or processed in response to a signal and it is functional without the need to be translated. Moreover, regulatory RNA molecules can be co-degraded with their targets on binding to each other, enabling them to impart rapid responses to a change in cellular status with quantitative characteristics that are distinct from protein-based cellular regulation[2,3].

Despite these advantages, RNA molecules must overcome several limitations to ensure that they specifically regulate cellular targets through direct base-pairing interactions. Although regulatory RNAs range in length from twenty to hundreds of nucleotides, target binding usually involves partial complementarity with only a short stretch of nucleotides[4]. This increases the potential for unintended interactions between RNA species in a cell. Therefore, solving the problem of target specificity requires that regulatory RNAs contain functional sequences that are used to search for, and interact with, regulated transcripts. However, as RNA sequences are not chemically diverse (RNA consists of four nucleotides that come in two sizes and consist of planar bases containing hydrogen bond donors and acceptors)[5], their secondary and tertiary structures are important in defining the sequences involved in target recognition. Functional secondary and tertiary structures compete with energetically similar alternative conformations, which in some instances sequester the sequences that define target specificity; this favours non-target interactions and limits the rate of regulatory RNA–target associations. In addition, interactions involving naked RNA and DNA oligonucleotides must overcome thermodynamic and kinetic barriers to efficiently regulate biological processes[5–8]. For example, duplexes containing 8 nucleotides are unstable, whereas a 20-nucleotide RNA–RNA duplex is nearly irreversible under physiological conditions[8].

The ubiquitous nature of RNA as a regulatory molecule suggests that RNA-based systems have evolved important mechanisms to ensure target specificity despite the thousands (if not millions) of expressed non-target RNAs that share a degree of complementarity. One way to achieve specificity is for RNAs to be subdivided into biochemically distinct regions, with specific subregions that interrogate potential RNA or DNA targets. Indeed, this feature has been identified in several RNA-based systems using molecular, genomic and computational approaches. These subregions are often referred to as 'seed sequences' — short stretches of nucleotides that are involved in the initial search by RNAs for, and binding of RNAs to, their targets and that disproportionately contribute to regulation compared with other regions of the RNA[4,9].

Early studies of RNA-based regulatory systems suggested that the folding of RNA into secondary and tertiary structures to expose specific sequences used for binding to their targets was a general principle of RNA-based recognition. Examples of these structural

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

motifs include the anticodon loop in tRNA[10] and loop–loop kissing interactions in antisense RNAs from plasmids and phages[11]. However, almost all RNAs associate with RNA-binding proteins, which have important structural, regulatory or catalytic roles in defining the behaviour of RNA. As the mechanisms of more RNA-based regulatory systems are elucidated, it is becoming clear that regulatory RNAs associate with specific proteins, enabling them to function as guides to bring each protein to its target site[8,12–16]. Such dedicated RNA-binding proteins alter the properties of nucleic acid hybridization, including how complementary nucleic acid sequences find each other, bind and dissociate. Importantly, several recent structural studies have revealed how the seed sequence is presented within the context of the respective ribonucleoprotein (RNP) particle to ensure its precise targeting and regulation[17–27], and single-molecule studies have begun to reveal the dynamics of the targeting reaction and to confirm the role of the seed sequence in RNA–target interactions[8,28–30].

This Review focuses on seed pairing mechanisms in three distinct systems that use RNA as a guide, namely small interfering RNA (siRNA) and microRNA (miRNA)-mediated gene silencing, bacterial CRISPR–Cas adaptive immunity and Hfq-dependent small RNA (sRNA)-guided gene regulation (FIG. 1). We briefly introduce each system before describing how they handle the issue of specificity and highlighting common principles and major differences between them.

Targeting by miRNAs and siRNAs

The discovery of miRNAs and siRNAs revolutionized the idea that RNA is a regulatory molecule in eukaryotic organisms. Computational and experimental approaches suggest that metazoan genomes encode hundreds of miRNAs[31,32], and it is estimated that the most conserved miRNAs control the expression of approximately two-thirds of human genes[33]. Although the origin and precursors of miRNAs and siRNAs differ, they share in part common biogenesis pathways and they are both loaded onto Argonaute (AGO) proteins, which they guide to mRNA targets for RNA silencing[16,34,35].

Biogenesis. miRNAs are sRNAs of ~21–23 nucleotides in length that are encoded by the genome and some viruses. The majority of metazoan miRNAs originate from independent loci, although some are processed from within introns. Their primary transcripts contain stem–loop structures, and consecutive endonuclease cleavages by two enzymes, Drosha in the nucleus and Dicer in the cytoplasm, generate a miRNA duplex containing a 5′ monophosphate and a 3′ hydroxyl. The mature miRNA strand is loaded into the AGO protein within the RNA-induced silencing complex (RISC)[34,36] (FIG. 1a). The miRNA within RISC associates with target mRNAs through base-pairing, leading to translational repression and/or mRNA destabilization[37] or, in some cases when the miRNA is highly complementary to its target, mRNA cleavage (FIG. 1a,b). siRNAs differ from miRNAs in that they are excised from long, fully complementary, double-stranded RNAs; these can originate from transgenes, viral infections and repetitive elements[34]. After the siRNA precursors are cleaved by Dicer into duplexes of ~21 nucleotides they are loaded into the siRISC (FIG. 1b), from which they recognize and associate with target transcripts through base-pairing, with high levels of complementarity; this enables the endonucleolytic cleavage of the scissile bond opposite the tenth and eleventh guide nucleotide by AGO within siRISC[34].

To identify physiological roles of miRNA and siRNA, early studies focused on predicting and identifying their target transcripts in the cell. Several fundamental observations provided molecular insight into how they select their targets. Sequence analysis revealed that the highly conserved 5′ regions of miRNAs are complementary to sequences in 3′ untranslated regions (3′ UTRs) of mRNA that had previously been linked to post-transcriptional regulation[38–42]. Distinct contributions of specific siRNA and miRNA regions to target regulation support the presence of a seed region; the 5′ nucleotides contribute more to binding and target regulation than the nucleotides at the 3′ end[43–46]. Consequently, target transcripts can be identified by evaluating the conservation of direct Watson–Crick base-pairing between miRNA seed nucleotides (specifically, nucleotides 2–7) (FIG. 1e) and sequences in the 3′ UTRs of mRNA. Targeting can also be facilitated by additional sequence elements, such as an unpaired adenosine in the target sequence just upstream of the seed sequence[41] and, in rare cases, supplementary pairing between the target and the 3′ end of the miRNA, involving miRNA nucleotides 13–16 (REF. 47).

siRNA seed sequence presentation. Initial insights into the molecular basis behind the importance of the seed sequence were obtained from crystal structures of the bacterial Ago protein from *Thermus thermophilus*. Even though these structures used guide DNAs, they revealed how guide strands mediate RNA recognition[26,27,48] (FIG. 2a). *T. thermophilus* Ago containing a 5′-phosphorylated 21-nucleotide DNA guide strand maintained the same overall bilobal structure as the apoprotein[48] (FIG. 2a,b). The DNA guide is nestled within a basic channel spanning the lobes that harbour the PAZ domain and the PIWI domain, and there are extensive contacts between its backbone phosphates and basic residues in the domains and linkers present in the protein. The trajectory of the guide is defined by the anchoring of its 5′ monophosphate into a pocket within the Mid domain and of its 3′ end in the PAZ domain. The conformation of the seed sequence at nucleotides 2–6 reveals how its presentation within the *T. thermophilus* Ago RNP is important for RNA recognition. Specifically, in agreement with earlier biochemical and biophysical studies[45,49], these nucleotides are stacked forming an A-form helix with their Watson–Crick edges exposed to the solvent, a favourable, pre-organized conformation for interacting with a target mRNA (FIG. 2a,b). Subsequent nucleotides are threaded through a narrow channel in the *T. thermophilus* Ago protein

Loop–loop kissing interactions
Watson–Crick base pairing between the loop nucleotides of two RNA stem loops.

Adaptive immunity
A specific response to an infection by a pathogen based on prior exposure to that pathogen.

RNA-induced silencing complex
(RISC). A ribonucleoprotein complex of an Argonaute protein and an RNA.

Scissile bond
A covalent bond that can be broken by an enzymatic reaction.

PAZ domain
(PIWI–Argonaute–Zwille domain). A domain present in Argonaute proteins that is involved in binding to the 3′ end of the guide.

PIWI domain
(P-element-induced wimpy testis domain). A domain present in Argonaute proteins that contains an RNase H-like active site.

A-form helix
A secondary structure motif found in RNA in which bases are tilted with respect to the helix axis.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

and are therefore inaccessible to mRNA. Introducing a kink into the guide strand by positioning two Arg residues between nucleotides 10 and 11 renders the binary *T. thermophilus* Ago protein–guide complex catalytically inactive and prevents the formation of the undistorted helical conformation that is required for mRNA cleavage[26,27]. However, on the addition of a 20-nucleotide target RNA to the *T. thermophilus* Ago protein–guide complex, the seed region forms an A-form DNA–RNA helix with the target and undergoes a conformational change that relieves the non-helical geometry, enabling cleavage of the target

