Chapter 5
Circular RNAs Are a Large Class of Animal RNAs with Regulatory Potency

5.1 The circRNA CDR1as As a Potential Sponge for miR-7

While screening argonaute (AGO) PAR-CLIP data produced by the Tuschl lab [9] and the Landthaler lab [17] for unusual and interesting sites of miRNA binding, Anna Elefsinioti discovered a very densely bound region (Fig. 5.1a) antisense to the annotated gene Cerebellar Degeneration Related 1 (CDR1) which was first described as an auto-antigen in patients with cerebellar degeneration [7]. Interestingly, [10] had previously described an antisense transcript from this locus to undergo circularization, presumably in an unusual splicing reaction (in what follows, the circular antisense transcript will be referred to as CDR1as). In addition to a highly complementary miRNA binding site for miR-671 described by [10], we found that CDR1as harbors 74 seed complementary sites for miR-7. Upon inspection of orthologous sequences from other vertebrates (Fig. 5.2a), spanning the range from human to lamprey (Table 5.1), we found these seed sites to be specifically conserved during evolution, while interspaced sequences were more variable (Fig. 5.2b). Furthermore, base pairing to mature miR-7 was predicted to drop sharply at positions 10, 11 which are pivotal for the catalytic slicing of AGO2, required for siRNA like silencing [1, 6, 26], and increased again towards the 3′ end of the miRNA (Fig. 5.2c). This profile is consistent with strong binding by miR-7 without inducing enzymatic cleavage by the miR-7 RISC complex, and indicates evolutionary optimization for a sponge-like function of CDR1as. In addition, no other animal miRNA apart from miR-7 is capable of explaining the broad binding of AGO along the entire CDR1as sequence, observed in the PAR-CLIP experiments.

We studied the cellular localization of CDR1as with single molecule fluorescent RNA hybridization and found CDR1as to be abundantly present in the cytoplasm (Fig. 5.1b), consistent with a function as miR-7 sponge. Confirming [10] we found CDR1as to exist almost exclusively as a circular molecule. Northern blotting did not
Fig. 5.1 The circRNA CDR1as is bound by the miRNA effector protein AGO, and is cytoplasmic. 
a CDR1as is densely bound by AGO (red) but not by unrelated proteins (black). Blue boxes indicate miR-7 seed matches. nt, nucleotides. 
b CDR1as RNA is cytoplasmic and disperse (white spots; single-molecule RNA FISH; maximum intensity merges of Z-stacks). siSCR, positive; siRNA1, negative control. Blue, nuclei (DAPI); scale bar, 5 mm (see also Supplementary Fig. 10 of [17] for uncropped images). 
c Northern blotting detects circular but not linear CDR1as in HEK293 RNA. Total, HEK293 RNA; circular, head-to-tail probe; circ+lin, probe within splice sites; IVT lin., in vitro transcribed, linear CDR1as RNA. 
d Circular CDR1as is highly expressed (qPCR, error bars indicate standard deviation). e CDR1as as a circular miR-7 sponge. Blue, seed matches; dark red, AGO PAR-CLIP reads; bright red, crosslinked nucleotide conversions. Adapted from [17]

detect additional linear species of the molecule (Fig. 5.1c) and quantitative PCR with head-to-tail splicing specific, divergent primers reported circular CDR1as expression at around 10–20 % of the highly expressed house keeping gene GAPDH (Fig. 5.1d). The data were suggesting that CDR1as, as an abundant and cytoplasmic, circular RNA molecule, could densely bind to miR-7 RISC and consequently represent a potent endogenous sponge for miR-7 (Fig. 5.1e).
5.2 CDR1as and miR-7 Are Co-expressed in Neuronal Tissue

In order to interact with miR-7, CDR1as needs to be co-expressed in the same tissue. In mice, miR-7 is known to be highly expressed in neuro secretory cells, pituitary gland and the pancreas [16]. We measured the expression of mature miR-7 and circular CDR1as (with divergent, head-to-tail specific primers) in various mouse tissues and indeed found miR-7 and CDR1as co-expressed to high levels in neuronal tissues, especially the cerebellum and mid brain [14, 20] (Fig. 5.3a). Closer investigation with
Table 5.1  Abundant miR-7 seed matches in vertebrate CDR1as sequences

