The mysterious RAMP proteins and their roles in small RNA-based immunity

Ruiying Wang¹ and Hong Li^{1,2*}

¹Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306

²Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306

Received 7 January 2012; Revised 26 January 2012; Accepted 2 February 2012

DOI: 10.1002/pro.2044

Published online 9 February 2012 proteinscience.org

Abstract: A new class of prokaryotic RNA binding proteins called Repeat Associated Mysterious Proteins (RAMPs), has recently been identified. These proteins play key roles in a novel type immunity in which the DNA of the host organism (e.g. a prokaryote) has sequence segments corresponding to the sequences of potential viral invaders. The sequences embedded in the host DNA confer immunity by directing selective destruction of the nucleic acid of the virus using an RNA-based strategy. In this viral defense mechanism, RAMP proteins have multiple functional roles including endoribonucleotic cleavage and ribonucleoprotein particle assembly. RAMPs contain the classical RNA recognition motif (RRM), often in tandem, and a conserved glycine-rich segment (G-loop) near the carboxyl terminus. However, unlike RRM that bind single-stranded RNA using their β -sheet surface, RAMPs make use of both sides of the RRM fold and interact with both single-stranded and structured RNA. The unique spatial arrangement of the two RRM folds, facilitated by a hallmark G-loop, is crucial to formation of a composite surface for recognition of specific RNA. Evidence for RNA-dependent oligomerization is also observed in some RAMP proteins that may serve as an important strategy to increase specificity.

Keywords: RNA-mediated Immunity; RNA-binding proteins; RNA processing; RNA-protein interactions; CRISPR

Introduction

RNA molecules are traditionally known as adaptors, messengers, and subunits of the peptidyl transferase in the process of protein synthesis. However, recent

discoveries of new RNA function, and advances in genome-wide RNA sequencing, strongly suggest that RNA serves in a wide range of processes beyond protein synthesis. It is estimated that 50% to 90% of the human genome is transcribed among which only 2% to 3% is further translated into proteins.^{1,2} Although most of the noncoding RNA has not been characterized, some has been found to function directly in gene silencing,^{3,4} RNA modification,^{5,6} chromatin maintenance,⁷ protein targeting,⁸ histone modifications,⁹ stem cell differentiation,¹⁰ and epigenetic controls.¹¹ Together with the traditional roles, the new roles of RNA highlight their essentiality as well as evolutionary advantage.

Abbreviations: Cas proteins, CRISPR-associated proteins; CRISPR, clustered regularly interspaced symmetric polydromic repeat; crRNA, CRISPR RNA; RAMP, repeat-associated mysterious proteins; RRM, RNA recognition motif; RNP, ribonucleoprotein particles.

*Correspondence to: Hong Li, Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306. E-mail: hong.li@fsu.edu.

Grant sponsor: National Science Foundation; Grant number: MCB-0817638 (to H.L.).

Figure 1. Schematics of CRISPR-mediated RNA immunity pathway and involvement of Repeat-Associated Mysterious Proteins (RAMPs) in two of the three functional steps. A number of RAMPs are required for CRISPR-RNA processing and target silencing. Other Cas proteins that are known or predicted to be required for spacer incorporation, CRISPR expression and interference steps are not depicted in this model.

It is to be expected that RNA will interact with proteins along their functional pathways. Unlike DNA that is predominantly double stranded, RNA may be single stranded or of complex structural motifs.¹² As such, RNA-binding proteins are required to interact with multiple RNA motifs and they do so by an impressive number of strategies. Hallmarks of protein-RNA interactions include stacking of aromatic side chains, hydrogen bonding with peptide main chain atoms (peptide backbone atoms provide a rich source of hydrogen bond potential), selectivity of 2' hydroxyl groups and sugar puckering that are unique to RNA.^{13–17} In all cases, the conformational flexibility of RNA plays an important role in these interactions.