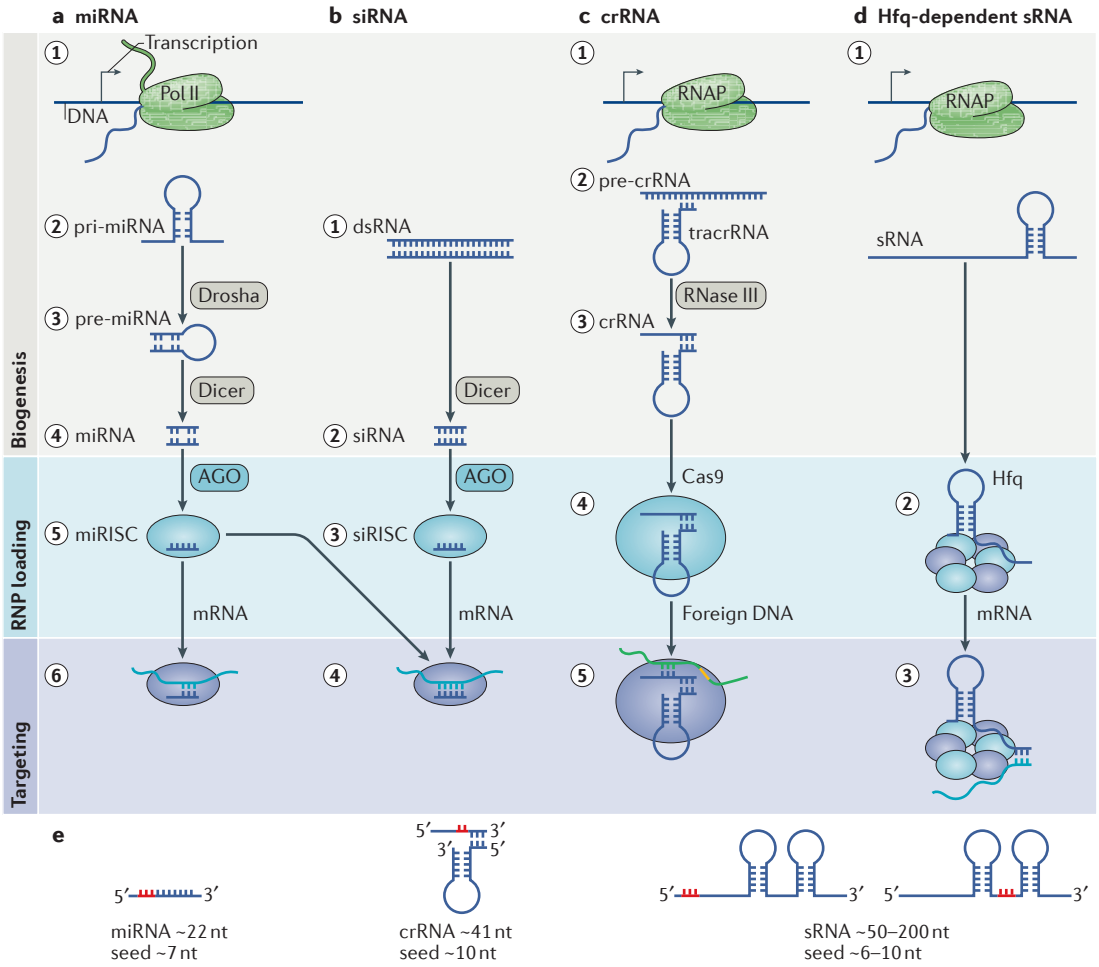


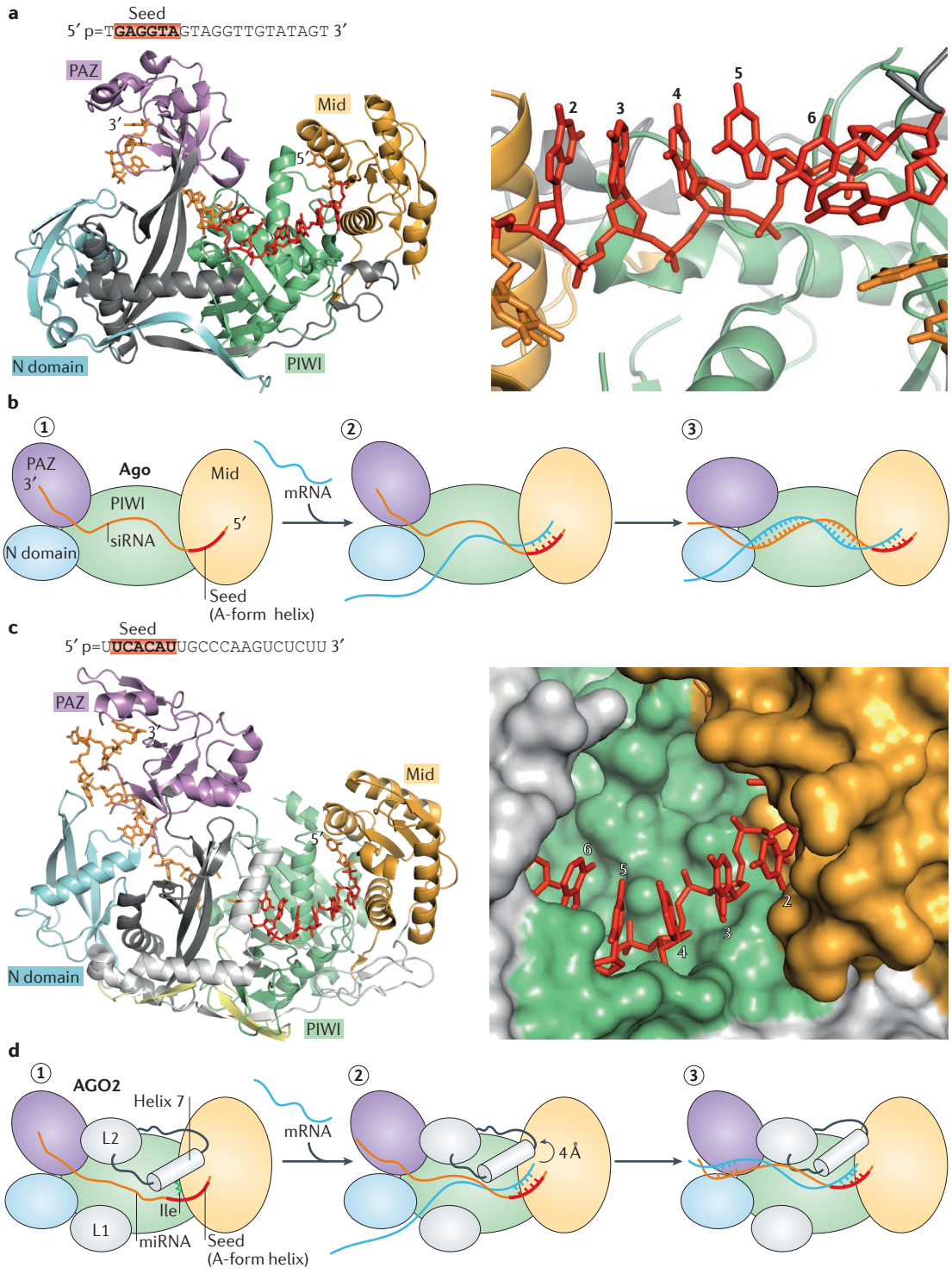
Figure 1 | **The biogenesis, loading and targeting of RNA in distinct RNA-based regulatory systems. a** | MicroRNAs (miRNAs) are generally transcribed by RNA polymerase II (Pol II) (step 1) to produce primary miRNAs (pri-miRNAs) (step 2); these are cleaved by Drosha in the nucleus to produce a stem–loop structure — the pre-miRNA (step 3) — which is exported to the cytoplasm and processed by Dicer to form the mature miRNA duplex (step 4), a single strand of which is loaded onto an Argonaute (AGO) protein as part of the miRNA-induced silencing complex (miRISC) (step 5). miRISC binds to target mRNAs through partial complementary base-pairing (step 6), which leads to changes in mRNA stability and/or translational repression or, in cases of high complementarity (indicated by the arrows), mRNA cleavage (not shown). **b** | Small interfering RNAs (siRNAs) are generated from double-stranded RNA (dsRNA) precursors (step 1), which are cleaved by Dicer into siRNA duplexes (step 2) and loaded onto AGO proteins as part of the siRISC (step 3). The siRISC binds to target mRNAs by highly complementary base-pairing (step 4), which leads to transcript cleavage (not shown). **c** | The type II CRISPR RNA (crRNA) locus contains repeat sequences that are interspersed with sequences from foreign DNA. It is transcribed by RNA polymerase (RNAP) (step 1) to produce a pre-crRNA; a small transactivating crRNA (tracrRNA) interacts with the spacer sequence of crRNA to form a duplex (step 2). The duplex is cleaved by RNase III to produce the mature crRNA (step 3). The mature crRNA is loaded into Cas9 nuclease (step 4), from which it binds to a complementary foreign DNA (green) containing a protospacer-adjacent motif (PAM; yellow) (step 5), resulting in DNA cleavage and destruction (not shown). **d** | Hfq-dependent small RNAs (sRNAs) are transcribed from independent genes or processed from longer RNA transcripts (step 1). They associate with the RNA chaperone Hfq (step 2), which facilitates base-pairing to their target mRNAs (step 3). On targeting an mRNA, the sRNA can positively or negatively affect RNA stability or translation (not shown). **e** | Properties of miRNAs, crRNAs and Hfq-dependent sRNAs. miRNAs are ~22 nucleotides (nt) in length and contain an ~7-nt 'seed sequence' towards their 5′ end. Mature crRNAs are ~41 nt in length and contain a seed sequence of ~10 nt towards the 3′ end of the crRNA (which is part of the crRNA–tracrRNA complex). Bacterial Hfq-dependent sRNAs vary in length from 50 to 200 nt and they often contain seed sequences of ~6–10 nt, which are generally found at the 5′ end but can be found throughout the sRNA. The position of each seed sequence is indicated in red.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Entropy penalty
The thermodynamic cost associated with a loss of conformational entropy on the immobilization of a molecule in a fixed configuration.

mRNA[26,27] (FIG. 2b). The importance of mRNA–seed sequence complementarity is highlighted by the fact that cleavage is impaired on the introduction of mismatches or insertions within the seed region in the guide DNA[27]. Therefore, the pre-organization of the seed in an A-form helix within *T. thermophilus* Ago facilitates its interaction with the target mRNA and offsets the entropy penalty of binding target mRNA, enhancing the interaction affinity up to 300-fold[50].

miRNA seed sequence presentation. Insight into the presentation of miRNA seed sequences and their role in target recognition awaited the structural elucidation of eukaryotic AGO proteins. Crystal structures of human and yeast AGO proteins bound to cellular guide RNAs[23,24] (FIG. 2c) or the human miRNA miR-20a[25] revealed that these AGO proteins form a similar bilobal domain protein architecture to, and bind RNA in the same way as, that of *T. thermophilus* Ago (FIG. 2c). Despite the



© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

High-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP). A sequencing method based on the ultraviolet crosslinking of RNA–protein complexes that is used to identify RNA ligands of RNA-binding proteins.

Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP). A sequencing method that uses photoactivatable ribonucleosides to crosslink RNA and protein to identify RNAs associated with a particular RNA-binding protein.

heterogeneity in RNA sequences bound by the yeast and human AGO proteins in the binary complex, nucleotides 2–6 of the guide strand in human AGO2 unambiguously adopted a similar splayed helical conformation to that of *T. thermophilus* Ago, with Watson–Crick edges exposed to the solvent (FIG. 2c). After nucleotide 6 in the seed sequence, the trajectory of the guide differs because human AGO inserts an Ile side-chain from helix-7, between the two faces of bases 6 and 7, inducing a kink that breaks the A-form helix structure of the guide (FIG. 2d). Formation of the guide RNA–target RNA duplex coincides with helix-7 moving away from the guide to avoid a steric clash with the target RNA region that pairs to nucleotides 6–7 in the guide RNA; this relieves the kink at this guide position and facilitates the formation of an A-form helix that can encompass the entire seed region of the miRNA[18] (FIG. 2d). The target-bound conformation of the miRNA is stabilized when helix-7 interacts with the minor groove of the RNA duplex. The degree of stability probably depends on the extent of complementarity between the seed and the target, which dictates the structure of the duplex minor groove and the creation of a new binding surface for helix-7, thus functioning as a checkpoint for correctly identifying targets[18]. The repositioning of helix-7 also widens the channel between the N and PAZ domains, enabling nucleotides 14–18 of the miRNA to adopt a stacked A-form state with Watson–Crick edges that are exposed to the solvent and opening the 3′ region of the guide for potential supplementary base-pairing[18] (FIG. 2d). Thus, the eukaryotic structures suggest that there is a stepwise model for miRNA targeting and correct target validation.

◀ Figure 2 | **Presentation of seed sequences in miRNA and siRNA. a** | Crystal structure of *Thermus thermophilus* Argonaute (Ago) containing a 21-base DNA guide (shown in orange and red, where red denotes the 'seed sequence'; the sequence of the guide is shown above the structure), the 5′ phosphate (p) of which is bound within a pocket in the Mid domain and the 3′ end of which is bound in the PAZ domain (left)[27] (Protein Data Bank identifier (PDB ID): 3DLH; see Further information). A close-up of the 5′ region of the DNA guide with nucleotides 2–6 of the seed sequence presented in a near A-form helical conformation is shown on the right. **b** | Schematic showing the presentation of the small interfering RNA (siRNA) seed sequence and the subsequent targeting of mRNA. Ago binds the guide so that its 5′ end is bound in a pocket in the Ago Mid domain and its 3′ end is bound by the PAZ domain of Ago (step 1). The seed sequence (nucleotides 2–7, coloured red) is presented in an A-form helix conformation and this is subsequently bound by an mRNA target (light blue) (step 2). Duplex formation is accompanied by a conformational change in Ago that relieves the non-helical geometry of the guide and produces a catalytically active state (step 3). **c** | Crystal structure of human AGO2 (left) containing a 21-nucleotide superoxide dismutase 1 (*SOD1*) guide RNA (shown in orange and red, where red depicts the seed sequence; the sequence of the guide is shown above the structure)[18] (PDB ID: 4W5N). A close-up of nucleotides 2–5 of the seed region of the guide RNA, which are presented in a near A-form helix conformation and exposed to solvent, is shown on the right. **d** | Schematic of microRNA (miRNA)-mediated targeting. AGO2 binds to miRNA (orange and red) with the 5′ end bound in a pocket in the Mid domain and the 3′ end bound by the PAZ domain; nucleotides 2–6 of the seed sequence (coloured red) are presented as an A-form helix. Helix 7 of AGO2 (shown as a cylinder) inserts an Ile side chain (green bar) between nucleotides 6 and 7, breaking the helical configuration of the RNA and sterically hindering guide–target duplex formation (step 1). Base-pairing between the seed region and the complementary mRNA target coincides with movement of helix 7, which allows guide–target duplex formation. Helix 7 interacts with the minor groove of the duplex to monitor shape complementarity (step 2). The conformational change in AGO2 allows base-pairing between the seed region and the target, as well as potential 3′ supplementary base-pairing with the 3′ region of miRNA (step 3).

Insight into the dynamics of the target-search process, and support for the stepwise model of miRNA targeting, has come from single-molecule experiments; these studies monitored eukaryotic AGO loaded with a fluorescently labelled guide RNA that was bound to an immobilized, labelled RNA target[8,28,51]. Incorporation of the guide RNA (let-7a or miR-21) into an RNP greatly increases the rate constant with which it binds to the target, as compared with that of a naked guide oligonucleotide, to the extent that finding the target becomes limited by diffusion[8,28,51]. Consistent with the presentation of the seed region in an A-form helix that favours target binding, acceleration of target binding by AGO2 is dependent on complementarity between the seed sequence and the target[8]. The observation from the human AGO2 structures that only guide nucleotides 2–6 are initially presented in a helical configuration, for the primary interrogation of targets, is supported by the fact that different mismatches within the seed have different effects on guide–target binding; mismatches at nucleotides 2–5 of the guide reduce the association rate more than mismatches at positions 6–8 (REFS 8,18,28). This further indicates that the arrangement of 5′ seed nucleotides in a helical form facilitates the search for the target and that, when a potential sequence match has been found, the conformational change involving helix-7 enables the rest of the seed to be interrogated. Complementarity between the seed sequence and target also determines the rate at which AGO dissociates from the target, enabling the complex to discriminate between seed-matched and mismatched targets[8]. In the absence of a matched seed sequence, the RNP finds targets more slowly and binds to them less stably[8]. The interrogation of mRNAs for target binding sites is facilitated by a one-dimensional scanning mechanism that uses sub-seed interactions before stably binding to the mRNA[28].

Genome-wide studies describing the miRNAs that are bound to AGO proteins on a global scale also support an important role for seed sequences in recognizing mRNA targets. For example, experiments in which AGO and miRNA were crosslinked by high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP)[52] and by photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP)[53] revealed hundreds of AGO-binding sites in transcripts from the mouse neocortex and human embryonic kidney cells, respectively. These binding sites were enriched for sequences that were complementary to the seeds of the most abundant miRNAs in the relevant cell types. Transcripts containing direct binding sites were more likely to be regulated on overexpression of the complementary miRNA. However, in a minority of cases (27% for HITS-CLIP and 7% for PAR-CLIP), AGO-binding sites did not contain exact seed matches, suggesting that they may contain bulges or mismatches in the seed region. Based on shape complementarity between helix-7 in human AGO2 and the miRNA–mRNA duplex, it is expected that RISC would have a lower affinity for these miRNA–target RNA interactions[18]. However, modification of the

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

crosslinking approach to ligate miRNA–target RNA duplexes in several related approaches called crosslinking, ligation and sequencing of hybrids (CLASH), as well as the *in vivo* PAR-CLIP with ligation[54] and covalent ligation of endogenous AGO-bound RNAs (CLEAR)-CLIP, suggests that non-canonical seed interactions involving G:U wobble pairs, mismatches and bulges are more frequent than canonical exact complementary seed pairings (which make up only 37% of seed interactions[55]) and hint at a greater involvement of the miRNA 3′ region in determining interactions than previously thought[55,56]. Seedless interactions were determined in 16% of cases[55], although these were poorly conserved and weakly regulated by miRNAs. In general, although some non-seed interactions modestly contribute to target regulation[54–56], high-throughput analyses indicate that many of the non-canonical sites bound by miRNAs do not result in the regulation of the target[57].

Taken together, incorporation of siRNA and miRNA into an RNP complex with AGO enables the presentation of the seed sequence in a pre-organized A-form helix. This facilitates the RNA-mediated target-search process by favouring interactions with the target and dramatically alters the thermodynamic and kinetic properties of guide–target hybridization. Conformational changes within the protein and guide RNA on guide–target pairing provide a basis for specificity in gene regulation.

CRISPR–Cas-mediated defence

Many bacteria and most archaea contain an RNA-mediated silencing mechanism that provides adaptive immunity against bacteriophages and plasmids[58]. This mechanism uses CRISPR RNAs (crRNAs) generated from genomic CRISPR loci for the sequence-specific detection and elimination of invading viruses and plasmids. crRNA sequences originate from previous invading nucleic acids and act as guides for various CRISPR-associated Cas proteins, thus providing an inherited adaptive immunity against re-infections. CRISPR loci can be classified into five main systems, depending on their Cas protein composition[59], but all contain a nuclease. Type I and type III systems use crRNA that is incorporated into a multisubunit complex to execute their interference activity. Type II CRISPR–Cas systems are one of the simplest, requiring only two RNAs, crRNA and transactivating crRNA (tracrRNA), and a single Cas protein, Cas9, to function as RNA-guided endonucleases (FIG. 3). A simplified chimeric single-guide RNA (sgRNA) (FIG. 3a) containing functionally important regions of crRNA and tracrRNA can be used to target any sequence to generate double-strand breaks, and it is used in various genome-editing approaches[60,61]. Below, we focus on the type II system owing to its simplicity and the amount of structural information available on its mechanism of RNA recognition.

Identifying foreign DNA using crRNA. CRISPR–Cas adaptive immunity consists of three phases: immunization, where new spacer sequences from invading nucleic acids are incorporated into the CRISPR array to function as a molecular memory of prior infections; expression, which involves the transcription and processing of crRNAs; and interference, whereby the crRNAs are used as guides to identify foreign nucleic acids and facilitate their cleavage and clearance by associated nucleases[58]. RNA plays a fundamental part in the expression and interference phases. Type II CRISPR–Cas systems transcribe their loci, which contain repeats and spacers, into a full-length pre-crRNA. The subsequent processing of this transcript by endogenous RNase III requires duplex formation between the anti-repeat sequence in the tracrRNA and a repeat sequence within the pre-crRNA[62–64] (FIG. 1c).The resulting mature crRNA contains a 20-nucleotide spacer-derived guide sequence (with complementarity to foreign DNA) at the 5′ end and a 21-nucleotide repeat sequence that base-pairs to the 5′ end of the tracrRNA (FIG. 1c), which probably has a structural role and is required for target DNA binding and site-specific cleavage by Cas9 (REF. 62).