<table>
<thead>
<tr>
<th>Species</th>
<th>Length (nt)</th>
<th>Total</th>
<th>6mer</th>
<th>7mer</th>
<th>8mer</th>
<th>9mer</th>
<th>10mer</th>
<th>11mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>1,489</td>
<td>74</td>
<td>1</td>
<td>65</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhesus macaque</td>
<td>1,542</td>
<td>76</td>
<td>1</td>
<td>68</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>2,931</td>
<td>139</td>
<td>9</td>
<td>49</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>811</td>
<td>36</td>
<td>1</td>
<td>31</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naked mole-rat</td>
<td>1,105</td>
<td>52</td>
<td>1</td>
<td>45</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cow</td>
<td>873</td>
<td>34</td>
<td>1</td>
<td>31</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horse</td>
<td>1,269</td>
<td>61</td>
<td>1</td>
<td>54</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>1,316</td>
<td>65</td>
<td>1</td>
<td>56</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elephant</td>
<td>1,333</td>
<td>54</td>
<td>1</td>
<td>50</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lizard</td>
<td>1,068</td>
<td>17</td>
<td>3</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamprey</td>
<td>1,177</td>
<td>38</td>
<td>3</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Orthologous sequences to human CDR1as were obtained from the UCSC genome browser for various vertebrate species (methods 7.8). In each sequence, miR-7 seed matches were very abundant and the majority of matches provided at least 7 nucleotides of complementarity to miR-7, indicating strong binding. The rat genome rn4 contained gaps in the assembly, which were removed, but the number of ‘N’s indicates the length may be similar to the mouse Cdr1as sequence. Zero entries have been omitted.

Fig. 5.3  CDR1as and miR-7 have overlapping and specific expression in neuronal tissues. a Among mouse tissues and MIN6 cells (qPCR, relative to cerebral cortex expression; error bars indicate standard deviations) neuronal tissues co-express miR-7 and CDR1as. b In situ staining of CDR1as and miR-7 in mouse embryo brain E13.5 (U6 and miR-124, positive control; scrambled probe, negative control). Scale bar, 1 mm. From [17]

in situ hybridization revealed specific and spatially overlapping expression patterns of miR-7 and CDR1as in the mid brain of developing mice (E13.5) (Fig. 5.3b), strongly suggesting that CDR1as functionally interacts in vivo with miR-7 in the
CDR1as and miR-7 Are Co-expressed in Neuronal Tissue

5.2 CDR1as and miR-7 Are Co-expressed in Neuronal Tissue

brain. Consistent with this hypothesis, HEK293 cells in which we found CDR1as highly expressed (Fig. 5.1f) are of neuronal origin [21], and miR-7 targets act in neuronal function and synapse formation [4, 12].

To study the functionality of CDR1as interacting with miR-7, a knock out of CDR1as would be the experiment of choice, but this would inadvertently lead to the loss of the CDR1 protein, encoded on the other genomic strand. As CDR1 is not well characterized, the interpretation of such a loss of function experiment would be very difficult.

5.3 CDR1as Expression in Zebrafish Phenocopies miR-7 Inhibition

The zebrafish Danio rerio has, according to careful bioinformatic analysis of the genome (including the trace archives), lost the entire CDR1 locus, while miR-7 is present and expressed in the embryonic brain [13]. We decided to study the miR-7 phenotype in zebrafish by morpholino inhibition and then compare it to the effect of expressing CDR1as sequence in the fish. We reasoned that such a gain of function experiment would allow to disentangle the function of CDR1as non-coding RNA from the CDR1 protein. If indeed, CDR1as sequence acted as a sponge for miR-7 in the zebrafish brain, we would expect a similar phenotype for either inhibition of miR-7 by morpholino, or expression of CDR1as sequence.