Most recently a new class of prokaryotic RNA binding proteins called Repeat Associated Mysterious Proteins (RAMPs), has been discovered.^{18–20} Structural studies of this new class of proteins carried out to date suggest that they use established methods of protein-RNA recognition. However, RAMPs are unique among the known RNA binding proteins in their versatility in recognizing multiple RNA motifs and in possessing enzymatic activities. This review summarizes the biological processes involving this class of RNA binding proteins and their structural features. It will also compare the protein-RNA interaction features to those learned from other RNA-binding proteins.

The CRISPR Immunity System

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) (Fig. 1) are found in many bacterial and all archaea genomes.^{21–28} They confer

acquired immunity to these organisms against invading genetic elements. CRISPR loci typically consist of identical repeats of 21 to 47 base pairs interspaced by variable spacer sequences of 20 to 72 base pairs. The overall motif also includes a set of CRISPR-associated (*cas*) genes that encode Cas proteins (Fig. 1). The spacer sequences are complementary to virus or plasmid sequences and it is through these sequences that the host “recognizes” the invading pathogen. The repeat-spacer array is transcribed and processed to ~39 to 60 nucleotide long CRISPR RNA (crRNA). The processed crRNA then partners with selected Cas proteins in order to obstruct the invasion of the virus or plasmid that have complementary sequence to the crRNA. Cas proteins play important roles in incorporation of new spacer sequences into the existing CRISPR loci, expression and maturation of crRNA and the actual process of interference.

CRISPR Immunity Processes Involving RAMP Proteins

RAMP proteins comprise a major family of Cas proteins. They play central roles in the CRISPR immunity system due to their abundant presence, and in selected cases, demonstrated enzyme activities. Each CRISPR system has at least one but may have as many as nine RAMP proteins distributed among its unique set of Cas proteins. In addition, RAMP proteins are also found outside the CRISPR loci with no apparent link to CRISPR function.¹⁸ The number of RAMP proteins is strongly correlated with the number of CRISPR units (unique inserts) as well as the

variance of insert length, supporting their potential roles in binding crRNA and related RNA.¹⁸ Early identification of RAMP proteins was based on five distinctive motifs.^{18,20} The first four of these motifs (I–IV) are selectively present in different RAMP subfamilies. Motif V contains a C-terminal glycine-rich (G-rich) sequence that is the signature motif of RAMPs and is universally present in all RAMP subfamilies. Two established CRISPR pathways described below include RAMP proteins as essential components.

The expression and maturation of crRNA (Fig. 1) is one of the key functional steps of CRISPR immunity. The repeat and spacer region of the CRISPR locus is transcribed constitutively and unidirectionally from the leader sequence to yield precursor CRISPR RNA (pre-crRNA). Processed stable RNA (crRNA), however, is short, ranging from 39 to 62 bases with 39 being the dominant form. Processing of the crRNA is carried out largely by an endonuclease that belongs to the RAMP superfamily (Fig. 1). In *Escherichia coli*, the endonuclease is part of the CRISPR-associated complex for antiviral defense (Cascade) complex that comprises five proteins and a crRNA.²⁹ Although the Cascade complex is believed to be responsible for processing pre-crRNA, the CasE (or Cse3) subunit alone contains the wild-type processing activity. The biological significance of CasE is demonstrated by the lack of CRISPR immunity in *E. coli* when CasE is absent.²⁹ Csy4 and Cas6, also RAMP proteins, were subsequently identified to be the crRNA processing endonuclease in *Pseudomonas aeruginosa*³⁰ and *Pyrococcus furiosus*,³¹ respectively. In addition, some organisms contain multiple copies of noncatalytic homologs of the processing endonuclease that bind to but do not cleave crRNA.³² The biological significance of the noncatalytic homologs of the processing endonuclease is unclear at this time. The final product of a mature crRNA comprises a spacer preceded by the last eight nucleotides of the repeat and with a heterogeneous 3' end. Structures of CasE and Cas6 in isolation have been determined and those of Csy4, CasE, Cas6, and a noncatalytic homolog of Cas6, Cas6nc, bound with RNA are now known.^{30,32–35}