The Cas9–crRNA–tracrRNA ternary complex must identify invading DNA molecules to cause inference. Identifying targets is a stepwise process involving not only base-pairing between DNA and the crRNA sequence (the complementary site in the DNA is called the protospacer) but also recognition of a short sequence motif called the protospacer-adjacent motif (PAM) on the non-complementary strand of the DNA by Cas9. The PAM is required for Cas9 to recognize and stably bind to DNA (FIGS 1c,3c); its sequence varies between CRISPR–Cas systems but in *Streptococcus pyogenes*, for example, it is composed of a 5′-NGG sequence (where 'N' is any nucleobase) immediately flanking the protospacer[62]. Biochemical experiments revealed that mismatches introduced in the 3′ region of the guide sequence of crRNA prevent Cas9-mediated cleavage, leading to the proposal that there is a seed region within the guide that is essential for target recognition[65] (FIG. 3a). The existence of a seed sequence is further supported by the requirement of a contiguous stretch of 13 complementary base-pairs between the 3′ region of the RNA guide and the complementary DNA target strand proximal to the PAM; by contrast, 6 contiguous mismatches are tolerated between complementary DNA and crRNA at the 5′ end of the guide. When the target DNA has been identified, it is cleaved by Cas9 using two domains that are homologous to the HNH endonuclease and RuvC endonuclease: the HNH domain cleaves the DNA strand complementary to the crRNA guide and the RuvC domain cleaves the non-complementary strand[65].

crRNA seed presentation. Insight into how the seed region of the guide RNA contributes to target DNA recognition and binding was first obtained from the crystal structure of Cas9 bound to sgRNA without a target[17]. Apo-Cas9 forms a bilobal structure with one lobe containing the HNH, RuvC and carboxy-terminal domains and the other comprising a large helical domain. Binding of the sgRNA induces conformational changes in the protein[20]; the C-terminal domain (which interacts with the PAM) forms a nucleic acid-binding groove that can accommodate the DNA duplex in which PAM is located (FIG. 3b). In this conformation, the sgRNA is bound in

Crosslinking, ligation and sequencing of hybrids (CLASH). A sequencing method used to identify RNAs and their targets by ligating them together when they are bound to a specific RNA-binding protein.

HNH domain
A nuclease domain within Cas9 that is related to McrA-like restriction endonucleases.

RuvC domain
A nuclease domain within Cas9 that is related to the RuvC endonuclease that cuts Holliday junctions during homologous recombination.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

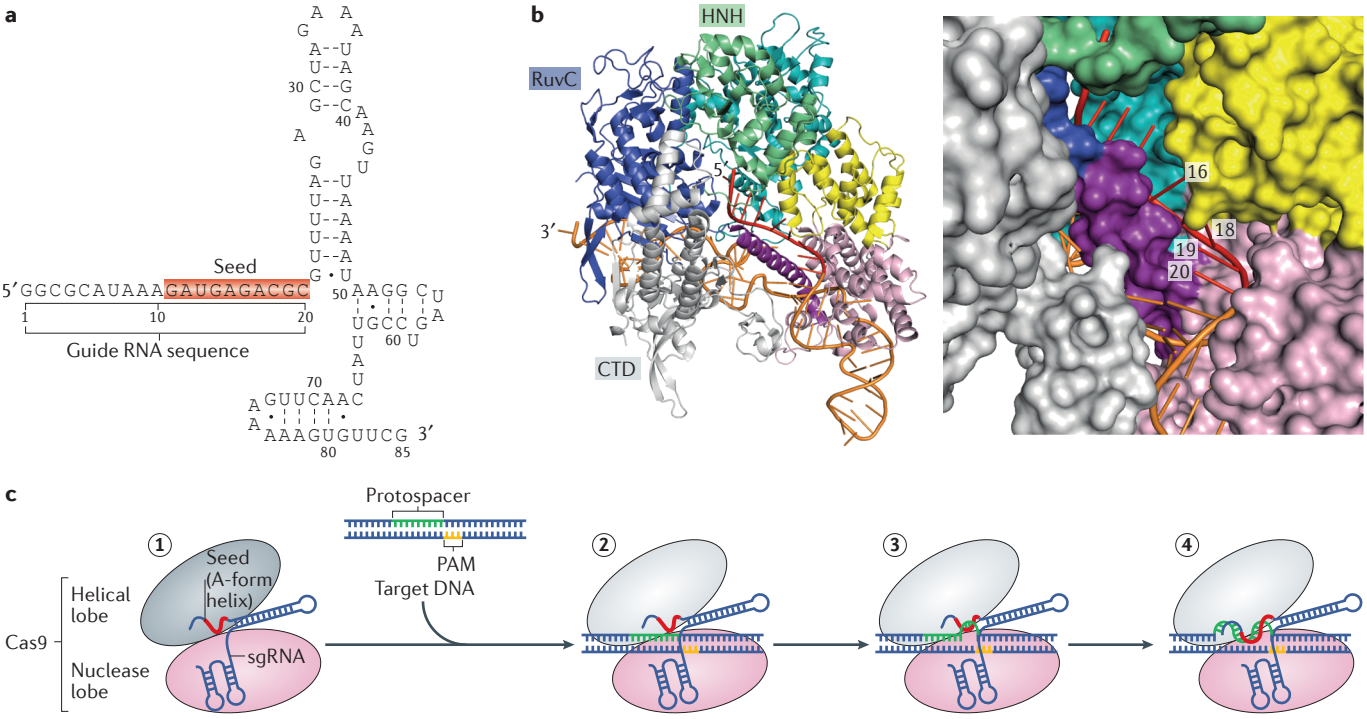


Figure 3 | **crRNA seed presentation. a** | Secondary structure of the single-guide RNA (sgRNA) that is present in the Cas9 crystal structures in part **b**, which is composed of a CRISPR RNA (crRNA) and a transactivating crRNA (tracrRNA) that are fused together. The 'seed sequence' is highlighted in red. Canonical and non-canonical base-pairing is depicted as bars and dots, respectively. **b** | Ribbon diagram of the crystal structure of the *Streptococcus pyogenes* Cas9–sgRNA binary complex (left)[17] (Protein Data Bank identifier: 4ZT0; see Further information). The guide RNA is shown in orange and red, where red depicts the seed sequence; for Cas9, the nuclease lobe contains the HNH endonuclease (green), the RuvC endonuclease (blue) and the carboxy-terminal domain (CTD; grey), and the helical recognition lobe is shown in light purple, yellow and cyan. A close-up of the presentation of the guide RNA by Cas9, in the positively charged channel between the two lobes of Cas9, is shown in the surface representation (right); nucleotides 19–20 of the guide are exposed to solvent. **c** | Schematic of DNA targeting by the Cas9–crRNA complex. Cas9 binds to the crRNA–tracrRNA complex; the crRNA seed sequence (nucleotides 11–20, shown in red) is presented in an A-form helix with nucleotides 19–20 exposed to the solvent (step 1). Target binding is initiated when Cas9 recognizes a protospacer-adjacent motif (PAM) sequence (yellow) on the non-complementary DNA strand (step 2). Strand separation is induced by interactions between Cas9 and the DNA duplex; the PAM proximal seed nucleotides (nucleotides 19–20) nucleate target binding and crRNA–DNA duplex formation (step 3); the DNA protospacer sequence (which is complementary to the crRNA) is shown in green. The propagation of helix formation enables Cas9 to acquire an active catalytic state (step 4). Parts **a** and **c** are modified from Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A. Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015) (REF. 17). Reproduced with permission from AAAS.

a positively charged channel between the two lobes of Cas9 (REF. 17) (FIG. 3b). Of the 20 nucleotides within the guide, only the nucleotides located in the seed region towards the 3′ end of the crRNA (nucleotides 11–20) are visible within the narrow channel. This region forms an A-form helix, similar to that seen in AGO–guide structures, with nucleotides 11–13 and the PAM proximal nucleotides (that is, nucleotides 19–20) exposed to the solvent[17] (FIG. 3b,c). Similarly to the AGO structures, a kink is introduced into the guide RNA by the insertion of an amino acid — this time a Tyr — between nucleotides 15 and 16 of the seed; this is relieved on target binding, which causes the Tyr to rotate by 120° (REF. 17). This pre-organized helical conformation places the seed sequence in a configuration that is thermodynamically favourable for guide–target duplex formation.

Recognition of a functional target requires the initial detection of the PAM sequence by Cas9 and DNA strand separation to enable the formation of an R-loop between the crRNA guide and target DNA[30]. The crystal structure of *S. pyogenes* Cas9 bound to sgRNA in the presence of target DNA that contains a canonical PAM implies that a multistep process exists for target recognition[22] (FIG. 3c). On PAM recognition, strand separation is induced by interactions between Cas9 and the minor groove of the PAM-containing DNA duplex, whereby a Glu and Ser residue of Cas9 interact with the DNA phosphodiester group linking the PAM sequence and the protospacer. This interaction causes a rotation of the phosphate group and induces a kink that orients the target strand directly towards the guide RNA, which is already pre-ordered in a conformation favourable for duplex formation[22]. Therefore, recognition of the PAM sequence and the A-helical arrangement of the seed sequence are coupled to conformational changes that enable Cas9 to recognize functional targets[19,20,66,67].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

A stepwise model involving a seed sequence is supported by single-molecule studies using purified Cas9 in assays with tethered DNA coupled with total internal reflection microscopy (TIRFM)[30]. In this model, Cas9 scans the target DNA via a three-dimensional collision mechanism and, on encountering a PAM sequence, interrogates the surrounding sequence. Sequence complementarity at the 3′ end of the guide sequence adjacent to the PAM is required for DNA binding, highlighting the importance of the seed sequence. Even a two-base-pair mismatch between crRNA and the DNA flanking the PAM abrogates binding, which is in accordance with structural data showing that these two seed nucleotides (nucleotides 19–20) are exposed to solvent and most likely form a nucleation site for base-pairing when strand separation has occurred during PAM recognition[17,30]. The strict requirement for complementarity immediately adjacent to the PAM suggests that unwinding of the target duplex is initiated in its vicinity and that it progresses in a stepwise manner until a single turn of an A-form helix has formed[30]. This is facilitated by recognition of the PAM, strand separation and presentation of a pre-organized A-form helical seed sequence[30]. The role of the PAM in forming the R-loop is also supported by single-molecule DNA supercoiling experiments in which PAM mutations decrease the rate of R-loop formation[29].