The loss of miR-7 function by morpholino inhibition (miR-7 MO) had a pronounced and reproducible effect on the developing zebrafish brain, we observed a significant reduction of the volume of the mid brain, while other brain structures such as the telencephalon remained largely unaffected. Importantly, at the intermediate dose of 9 ng miR-7 MO, the overall morphology of the animal was intact (Fig. 5.4a, b). The reduced mid brain phenotype was further characterized by 3D volumetric reconstruction for select animals. A representative image is shown in Fig. 5.4c. Next we tested the effect of CDR1as expression in the fish, by injecting DNA plasmids that express either an artificial linear, or circular CDR1as RNA (plasmid courtesy of the Kjems lab), albeit at lower efficiency than the native locus in human cells. We observed reduced mid brain volumes, especially for the circular CDR1as expressing plasmids, that could be partially rescued by concomitant injection of miR-7 precursor RNA (Fig. 5.4d–f). The experiments were performed on several hundreds of animals (including another genetic background) and the results were reproducible and highly significant (Fig. 5.4g, h). From the similar phenotype of CDR1as expression and miR-7 inhibition by morpholinos, we conclude that CDR1as can act as a potent sponge for miR-7 in the developing zebrafish brain. However, the rescue with miR-7 precursor over expression was only partial, hinting at the possibility of additional functions of CDR1as.
Fig. 5.4 In zebrafish, knockdown of miR-7 or expression of CDR1as causes mid brain defects. a, b Neuronal reporter [(Tg(huC:egfp))] embryos (top, light microscopy) 48 h post fertilization [bottom, representative confocal z-stack projections; blue dashed line, telencephalon (TC) (control); yellow dashed line, mid brain (MB)]. Embryos after injection of 9 ng miR-7 morpholino (MO) (b) display a reduction in mid brain size. Panel a shows a representative embryo injected with 15 ng control morpholino. c Three-dimensional volumetric reconstructions. d Empty vector control. e Expression vector encoding human circular CDR1as. f Rescue experiment with miR-7 precursor. g Phenotype penetrance (% of embryos, miR-7 MO, n = 135; uninjected, n = 83; empty vector, n = 91; linear CDR1as, n = 258; circular CDR1as, n = 153; circular CDR1as plus miR-7 precursor, n = 217). Phenotype distribution derived from at least three independent experiments. Scale bar: 0.1 mm. ** P < 0.01; *** P < 0.001 in Student’s t-test for normal mid brain, reduced mid brain. h Phenotype quantification (Methods). Error bars indicate standard deviation n = 3 per group. From [17]
Identifying Animal Circular RNAs from Deep Sequencing Data

Intrigued by the circularity of the naturally occurring CDR1as RNA, we asked how many other circular RNAs exist in animals, and what are the properties of circular RNAs (circRNAs) that distinguish them from linear RNA? Previous work had already identified naturally occurring circRNAs in human leukocytes [19], archaea [5], and the Sry transcript expressed in mouse testis [3]. In prokaryotes, circular RNAs were known to arise from ribosomal RNA precursors [8]. However, no biological function had been assigned to any of the previously described circRNAs (see also Sect. 1.5.4).

We reasoned that head-to-tail splicing could be detected in RNA-seq data by specifically searching for spliced reads which span the participating exons in reversed order (Fig. 5.5a). Spliced reads, arising from cDNA fragments that span normal, linear exon-exon junctions, have been used before to reconstruct exon-intron structures from deep sequencing data [23]. We chose a similar approach, splitting putatively spliced reads, that do not align contiguously to the genome, into three parts. The outermost 20 nt from each candidate read were aligned individually to the genome in order to find exonic anchor positions on both sides of the junction. Subsequently, the remaining, internal part of the read was used to extend the anchor alignments until (a) the complete read aligned, (b) the breakpoint was unambiguously determined by the sequence, (c) the inferred exon boundaries were flanked by canonical GU/AG splice signal di-nucleotides. This procedure was able to detect normal, linear and head-to-tail splicing with a low false discovery rate <1 % (see [17] Supplementary Information and Full Methods for details on permutation tests and synthetic data).