The machineries responsible for CRISPR interference also include several RAMP proteins. In *Pyrococcus furiosus*, crRNA-associated ribonucleoprotein particles (RNPs) called CMR complex, were isolated and demonstrated to have crRNA-guided RNA, but not DNA, cleavage activity³⁶ (Fig. 1). The proteins associated with the RNPs were shown by mass spectrometry to be encoded by the RAMP module genes (the modularly structured genes coding for RAMPs) *cmr1*–*cmr6*. The RNA cleavage activity can also be reconstituted using recombinant Cmr proteins and synthetic crRNAs. Interestingly, four of the six Cmr proteins (Cmr1, Cmr3, Cmr4,

Cmr6) are RAMP proteins, all of which are essential to RNA silencing activity.³⁶ In *E. coli*, the Cascade complex functions with Cas3 helicase to defend against invading DNA.²⁹ The Cascade complex comprises CasA, CasB, CasC, CasD, CasE (also called Cse1, Cse2, Cse4, Cas5e, Cse3) and crRNA.^{37,38} Three of the five proteins (CasC, CasD, and CasE) are RAMP proteins³⁹ (Fig. 1). Despite the prominent presence of RAMP proteins in CRISPR interference complexes, there is not yet sufficient structural data to define their exact functional roles.

The Structure of RAMPs

The basic fold of RAMP proteins is the same as that of the well-recognized RNA recognition motif (RRM) (also known as the ferredoxin fold).⁴⁰ This fold is characterized by two sequentially connected units of $\beta\alpha\beta$ to give an overall fold of the form $\beta\alpha\beta\beta\alpha\beta$ (Fig. 2). The ferredoxin-like domain is a wide-spread folding unit found in proteins that bind nucleotides or nucleic acids.⁴⁰ In general, RAMP proteins distinguish themselves from other ferredoxin-fold proteins by juxtaposing additional secondary elements on the basic $\beta\alpha\beta\beta\alpha\beta$ unit and by arranging two such domains in tandem (Fig. 2). The N-terminal domain of all RAMP proteins tends to closely resemble the classic ferredoxin fold whereas the C-terminal domain is often interrupted by insertions, and in one case (Csy4), deviates in the conformation of the fold.³⁰ As will be discussed below, variations in the C-terminal domain are correlated with the classification of the CRISPR repeats, suggesting their roles in crRNA binding specificity.

The hallmark G-rich motif forms a loop plus a portion of the C-terminal strand of the last β -hairpin unit nestled between the two ferredoxin domains (Fig. 2). The central location of the G-loop implicates its essential role in maintaining the integrity of the protein. The RNA-bound Csy4 and Cse3 structures suggest its involvement in recognition of hairpin substrates.^{30,34,35} Mutation of three conserved glycine residues within the G-rich motif of Cas6 abolished RNA cleavage but not binding activity.³³ Thus, the G-rich loop plays a dual functional role in maintaining the dual ferredoxin fold and RNA recognition.

RNA Recognition

Structural and complementary mutational studies have revealed a surprisingly diverse range of RNA recognition methods by RAMP proteins (Fig. 3) despite their high degree of structural homology. The most striking difference among the four known RAMP proteins is the RNA structures that they recognize. Csy4 and Cse3 recognize hairpin RNA structures, whereas Cas6 and Cas6nc recognize single-stranded RNA.

Figure 2. Protein fold and topology of a representative RAMP protein (Cas6). The Cas6-family of RAMP proteins comprise a dual ferredoxin fold characterized by the $\beta\alpha\beta\beta\alpha\beta$ topology and a conserved G-loop ("x" denotes any residue and parenthesis indicates varying number of residues, "g" denotes a less conserved glycine residue). Insertions between two consecutive secondary elements (e.g. $\alpha 2$ of the N-terminal domain) are often found in various RAMP proteins but they do not alter the basic topology. An interactive view is available in the electronic version of the article.