The evolving picture of the CRISPR mechanism is that the presentation of the crRNA seed sequence in a pre-organized A-form helix within Cas9 places it in an optimal configuration for interaction with the target on identification of a PAM sequence. The initial interrogation involves two of the seed sequence nucleotides, which are exposed to the solvent, before conformational changes within the protein facilitate interactions with the remainder of the guide.

Hfq-dependent sRNA

Almost all bacteria contain sRNAs; these participate in post-transcriptional gene regulation to regulate the majority of cellular pathways, including stress-response pathways, and to control important behavioural phenotypes, such as virulence in many pathogens[68–70]. *Escherichia coli* and *Salmonella enterica*, two enterobacterial model organisms in which bacterial RNA-mediated regulation has been extensively described, contain 200–300 sRNAs[71–74].

Bacterial sRNAs are more heterogeneous in size and structure than miRNAs and crRNAs, ranging from 50 to 200 nucleotides and containing complex secondary and tertiary structures[68]. They can be directly transcribed from independent transcription units that are located in intergenic regions and function as primary transcripts containing a 5′ triphosphate group and a 3′ stem loop (FIG. 1d). In addition, they can be processed from longer transcripts or the 3′ UTRs of mRNAs, which results in sRNAs with a 5′ monophosphate group[68,75,95]. The major class of bacterial sRNAs function through base-pairing with multiple mRNA targets via short seed sequences that are present in unstructured regions of the RNA (FIG. 1e). The majority of these *trans*-encoded sRNAs in *S. enterica* and *E. coli* bind to the Sm/Lsm superfamily RNA chaperone protein Hfq, which is required for their intracellular stability and their ability to anneal to targets, and they are thus called Hfq-dependent sRNAs[12,76,77]. Hfq-dependent sRNAs target mRNAs to affect their stability or translation. The binding of sRNAs to mRNAs near the Shine–Dalgarno sequence in their 5′ UTR regulates their access to ribosomes[68,69,78], and sRNAs negatively and positively influence mRNA stability by recruiting the major constitutive mRNA decay ribonuclease RNaseE or inhibiting its cleavage activity, respectively[79–82].

Seed sequences in sRNA. Interactions between Hfq-dependent sRNA and mRNA usually involve partial complementarity between single-stranded regions[68]. Evidence suggests that many sRNAs contain short seed sequences that disproportionately contribute to the search for, and interaction with, targets compared with other regions of the sRNA[83–87]. For example, the presence of a seed region in a sRNA was first suggested for sugar transport-related sRNA (SgrS), an ~220-nucleotide-long sRNA that is induced during glucose-phosphate stress and binds to the glucose-specific phosphotransferase *ptsG* mRNA, which encodes a membrane component of the phosphoenolpyruvate phosphotransferase system[88]. Of a 32-nucleotide region in SgrS that displays partial complementarity with the translation-initiation region of *ptsG* mRNA, only 6 nucleotides in its 3′ region are essential for its regulatory activity[83]. The importance of this six-nucleotide region is further highlighted by the ability of the SgrS seed sequence to distinguish between two horizontally acquired virulence factor mRNAs, *Salmonella* outer protein D (*sopD*) and *sopD2,* at the level of a single hydrogen bond[89]. Similarly, the ~80-nucleotide-long RybB sRNA, which is induced when the bacterial envelope is subject to stress, contains a highly conserved 5′ seed region of seven nucleotides that is responsible for the targeting of multiple outer membrane protein (*omp*) mRNAs in *S. enterica* and *E. coli*. The seed region was identified using independent genetic, molecular and biochemical approaches[84,85]. The 5′ region of RybB that contains the seed sequence is sufficient to repress *omp* mRNAs during the envelope stress response even when grafted onto an unrelated RNA scaffold, demonstrating that the seed sequence is an important element in RNA mediated silencing[84,90]. Finally, the sRNA quorum regulatory RNA 4 (Qrr4), which is involved in quorum sensing in *Vibrio* bacteria, harbours a subset of nucleotides within the region predicted to base-pair with mRNA that are important for target silencing. Intriguingly, the identity of the crucial residues differs according to the identity of the target, providing a basis for discrimination[86].

The sizes of sRNAs and the relative location of the seed sequences within them are more heterogeneous than described for miRNAs and crRNAs (FIG. 1e). In addition, some Hfq-dependent sRNAs, such as Spot42 or GcvB, which are associated with catabolite repression and are involved in amino acid transport and metabolism, respectively, use multiple or extended seed regions to recognize target mRNA[91–93]. Similarly to in miRNAs,

Total internal reflection microscopy
(TIRFM).A microscopy technique that uses an evanescent wave to specifically excite fluorophore-labelled molecules close to a surface.

Sm/Lsm superfamily
A large family of proteins present in all three domains of life and involved in RNA processing and degradation.

Shine–Dalgarno sequence
A 5–10-nucleotide sequence upstream of the initiation codon involved in defining where bacterial translation initiates.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the seed sequences in sRNAs are often highly conserved, further highlighting their functional importance[94]. An analysis of seed-sequence location in 18 well-studied *E. coli* and *S. enterica* sRNAs suggests that, in around one-third of cases, seed sequences are found in the 5′ region of the sRNA[84]. Furthermore, the seed region of a sRNA can be internally located within a primary transcript (either a longer sRNA or an mRNA) and become exposed at the 5′ end of a transcript on endonucleolytic cleavage. Examples of such processing include the ~120-nucleotide sRNA ArcZ[95], which is involved in the general stress response and is processed to an ~50-nucleotide-long sRNA[94], and CpxQ, which is liberated from the 3′ UTR of the *cpxP* mRNA (which encodes a chaperone protein) and is associated with the cellular response to inner membrane stress[96].

sRNA seed presentation. Hexameric Hfq contains at least three different binding surfaces for RNA. It typically binds sRNAs via U-rich sequences, which are often associated with RHO-independent transcription terminators, on its proximal face[21,97–99] and binds to mRNAs and some sRNAs via A-rich sequences on its distal face[100]. A third binding surface is the rim of the Hfq hexamer, which also binds U-rich sequences[21,97,101–103]. To fully understand the importance and presentation of the seed sequence by Hfq, high-resolution structural information is required. However, obtaining binary and ternary complexes containing a natural, full-length sRNA has proven challenging, probably owing to the large size and heterogeneous organization of sRNA. Nevertheless, a binary complex between Hfq and the sRNA RydC has been described, and provides the first glimpse of how a sRNA seed region is presented[21]. RydC is a 65-nucleotide sRNA that folds into a complex pseudoknot tertiary structure (FIG. 4a). RydC binds with Hfq to the 5′ UTR of the cyclopropane fatty acid (CFA) synthase mRNA, stabilizing the transcript, increasing the level of CFA synthase and thus changing the fatty acid composition of membranes[82]. The U-rich RHO-independent terminator that is located at the 3′ end of RydC binds in a recessed channel on the proximal face of the Hfq hexamer[21] (FIG. 4b). The sRNA exits the channel at the rim, where it forms extensive RNA–protein interactions. The 5′ seed region of sRNA, which spans nucleotides 2–12, binds to the rim of the Hfq hexamer, with the 3′ region of the seed (nucleotides 8–12) in an extended conformation as a result of its interaction with residues 8–17 in Hfq (FIG. 4b,c). Notably, this interaction involves Arg16 and Arg17, which were previously shown to interact with RNA and which are essential for duplex formation and for the strand exchange activity of Hfq that accelerates dynamic base-pair formation[103]. This is also consistent with chemical footprinting and small-angle X-ray scattering (SAXS) data, which show that Hfq interacts with several sites in an mRNA to fold it into a compact tertiary structure, meanwhile placing the sRNA target site at the Hfq rim in the vicinity of Arg16 and Arg17 (REF. 102). Unfortunately, the remaining 5′ portion of the RydC seed sequence is not visible in the crystal structure, potentially owing to structural heterogeneity and dynamics[21]. However, this portion of the seed could form a duplex with the target-recognition site in the mRNA while the sRNA is bound to the rim of Hfq. Indeed, it has been proposed that this interaction can be mimicked by the packing of a duplex region on the rim in the RydC–Hfq crystal lattice. As the sRNA interacts predominately with the proximal face of Hfq and mRNAs interact predominately with the distal face of Hfq, a model can be proposed in which the sRNA and mRNA are held in proximity and the sRNA–Hfq complex can scan the mRNA until the target site base-pairs with the seed and engages in the formation of a short duplex on the rim of Hfq[21,103] (FIG. 4c).

Recent global *in vivo* crosslinking datasets describing the RNA ligands that associate with Hfq provide some insight into how sRNA and mRNA bind to the chaperone[87,104,105]. In *S. enterica*, Hfq binds to more than 550 mRNAs, mainly in the 5′ UTR and 3′ UTR, and although Hfq binds throughout different regions of ~90 sRNAs that have been assessed it preferentially associates with the 3′ end of sRNA[104]. Analysis of sRNA–mRNA pairs suggests that Hfq binds to mRNA upstream of its sRNA-interaction site and that it binds on the 3′ side of the seed sequence in sRNA, consistent with the model that mRNA and sRNA bind to opposite faces of Hfq before duplex formation occurs. Although this model assumes that seed regions located at the 5′ end of sRNA are presented on the rim of Hfq (FIG. 4c), it is unclear whether sRNAs with seed regions located internally or towards the 3′ end would present them in the same manner. However, in some cases, cleavage of full-length sRNAs places the seed sequence at the 5′ end of the processed RNA, which may facilitate seed presentation[94,106].