Standard mRNA sequencing employs biochemical enrichment for poly-adenylated RNAs to deplete the vast amount of ribosomal and transfer RNA from the cellular extract to ensure efficient profiling of mRNA expression. However, circRNAs do not have ends and consequently lack poly-A tails. We therefore ran our algorithm on RNA-seq data produced after depletion of ribosomal RNA (termed “ribominus” or “ribozero”), a method that retains non-polyadenylated RNA. We investigated published RNA-seq data from human leukocytes [19], mouse head, brain, and differentiation induced stem cells [11, 25], as well as our own data for human HEK293 cells and the early stages of C. elegans embryonic development (by Marlon Stöckius, unpublished). At careful quality cutoffs (at least two independent, distinct and uniquely aligning head-to-tail spanning reads with unambiguous identification of the splice site from the sequence) we discovered 1,950 human circRNAs, 1,903 mouse circRNAs, and 724 nematode circRNAs (Fig. 5.5b). These were oftentimes detected in specific cell types or tissues, suggesting that their expression is regulated during differentiation and development. After annotation of human circRNAs against the RefSeq database and a catalogue of non-coding RNAs [2, 15, 18, 24], we found 85 % of human circRNAs to align sense to known genes. Their splice sites typically
Circular RNAs Are a Large Class of Animal

Fig. 5.5 Detection, classification and evolutionary conservation of circRNAs.  
(a) The termini of junction-spanning reads (anchors) align sequentially to the genome for linear (top) but in reversed orientation for head-to-tail spliced reads (bottom). Spliced reads must distribute completely to anchors, flanked by AG/GU (Methods).  
(b, c) circRNAs in human cell types (b) and nematode stages (c).  
(d) Genomic origin of human circRNAs. A total of 96% of circRNAs overlap known transcripts.  
(e) Examples of human circRNAs. The AFF1 intron is spliced out. Sequence conservation: placental mammals phyloP score (methods), scale bar, 200 nucleotides.  
(f) A total of 223 human coding sequence circRNAs with mouse orthologs (green) and controls (black) with matched conservation level. Insert: mean conservation for each codon position (gray), controls (black); x axis, codon positions; y axis, placental mammals phyloP score; Third codon positions are significantly more conserved ($P < 4 \times 10^{-10}$, Mann-Whitney U-test, n = 223). From [17].
Table 5.2 Most experimentally tested circRNAs are validated

<table>
<thead>
<tr>
<th>Sample</th>
<th>Validation experiment</th>
<th>Validation success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human (HEK293)</td>
<td>Head-to-tail splicing</td>
<td>19 of 23</td>
</tr>
<tr>
<td></td>
<td>Circulary</td>
<td>21 of 21</td>
</tr>
<tr>
<td></td>
<td>Expression &gt; 3 % vinculin</td>
<td>12 of 21</td>
</tr>
<tr>
<td></td>
<td>Expression specificity (leukocyte specific)</td>
<td>5 of 7</td>
</tr>
<tr>
<td>Mouse (adult brain)</td>
<td>Head-to-tail splicing</td>
<td>3 of 3</td>
</tr>
<tr>
<td></td>
<td>Circulary</td>
<td>3 of 3</td>
</tr>
<tr>
<td></td>
<td>Expression &gt; 1 % β-actin</td>
<td>2 of 3</td>
</tr>
<tr>
<td><em>C. elegans</em></td>
<td>Head-to-tail splicing</td>
<td>15 of 20</td>
</tr>
<tr>
<td></td>
<td>Circulary</td>
<td>13 of 13</td>
</tr>
<tr>
<td></td>
<td>Expression &gt; 1 % eif3.d</td>
<td>12 of 15</td>
</tr>
</tbody>
</table>

In total, 50 circRNA candidates in three different organisms were tested. Assays included qPCR for head-to-tail splicing and expression level, Sanger sequencing for exact splice site, RNase-R treatment for circularity. Table adapted from [17]

span one to five exons and overlap coding exons (84 %), but only in 65 % of these cases are both splice sites that participate in the circularization known splice sites (see [17], Supplementary Table 5.2). The latter point stresses the advantage of our detection strategy, which in contrast to [19] does not rely on existing exon annotations. Figure 5.5e shows some examples of human circRNAs, demonstrating the variety of genomic origins of these molecules.