In the Csy4-RNA complex (Fig. 3), the C-terminal domain interacts with RNA specifically while the ferredoxin-like N-terminal domain cleaves the RNA.³⁰ Its $\beta 6$ – $\beta 7$ hairpin presses against the end of the nucleic acid and the $\alpha 1$ -helix inserts into the major groove of the RNA hairpin. Five arginine residues (Arg111, Arg114, Arg115, Arg118, and Arg119) are located on the RNA-binding side of the $\alpha 1$ -helix and provide stabilizing electrostatic interactions. These arginine residues also allow the protein to recognize the nucleobases of the hairpin. Structural and mutational studies show that the two base pairs upstream of the cleavage site play important roles

for substrate recognition.³⁰ The 3' end of the hairpin is the site of cleavage and it is placed between a conserved histidine (His29) and a serine (Ser148), both of which are required for cleavage. Despite the fact that the N-terminal ferredoxin fold does not directly participate in RNA recognition, it anchors the C-terminal domain and provides critical catalytic residues.

Similarly, the Cse3-type of processing endonucleases recognize and process hairpin RNA substrates.^{34,35} As for Csy4, the RNA substrate is bound to the opposite side of the β -sheet of the ferredoxin-fold and is positioned similarly with respect to the

Figure 3. Overview of the currently known RAMP-RNA complex structures. The four RAMP proteins are aligned by their N-terminal ferredoxin fold (teal) and their C-terminal domains (all have a ferredoxin fold except for Csy4) are colored light grey. The bound RNA molecules are colored red. The second complex of the Cas6nc-RNA homodimer is displayed in stick model for the protein and orange color for the RNA. Note that Cas6 and Cas6nc (Cas6 noncatalytic homolog) recognize single-stranded RNA while Csy4 and Cse3 bind hairpin RNA substrates. The schematic interaction of each protein-RNA complex is depicted directly below. The small scissor represents the active site for those that cleave RNA. An interactive view is available in the electronic version of the article.

Figure 4. Comparison of RNA recognition by RAMP (left) and RNA-recognition motif (RRM) (right). Two representative complexes from each class of proteins are displayed. The ferredoxin folds of all four proteins are aligned and colored in teal. RNA is colored in red. The PDB code for each complex is included in parentheses. Note that RAMP proteins use both sides of the ferredoxin folds to recognize RNA while RRM proteins use only their β -sheet face to interact with single stranded RNA.

N-terminal ferredoxin fold (Fig. 3). Cse3 also interacts with the major groove side of the RNA hairpin, although using a β -hairpin rather than an α -helix unit. Positively charged residues are again observed to stabilize the RNA phosphate backbone. Unlike Csy4, however, Cse3 requires the unpaired nucleotides immediately flanking the hairpin unit (nucleotide 5 and 22–24) for efficient binding and cleavage. Base-specific contacts are observed between highly conserved Cse3 residues and these four nucleotides.

The recognition of RNA by Cas6 and Cas6nc is drastically different from that for Csy4 and Cse3. Like Cse3, Cas6, and Cas6nc maintain a tandem ferredoxin fold in the presence and absence of bound RNA.^{32,33} But unlike Csy4 and Cse3, the RNA substrate for Cas6(nc) is unstructured and binds as a single strand to the β -sheet side of the ferredoxin folds (Fig. 3). The phosphate backbone of the RNA is stabilized by positively charged residues. The bases are inserted into the central groove for specific recognition. Mutational analysis indicates that only a small region of the RNA, residues 3 to 8, is recognized by Cas6 while the cleavage takes place after residue 22.³³ The fact that as many as five nucleotides can be deleted between nucleotides 12 and 18 without affecting RNA binding and cleavage further

establishes that no secondary structure of the RNA is required for binding or cleavage.³³ These analyses lead to a wrap-around model in which the 5' end of the RNA is anchored on the β -sheet side while the 3' end is cleaved by residues located on the opposite side of the β -sheet (Fig. 3).