In short, sRNAs seem to bind to the proximal face of Hfq with their seed sequence presented in an extended conformation at the rim, close to the site that has been implicated in the strand-annealing activity of the chaperone. Hfq probably stabilizes the structure of the sRNA, ensuring that the single-stranded seed sequence is available to interrogate mRNA targets that are recruited to the distal face of the chaperone.

Common principles of seed presentation

The incorporation of RNA guides into RNPs to ensure the recognition of functional targets has evolved in eukaryotic, bacterial and archaeal systems for post-transcriptional regulation and adaptive immunity[17,18,21,23–26,48]. Structural and mechanistic studies are beginning to reveal common features that are used by the different systems to impose specificity. These can be grouped together into two general principles: first, the presentation of the seed sequence in a conformation that facilitates the search for, and interaction with, target nucleic acids, and second, the coupling of target recognition and conformational changes within the RNP to ensure that the correct RNA or DNA is regulated (FIG. 5).

The guide RNA forms intimate interactions with the protein, exposing only a small subregion of the sequence to interrogate putative targets and placing the target near to the functionally active region of the protein. In miRNA, siRNA, crRNA and sRNA, the seed sequence is most important in determining target binding,

Chemical footprinting
Methods used to map RNA secondary and tertiary structures based on the accessibility of nucleotides to specific chemicals.

Small-angle X-ray scattering
(SAXS). An X-ray scattering method that provides information on the size and shape of biological molecules in solution.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

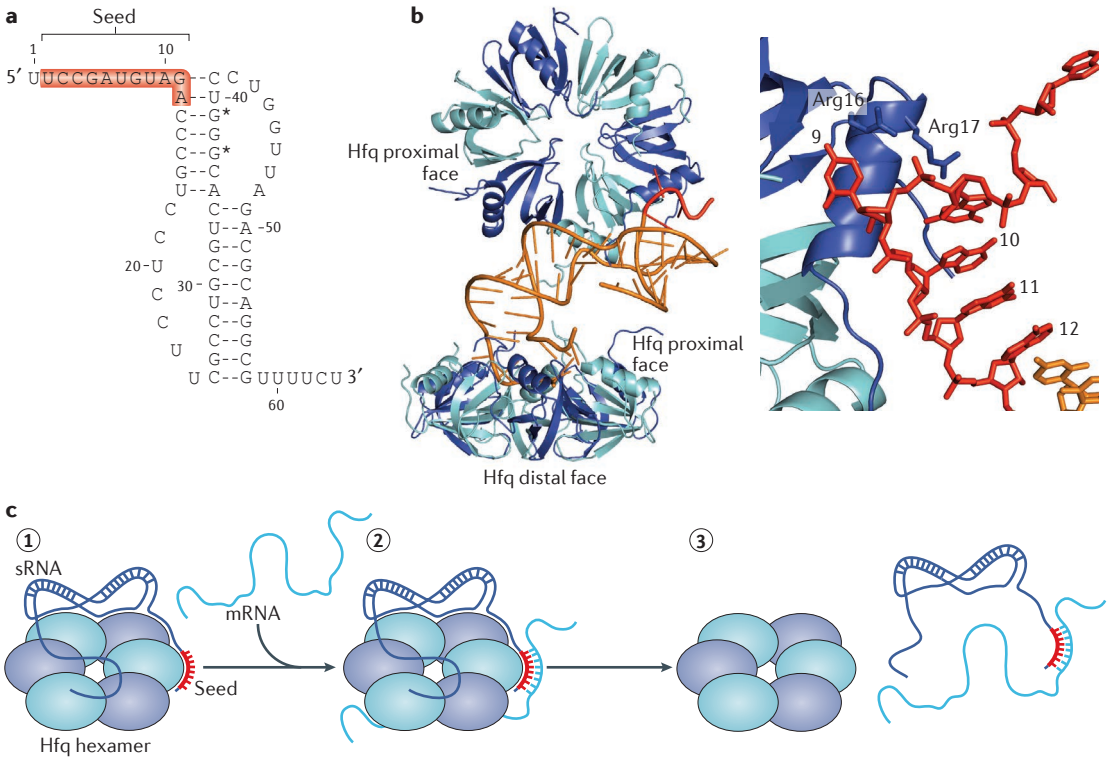


Figure 4 | **Hfq-dependent sRNA seed presentation. a** | Secondary structure of the RydC small RNA (sRNA) from *Salmonella enterica*. The seed region (highlighted in red) is located at nucleotides 2–12. **b** | Crystal structure of RydC (orange cartoon with bases depicted as sticks and the seed sequence shown in red) bound to the proximal faces of two neighbouring Hfq hexamers (protomers in cyan and blue) (left; Protein Data Bank identifier: 4v2s; see Further information)[21]. In the crystal lattice, the RNA bridges two adjacent Hfq hexamers; the structure probably encompasses the full set of interactions that would form in a 1:1 RydC–Hfq complex, which is the biologically relevant species in the cell. A close-up of the presentation of the 3′ region of the RydC seed sequence (stick representation, coloured red) in an extended conformation on the rim of the Hfq hexamer is shown on the right. **c** | Schematic depicting the targeting of mRNA by Hfq–sRNA. sRNA binds to the proximal face of the Hfq hexamer, with the 3′ region of sRNA bound in a recessed channel and nucleotides 8–12 of the 5′ seed sequence presented in an extended conformation at the rim of the hexamer, proximal to amino acids that are in the annealing activity of Hfq (step 1). mRNA binds to a distal face of the Hfq hexamer and is held by Hfq in proximity to the sRNA to enable it to scan for complementary seed sequences. A sRNA–mRNA duplex probably forms at the rim of the Hfq hexamer (step 2). On duplex formation, the sRNA and mRNA probably cycle off Hfq (step 3), which makes it available to bind to another sRNA. Part **a** is modified with permission from REF. 21.

even though its size and location within the RNA differs (FIG. 1e). For 21-nucleotide miRNA and siRNAs, nucleotides 2–8 make the largest contribution to the energy of target binding, with mismatches at the centre of the seed decreasing binding the most[8,28,43–45]. In the 20-nucleotide-long guide sequence of crRNA, 10 nucleotides in the seed region towards the 3′ end are required for target binding and cleavage; a 2-nucleotide seed subregion within this is essential for annealing to DNA next to the PAM sequence and for target recognition[30,65]. In SgrS, from a predicted 30-nucleotide region of partial complementarity, 6 nucleotides towards the 3′ end are essential for silencing[83–85].

A common principle emerging from studying miRNA-, siRNA-, crRNA- and sRNA-mediated target recognition is that the protein defines the trajectory and conformation of the guide RNA and exposes the seed sequence in a thermodynamically favourable configuration for interaction with putative targets. Presenting predominantly the seed sequence decreases the formation of non-productive intramolecular and intermolecular interactions that could arise owing to the inherently low information content of RNA bases, thereby increasing specificity.

Importantly, association of the guide RNA with a cognate protein pre-organizes the seed sequence in a conformation that favours duplex formation (FIG. 5). This is particularly evident for guide RNAs bound by AGO and Cas9, for which a subregion of the seed sequence is pre-organized in a near A-helix conformation[17,18,21,23–26,48] (FIG. 5a,b). Pre-immobilization of the guide strand by the protein decreases the loss in conformational flexibility on duplex formation[50]. This 'pre-payment' of the entropic cost associated with duplex formation enhances the affinity of the guide strand for the target when compared with that of a naked oligonucleotide[8,50]. Presentation of the seed in an A-form helix also probably decreases the rate-limiting barrier to duplex formation, which is the formation of an initial nucleus of base pairs before rapid helix propagation. In agreement with this,

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

AGO proteins increase the rate of target binding up to 250-fold compared with that of naked oligonucleotides[8,28]. This is dependent on sequence complementarity between seed nucleotides of the guide (which are presented in an A-helix conformation) and the target[8,28]. The seed sequence also plays a direct part in the mode by which protein complexes search for target sites on a nucleic acid. AGO[28] and Cas9 (REF. 30) use a lateral diffusion process and a three-dimensional collision mechanism to identify their targets, respectively. AGO uses the seed region to scan for matches to nucleotides 2–4 of the seed before converting transient interactions into more stable ones encompassing the entire seed sequence (that is, nucleotides 2–8)[28]. Cas9 scans for targets by searching for a short PAM sequence, which initiates strand separation and interrogation of the seed sequence that is also presented in an A-helix form[30].

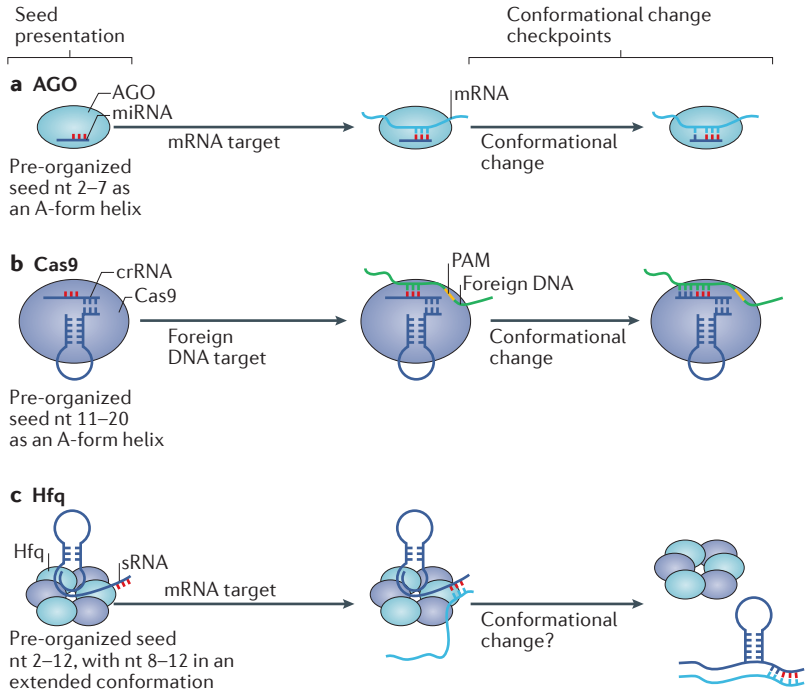


Figure 5 | **Common principles of the target-search process.** Target-search processes commonly involve presentation of the seed and conformational changes en route to target identification. Using a subregion of an RNA to initially interrogate targets produces a stepwise RNA–protein binding mechanism in which RNA–target interactions are coupled to a series of conformational changes in the ribonucleoprotein to ensure the identification of the correct target and commitment to the desired biochemical outcome. **a** | The microRNA (miRNA) seed sequence is presented in a near A-form helix conformation within the miRNA-induced silencing complex (miRISC) and is used to interrogate potential target mRNAs. On mRNA binding, conformational changes in Argonaute (AGO) expose further regions of the miRNA for interaction with the target and also check the extent of complementarity between miRNA and mRNA in a duplex. A similar principle applies to small interfering RNAs (siRNAs), which are not depicted here. **b** | The CRISPR RNA (crRNA) seed sequence is presented in a near A-form helix conformation within Cas9. Cas9 scans DNA for the protospacer-adjacent motif (PAM) sequence before undergoing conformational changes that induce strand separation and allow it to interrogate the flanking DNA sequence for complementarity to seed nucleotides (nt) 19–20; this is followed by the unwinding and binding of the remaining seed sequence. **c** | Small RNA (sRNA) bound to the proximal face of Hfq presents the seed sequence at the rim, which places it in an optimal position to interrogate a target mRNA that is bound to the distal face of Hfq and to stimulate sRNA–target mRNA duplex formation.