The finding that many circRNAs arise from head-to-tail splicing of known, coding exons raised the question of whether there could be additional functions carried out by the corresponding RNA sequences. As coding sequences overall are typically well conserved due to the strong selective pressure to keep the open reading frame intact, we analyzed the third codon position in particular. The base pairing of the third position in a codon to the tRNA anti-codon loop is less stable, and most synonymous codons differ in the third position. Consequently, third codon positions in open reading frames are much less constrained than the first and second positions. We reasoned that any function embedded in coding sequence, in addition to the encoded amino acid sequence, should manifest itself in an elevated conservation of the third codon position. Indeed, when comparing 223 circRNAs with conserved expression in mouse to carefully chosen control exons (see [17], Supplementary Fig. 5.1j, k and Full Methods) we found the third codon positions inside the circularized exons to be on average significantly more conserved (Fig. 5.5f), while first and second positions matched the chosen control exons (Fig. 5.5f, insert). The analysis thus indicates that at least a subset of circRNAs derived from coding exons may exert a biological function in addition to the known function of encoding a protein in the context of a linear mRNA.
Fig. 5.6  CircRNAs are stable transcripts with robust expression. a Human (hsa) ZRANB1 circRNA exemplifies the validation strategy. Convergent (divergent) primers detect total (circular) RNAs. Sanger sequencing confirms head-to-tail splicing. b Divergent primers amplify circRNAs in cDNA but not genomic DNA (gDNA). GAPDH, linear control, size marker in base pairs. c Northern blots of mock (2) and RNase-R (1) treated HEK293 total RNA with head-to-tail specific probes for circRNAs. GAPDH, linear control. d, e circRNAs are at least 10-fold more RNase-R resistant than GAPDH mRNA (d) and stable after 24 h transcription block (e) (qPCR; error bars indicate standard deviation). From [17]

5.5 Validation of circRNA Candidates

With the large number of putative circRNAs, derived from sequencing data, we wanted to confirm the existence of the corresponding circular RNAs in vivo. We selected in total ∼50 circRNA candidates and subjected them to validation experiments. First, we designed divergent primers for each predicted head-to-tail junction and tested for a PCR amplicon of the correct size. Amplicons were Sanger sequenced to confirm the reversed order of exonic sequences and the exact
breakpoint, they could not test whether the molecule is circular or linear. In particular, the head-to-tail splicing signature could also be explained by trans-splicing, genomic rearrangements reversing exon order, or by PCR artifacts.

However, true circularity should render an RNA highly resistant to exonuclease digestion [22]. Indeed, Northern blotting (without any PCR), confirmed the resistance of CDR1as, hsa-circRNAs 2, 3 and 16 to RNase-R treatment, while the highly abundant linear mRNA of GAPDH was efficiently degraded (Fig. 5.6c). We quantified the RNase-R resistance of circRNAs by qPCR and consistently found them to be at least 10-fold more resistant than GAPDH (Fig. 5.6d). As normal, cellular mRNA turn-over mechanisms rely on exonuclease activity, we hypothesized that circRNAs had the potential for high stability under cellular conditions. Indeed, 24 h after transcription block with Actinomycin D, most circRNAs levels were largely unaffected, comparable to, or exceeding the stability of GAPDH mRNA (Fig. 5.6e). Table 5.2 lists the results of all validation experiments and shows that the clear majority of tested circRNA candidates could be validated.

References

complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. PLoS One 6(11), e27288 (2011)


学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。
图书馆致力于便利、促进学习与科研，提供最强文献下载服务。