Evidence for RNA-Dependent Oligomerization of RAMPs

Structures of Cas6nc bound with a repeat RNA and its variants reveal plasticity of RNA recognition and suggest a model of RNA-dependent oligomerization of RAMP proteins.³² Cas6nc shares 27% sequence identity with Cas6 and binds RNA similarly to Cas6. It also interacts with single stranded RNA and anchors the 5' end of the RNA by the central groove formed by the two protein domains. However, Cas6nc differs in oligomerization from Cas6 (Fig. 3). In the Cas6nc-RNA complex, a single RNA binds across two Cas6nc proteins and vice versa, leading to a 2:2 homodimer (Fig. 3). The dimerization is RNA sequence sensitive because mutation of the fifth to seventh nucleotides led to 1:1 protein-RNA complexes similar to Cas6-RNA complex.³² These observations suggest that some RAMP proteins have an intrinsic ability to form oligomers with certain repeat RNAs. Additional *in vivo* work is required to establish the physiological significance of the RNA-dependent dimerization.

Relationship to the RNA Recognition Motif (RRM) Proteins

The other prominent family of RNA binding proteins that contain the ferredoxin fold is the RNA Recognition Motif, or RRM family.⁴⁰ RRM proteins bind RNA very differently than RAMP proteins (Fig. 4). First, most RRMs are capable of binding RNA with a single ferredoxin fold motif while RAMP proteins require the collaboration of both an N-terminal ferredoxin fold and a C-terminal domain that, in many cases, is also a ferredoxin fold. Second, RRMs use the protruding aromatic side chains of the β -sheet to interact with single stranded RNA (Fig. 4). Third, RRMs only interact with single-stranded RNA while RAMP proteins interact with both nonstructured and structured RNA. Another interesting distinction between RRMs and RAMPs is the fact that RAMPs are mostly prokaryotic while RRMs are predominantly eukaryotic. Therefore, it is possible that both RRMs and RAMPs are diversified from the common ancestral ferredoxin fold.

The recent discovery that a dual RRM protein, U2AF65, reorganizes its two RRM modules upon binding RNA^{41,42} provides a conceptual link between RRM and RAMP proteins in RNA-induced conformational change (Fig. 5). Two RRM motifs of U2AF65 are closely associated in the absence of RNA but are predominantly extended in the presence of RNA.⁴¹

Figure 5. Schematic representation of the RNA-induced changes in both RAMP and RRM proteins. RNA mediates formation of a dimer of Cas6nc (top) and shifts the equilibrium of conformations of a tandem RRM protein (bottom).

Similar to the change in oligomeric state observed for Cas6nc, binding of single-stranded RNA stabilizes protein conformations that are otherwise unpopulated. Changes in the conformation or state of a protein when a ligand is bound may provide a means of regulation.

Conclusion

RAMP proteins comprise a new class of RNA binding proteins found mostly in bacteria and archaea. RAMP proteins are functional components of the recently discovered CRISPR pathway that confers immunity to the prokaryotic host against invading nucleic acids. The two currently known functions associated with RAMP proteins are processing of CRISPR loci encoded crRNA and silencing of invading RNA molecules (Fig. 1). Additional functions of RAMP proteins are yet to be identified.

Current structural data on RAMP proteins are limited to members of a subfamily that are responsible for processing pre-crRNA. Findings from structural studies of these RAMP protein-RNA complexes already reveal a great diversity in RAMP-RNA interactions and, therefore, present a challenge to understand the recognition mechanism critical to crRNA processing and target silencing. There are 12 distinct families of repeat RNAs based on their sequences and structures.⁴³ We anticipate more diverse modes of RNA interactions by RAMP proteins as structural studies of this fascinating class of RNA binding proteins continue. What other types of RNA structures do RAMP proteins recognize? Do RAMP proteins interact with other proteins and if so, how is this protein-interaction mode related to that of RNA binding?