The presentation of a sRNA seed sequence follows similar principles; binding to Hfq constrains the 3′ end of the seed sequence in an extended single-stranded conformation that is required for target interrogation[21] (FIG. 5c). Although the conformation of the remaining seed sequence has not been resolved, it is constrained to the region of Hfq that has been implicated in stimulating target annealing[103]. In the absence of structural information on a ternary complex, the proposed model suggests that binding of the sRNA to the proximal face of Hfq and presentation of the seed at the rim places it in an optimal position to interrogate a target mRNA that is bound to the distal face of Hfq and to stimulate sRNA–target mRNA duplex formation. Similarly, RISC exhibits annealing activity that increases the efficiency of nucleic acid hybridization, and this also involves the seed sequence[44]. Single-molecule studies of SgrS bound to Hfq suggest that the target-search process is the rate-limiting step during sRNA-mediated regulation, but the role of the seed sequence remains to be investigated[107].

Presentation of a seed sequence in a conformation that facilitates binding seems to be a general principle and is not restricted to the RNA-based systems described in this Review. For example, the type-I CRISPR–Cas system in *E. coli*, which involves the multiprotein complex CRISPR-associated complex for antiviral defence (Cascade), organizes a 61-nucleotide crRNA into several consecutive 5-nucleotide-long segments. These segments are presented in an A-form conformation, which includes nucleotides 1–5 and nucleotides 7–8; both of these nucleotide sets have been experimentally shown to function as seeds during DNA targeting[108–111].

A common feature of seed sequences in these different systems is that they are in the range of 6–10 nucleotides in length. This means that they can reside within a single turn of an A-form helix, which contains ~10–11 bases; making the seed sequences longer would not be beneficial as the additional nucleotides would not be suitably orientated for base-pairing with the target. In addition, the discriminatory effect of single mismatches would be diminished in longer seed sequences, risking off-target effects, and having fewer nucleotides in a seed would compromise affinity and specificity. A second common principle of using a subregion of an RNA to interrogate targets is the linking of RNA–target interactions with a series of conformational changes in the RNP to ensure the identification of the correct target and commitment to the desired biochemical outcome (FIG. 5). For example, for miRNA, siRNA and crRNA, a match between the seed subregion and a potential target is required before subsequent interactions in the guide and target are possible. The use of a small number of nucleotides to initially scan for a functional target gives rise to a multistep binding mechanism, which provides various opportunities to check the validity of the target and thus minimise off-target effects. This suggests that RNPs may use several conformational checkpoints to further interrogate targets before placing the RNP in an active conformation. For example, the binding of *T. thermophilus* Ago to the target RNA involves nucleotides 2–6 of the seed in the guide strand; this binding induces a conformational

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

change in *T. thermophilus* Ago that relieves the kink in the guide RNA to allow base-pairing with the target RNA beyond the seed and the formation of catalytically active *T. thermophilus* Ago. Likewise, in human AGO2, nucleotides 2–5 in the seed region of miRNA initially scan putative targets; a correct match with the seed induces a conformational change in the protein that allows the rest of the seed sequence to interrogate targets. This opens up the remainder of the guide towards the 3′ end for supplementary base-pairing. Importantly, the movement of helix-7 enables the protein to sense the shape of the minor groove to ensure that a complementary duplex has been formed (FIG. 5a). In the CRISPR–Cas system, Cas9 scans DNA for the PAM sequence before inducing strand separation and thus the interrogation of the DNA sequence by seed nucleotides 19–20; this is followed by a conformational change that relieves a kink in the guide and allows binding of the remaining seed sequence to target DNA (FIG. 5b). By contrast, we currently do not know whether Hfq undergoes a conformational change to gain its RNA chaperone activity, and thus whether it is subject to the same conformational control steps as AGO proteins and Cas9 (FIG. 5c). However, a difference between the Hfq system and the siRNA and crRNA-containing RNPs is that its interaction with its RNA ligand is more dynamic, as sRNAs bind to Hfq through an active kinetic cycling process to ensure a rapid cellular response[112].

In short, it seems that using a small subregion of the seed for the interrogation of targets provides a multistep binding mechanism in which the RNP introduces multiple checkpoints to ensure the correct target is recognized.

Conclusions and future perspectives

In RNA-based silencing systems that mostly identify functional targets by seed pairing, incorporation of the guide within an RNP modifies its properties to avoid the kinetic and thermodynamic limitations of nucleic acid hybridization, ensuring it behaves more like an RNA-binding protein[8]. Interestingly, these new insights into seed presentation by RNPs highlight common principles that are involved in some RNA–RNA interactions. Of note, earlier work on codon–anticodon interactions and on antisense RNAs involved in the control of plasmid replication revealed that these RNAs interact through loop–loop interactions in the absence of proteins. These loops contain U-turn motifs, which also present flanking nucleotides in a pre-formed A-helix conformation to enhance the search for complementary sequences in their target RNAs[10,11,113,114].

A combination of structural, biochemical, biophysical and genome-wide approaches has provided extensive insights into how short RNA sequences can identify targets in the cell. This has revealed shared principles, which are especially clear for the miRNA, siRNA and crRNA systems. However, many open questions remain, including how are non-canonical RNA–target interactions specified, how do RNPs find their targets in the complex cellular environment and how does Hfq facilitate base-pairing between sRNAs and mRNAs? The use of short RNA guides has been described for several other systems, including the C-rich apical loop sequences in *Staphylococcus aureus* sRNAs[115] and also external guide sequences in RNase P-mediated regulation of gene expression[116]. It will be interesting to see whether the principles discussed in this article are relevant to these modes of regulation as well as to new RNA-based regulatory systems that are emerging from the relatively unexplored bacterial and archaeal world, for example new CRISPR–Cas proteins and global sRNA-binding proteins, such as ProQ[117,118]. Understanding how systems have evolved to identify specific functional targets has great implications for the application of RNA-based systems in biotechnology and therapeutics, such as the generation of specific RNA-based circuits for synthetic biology, the use of miRNAs or antisense inhibitors to treat various non-communicable and infectious diseases, and the improvement of tools for genome editing.

1. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution — trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
2. Levine, E. & Hwa, T. Small RNAs establish gene expression thresholds. *Curr. Opin. Microbiol.* **11**, 574–579 (2008).
3. Massé, E., Escorcia, F. E. & Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.* **17**, 2374–2383 (2003).
4. Bartel, D. P. MicroRNAs: Target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
5. Herschlag, D. RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **270**, 20871–20874 (1995).
6. Eguchi, Y., Itoh, T. & Tomizawa, J. Antisense RNA. *Annu. Rev. Biochem.* **60**, 631–652 (1991).
7. Zeiler, B. N. & Simons, R. W. in *RNA Structure and Function* Vol. 35, 437–464 (Cold Spring Harbor Laboratory Press, 1998).
8. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides. *Cell* **162**, 84–95 (2015).
9. Kunne, T., Swarts, D. C. & Brouns, S. J. Planting the seed: Target recognition of short guide RNAs. *Trends Microbiol.* **22**, 74–83 (2014).
10. Quigley, G. J. & Rich, A. Structural domains of transfer RNA molecules. *Science* **194**, 796–806 (1976).
11. Wagner, E. G., Altuvia, S. & Romby, P. Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.* **46**, 361–398 (2002).
12. Updegrove, T. B., Zhang, A. & Storz, G. Hfq: The flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).
13. Jiang, F. & Doudna, J. A. The structural biology of CRISPR–Cas systems. *Curr. Opin. Struct. Biol.* **30**, 100–111 (2015).
14. van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **12**, 479–492 (2014).
15. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014).
16. Meister, G. Argonaute proteins: Functional insights and emerging roles. *Nat. Rev. Genet.* **14**, 447–459 (2013).
17. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A. Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
Describes the crystal structure of Cas9 and reveals that the crRNA seed sequence is presented in an A-form helix configuration.
18. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science* **346**, 608–613 (2014).
19. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
20. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
21. Dimastrogiovanni, D. *et al.* Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *eLife* **3**, e05375 (2014).
Provides the crystal structure of Hfq bound to a full-length sRNA, with the seed sequence presented in an extended conformation.
22. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
23. Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037–1040 (2012).
24. Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. Structure of yeast Argonaute with guide RNA. *Nature* **486**, 368–374 (2012).
25. Elkayam, E. *et al.* The structure of human argonaute-2 in complex with miR-20a. *Cell* **150**, 100–110 (2012).
References 18, 24 and 25 report the first crystal structures of eukaryotic AGO proteins bound to RNA guides, with the seed sequence presented as an A-form helix.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

26. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754–761 (2009).
27. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
28. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A. Dynamic search process underlies microRNA targeting. *Cell* **162**, 96–107 (2015).
References 8 and 28 describe elegant single-molecule studies that provide evidence for the importance of the miRNA seed sequence in the search for targets.
29. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl Acad. Sci. USA* **111**, 9798–9803 (2014).
30. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
Reports a single-molecule study describing the stepwise interrogation of DNA targets by Cas9 and evidence for the importance of the seed sequence.
31. Fromm, B. *et al.* A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu. Rev. Genet.* **49**, 213–242 (2015).
32. Chiang, H. R. *et al.* Mammalian microRNAs: Experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24**, 992–1009 (2010).
33. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
34. Czech, B. & Hannon, G. J. Small RNA sorting: Matchmaking for Argonautes. *Nat. Rev. Genet.* **12**, 19–31 (2011).
35. Kuhn, C. D. & Joshua-Tor, L. Eukaryotic Argonautes come into focus. *Trends Biochem. Sci.* **38**, 263–271 (2013).
36. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell. Biol.* **10**, 126–139 (2009).
37. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015).
38. Lai, E. C. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**, 363–364 (2002).
39. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
40. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
41. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
42. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
References 39–42 provide the first evidence for the importance of the miRNA seed sequence in mRNA targeting.
43. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055–1067 (2012).
44. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130**, 101–112 (2007).
45. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* **11**, 599–606 (2004).
46. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
47. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
48. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008).
The first structural report describing the presentation of a guide seed sequence in an A-form helix conformation by a bacterial Ago protein.
49. Mallory, A. C. *et al.* MicroRNA control of PHABULOSA in leaf development: Importance of pairing to the microRNA 5′ region. *EMBO J.* **23**, 3356–3364 (2004).
50. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* **33**, 204–214 (2009).
51. Jo, M. H. *et al.* Human Argonaute 2 has diverse reaction pathways on target RNAs. *Mol. Cell* **59**, 117–124 (2015).
52. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* **460**, 479–486 (2009).
53. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
54. Grosswendt, S. *et al.* Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol. Cell* **54**, 1042–1054 (2014).
55. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
56. Moore, M. J. *et al.* miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.* **6**, 8864 (2015).
57. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4,** e05005 (2015).
58. Marraffini, L. A. CRISPR–Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
59. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
60. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR–Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
61. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR–Cas9. *Science* **346**, 1258096 (2014).
62. Deltcheva, E. *et al.* CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
63. Dugar, G. *et al.* High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genet.* **9**, e1003495 (2013).
64. Zhang, Y. *et al.* Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol. Cell* **50**, 488–503 (2013).
65. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
Provides key biochemical evidence for the presence of a seed sequence in a Cas9-containing CRISPR–Cas system.
66. Jiang, F. *et al.* Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
67. Nishimasu, H. *et al.* Crystal structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126 (2015).
68. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43**, 880–891 (2011).
69. Gottesman, S. & Storz, G. Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harb. Perspect. Biol.* **3,** a003798 (2011).
70. Wagner, E. G. & Romby, P. Small RNAs in bacteria and archaea: Who they are, what they do, and how they do it. *Adv. Genet.* **90**, 133–208 (2015).
71. Westermann, A. J. *et al.* Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature* **529**, 496–501 (2016).
72. Thomason, M. K. *et al.* Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.* **197**, 18–28 (2015).
73. Peer, A. & Margalit, H. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA* **20**, 994–1003 (2014).
74. Kroger, C. *et al.* An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).
75. Miyakoshi, M., Chao, Y. & Vogel, J. Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr. Opin. Microbiol.* **24**, 132–139 (2015).
76. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
77. De Lay, N., Schu, D. J. & Gottesman, S. Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J. Biol. Chem.* **288**, 7996–8003 (2013).
78. Papenfort, K. & Vanderpool, C. K. Target activation by regulatory RNAs in bacteria. *FEMS Microbiol. Rev.* **39**, 362–378 (2015).
79. Hui, M. P., Foley, P. L. & Belasco, J. G. Messenger RNA degradation in bacterial cells. *Annu. Rev. Genet.* **48**, 537–559 (2014).
80. Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C. K. & Vogel, J. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* **153**, 426–437 (2013).
81. Lalaouna, D., Simoneau-Roy, M., Lafontaine, D. & Masse, E. Regulatory RNAs and target mRNA decay in prokaryotes. *Biochim. Biophys. Acta* **1829**, 742–747 (2013).
82. Fröhlich, K. S., Papenfort, K., Fekete, A. & Vogel, J. A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J.* **32**, 2963–2979 (2013).
83. Kawamoto, H., Koide, Y., Morita, T. & Aiba, H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.* **61**, 1013–1022 (2006).
84. Papenfort, K., Bouvier, M., Mika, F., Sharma, C. M. & Vogel, J. Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA. *Proc. Natl Acad. Sci. USA* **107**, 20435–20440 (2010).
85. Balbontin, R., Fiorini, F., Figueroa-Bossi, N., Casadesus, J. & Bossi, L. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Mol. Microbiol.* **78**, 380–394 (2010).
References 83–85 describe the identification of seed sequences in bacterial sRNAs.
86. Rutherford, S. T., Valastyan, J. S., Taillefumier, T., Wingreen, N. S. & Bassler, B. L. Comprehensive analysis reveals how single nucleotides contribute to noncoding RNA function in bacterial quorum sensing. *Proc. Natl Acad. Sci. USA* **112**, E6038–E6047 (2015).
87. Melamed, S. *et al.* Global mapping of small RNA-target interactions in bacteria. *Mol. Cell* **63**, 884–897 (2016).
88. Vanderpool, C. K. & Gottesman, S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.* **54**, 1076–1089 (2004).
89. Papenfort, K., Podkaminski, D., Hinton, J. C. & Vogel, J. The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc. Natl Acad. Sci. USA* **109**, E757–E764 (2012).
90. Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H. & Vogel, J. Small RNA binding to 5′ mRNA coding region inhibits translational initiation. *Mol. Cell* **32**, 827–837 (2008).
91. Coornaert, A., Chiaruttini, C., Springer, M. & Guillier, M. Post-transcriptional control of the *Escherichia coli* PhoQ–PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet.* **9**, e1003156 (2013).
92. Sharma, C. M. *et al.* Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol. Microbiol.* **81**, 1144–1165 (2011).
93. Beisel, C. L. & Storz, G. The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol. Cell* **41**, 286–297 (2011).
94. Papenfort, K. *et al.* Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol. Microbiol.* **74**, 139–158 (2009).
95. Chao, Y. *et al. In vivo* cleavage map illuminates the central role of RNase E in coding and noncoding RNA pathways. *Mol. Cell* **65**, 39–51 (2017).
96. Chao, Y. & Vogel, J. A. 3′ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol. Cell* **61**, 352–363 (2016).
97. Sauer, E., Schmidt, S. & Weichenrieder, O. Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc. Natl Acad. Sci. USA* **109**, 9396–9401 (2012).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

98. Horstmann, N. *et al.* Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. *Nucleic Acids Res.* **40**, 11023–11035 (2012).
99. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P. & Brennan, R. G. Structures of the pleiotropic translational regulator Hfq and an Hfq–RNA complex: A bacterial Sm-like protein. *EMBO J.* **21**, 3546–3556 (2002).
100. Link, T. M., Valentin-Hansen, P. & Brennan, R. G. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc. Natl Acad. Sci. USA* **106**, 19292–19297 (2009).
101. Schu, D. J., Zhang, A., Gottesman, S. & Storz, G. Alternative Hfq–sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* **34**, 2557–2573 (2015).
102. Peng, Y., Curtis, J. E., Fang, X. & Woodson, S. A. Structural model of an mRNA in complex with the bacterial chaperone Hfq. *Proc. Natl Acad. Sci. USA* **111**, 17134–17139 (2014).
103. Panja, S., Schu, D. J. & Woodson, S. A. Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res.* **41**, 7536–7546 (2013).
104. Holmqvist, E. *et al.* Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*. *EMBO J* **35**, 991–1011 (2016).
105. Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D. & Gally, D. L. Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol. Cell* **55**, 199–213 (2014).
106. Papenfort, K., Espinosa, E., Casadesus, J. & Vogel, J. Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in *Salmonella*. *Proc. Natl Acad. Sci. USA* **112**, E4772–E4781 (2015).
107. Fei, J. *et al.* RNA biochemistry. Determination of *in vivo* target search kinetics of regulatory noncoding RNA. *Science* **347**, 1371–1374 (2015).
108. Zhao, H. *et al.* Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 (2014).
109. Mulepati, S., Heroux, A. & Bailey, S. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**, 1479–1484 (2014).
110. Jackson, R. N. *et al.* Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479 (2014).
111. Wiedenheft, B. *et al.* Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).
112. Fender, A., Elf, J., Hampel, K., Zimmermann, B. & Wagner, E. G. RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* **24**, 2621–2626 (2010).
113. Franch, T., Petersen, M., Wagner, E. G., Jacobsen, J. P. & Gerdes, K. Antisense RNA regulation in prokaryotes: Rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.* **294**, 1115–1125 (1999).
114. Brunel, C., Marquet, R., Romby, P. & Ehresmann, C. RNA loop–loop interactions as dynamic functional motifs. *Biochimie* **84**, 925–944 (2002).
115. Geissmann, T. *et al.* A search for small noncoding RNAs in *Staphylococcus aureus* reveals a conserved sequence motif for regulation. *Nucleic Acids Res.* **37**, 7239–7257 (2009).
116. Forster, A. C. & Altman, S. External guide sequences for an RNA enzyme. *Science* **249**, 783–786 (1990).
117. Attaiech, L. *et al.* Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc. Natl Acad. Sci. USA* **113**, 8813–8818 (2016).
118. Smirnov, A. *et al.* Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl Acad. Sci. USA* **113**, 11591–11596 (2016).

Acknowledgements

The authors are grateful to B. Luisi, A. Eulalio, G. Wagner and members of the authors' laboratories for discussions and comments on the manuscript. The authors also thank S. Geibel for help with the figures.

Competing interests statement

The authors declare competing interests: see Web version for details.

DATABASES

RCSB Protein Data Bank: http://www.rcsb.org/pdb/home/home.do

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.