Acknowledgments

The authors are grateful for insightful discussion with Drs. M. P. Terns and R. M. Terns.

References

- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G,

- Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308: 1149–1154.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korb J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Masingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Srinivasan M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyrales E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstein GM, Gibbs RA, Graves T, Fulton R, Mardis ER,

- Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.
3. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
4. Hannon GJ (2002) RNA interference. *Nature* 418:244–251.
5. Potter CS, Chu H, Frey B, Green C, Kisseberth N, Madden TJ, Miller KL, Nahrstedt K, Pulokas J, Reilein A, Tchong D, Weber D, Carragher B (1999) Leginon: a system for fully automated acquisition of 1000 electron micrographs a day. *Ultramicroscopy* 77:153–161.
6. Decatur WA, Fournier MJ (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem* 278:695–698.
7. Mitchell JR, Wood E, Collins K (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402:551–555.
8. Egea PF, Stroud RM, Walter P (2005) Targeting proteins to membranes: structure of the signal recognition particle. *Curr Opin Struct Biol* 15:213–220.
9. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long non-coding RNA as modular scaffold of histone modification complexes. *Science* 329:689–693.
10. Bertani S, Sauer S, Bolotin E, Sauer F (2011) The non-coding RNA *Misra* activates *Hoxa6* and *Hoxa7* expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell* 43:1040–1046.
11. Spitale RC, Tsai MC, Chang HY (2011) RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6:539–543.
12. Moore PB (1999) Structural motifs in RNA. *Annu Rev Biochem* 68:287–300.
13. Draper DE (1999) Themes in RNA-protein recognition. *J Mol Biol* 293:255–270.
14. Chen Y, Varani G (2005) Protein families and RNA recognition. *FEBS J* 272:2088–2097.
15. Cusack S (1999) RNA-protein complexes. *Curr Opin Struct Biol* 9:66–73.
16. Perez-Canadillas JM, Varani G (2001) Recent advances in RNA-protein recognition. *Curr Opin Struct Biol* 11:53–58.
17. De Guzman RN, Turner RB, Summers MF (1998) Protein-RNA recognition. *Biopolymers* 48:181–195.
18. Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* 30:482–496.
19. Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 1:e60.
20. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* 1:7.
21. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151:2551–2561.
22. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60:174–182.
23. Sorek R, Kunin V, Hugenholtz P (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6:181–186.
24. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* 34:401–407.
25. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11:181–190.
26. Karginov FV, Hannon GJ (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* 37:7–19.
27. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–170.
28. Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136:615–628.
29. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuys RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–964.
30. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329:1355–1358.
31. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP (2010) Binding and cleavage of CRISPR RNA by Cas6. *RNA* 16:2181–2188.
32. Wang R, Zheng H, Preamplume G, Shao Y, Li H (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. *Protein Sci* 21:405–417.
33. Wang R, Preamplume G, Terns MP, Terns RM, Li H (2011) Interaction of the Cas6 Riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* 19:257–264.
34. Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* 18:688–692.
35. Sashital DG, Jinek M, Doudna JA (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol* 18:680–687.
36. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139:945–956.
37. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* 18:529–536.
38. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJ, van der Oost J, Doudna JA, Nogales E (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477:486–489.
39. Makarova KS, Aravind L, Wolf YI, Koonin EV (2011) Unification of Cas protein families and a simple

- scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* 6:38.
40. Maris C, Dominguez C, Allain FH (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272: 2118–2131.
  41. Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcarcel J, Sattler M (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475:408–411.
  42. Gupta A, Jenkins JL, Kielkopf CL (2011) RNA induces conformational changes in the SF1/U2AF65 splicing factor complex. *J Mol Biol* 405:1128–1138.
  43. Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 8:R61.