

Expression determinants of mammalian argonaute proteins in mediating gene silencing

Paul N. Valdmanis^{1,2}, Shuo Gu^{1,2}, Nina Schürmann³, Praveen Sethupathy⁴, Dirk Grimm³ and Mark A. Kay^{1,2,*}

¹Department of Pediatrics, ²Department of Genetics, Stanford University, 269 Campus Drive CCSR 2110, Stanford, CA, USA, 94305, ³Cluster of Excellence CellNetworks, Department of Infectious Diseases/Virology, University of Heidelberg, Heidelberg, Germany and ⁴Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Received August 6, 2011; Accepted December 8, 2011

ABSTRACT

RNA interference occurs by two main processes: mRNA site-specific cleavage and non-cleavage-based mRNA degradation or translational repression. Site-specific cleavage is carried out by argonaute-2 (Ago2), while all four mammalian argonaute proteins (Ago1–Ago4) can carry out non-cleavage-mediated inhibition, suggesting that Ago1, Ago3 and Ago4 may have similar but potentially redundant functions. It has been observed that in mammalian tissues, expression of Ago3 and Ago4 is dramatically lower compared with Ago1; however, an optimization of the Ago3 and Ago4 coding sequences to include only the most common codon at each amino acid position was able to augment the expression of Ago3 and Ago4 to levels comparable to that of Ago1 and Ago2. Thus, we examined whether particular sequence features exist in the coding region of Ago3 and Ago4 that may prevent a high level of expression. Swapping specific sub-regions of wild-type and optimized Ago sequence identified the portion of the coding region (nucleotides 1–1163 for Ago-3 and 1–1494 for Ago-4) that is most influential for expression. This finding has implications for the evolutionary conservation of Ago proteins in the mammalian lineage and the biological role that potentially redundant Ago proteins may have.

INTRODUCTION

Argonaute (Ago) proteins are essential for one of the last steps in the microRNA (miRNA) biogenesis pathway—the recognition and silencing of target transcripts. Canonical miRNA biogenesis typically commences with RNA polymerase II-mediated transcription of primary

microRNA transcripts (pri-microRNAs) in the nucleus, followed by rapid cleavage by a nuclear microprocessor complex including Drosha, yielding a precursor miRNA (pre-miRNA). The pre-miRNA is then transported to the cytoplasm by Exportin-5, followed by Dicer-mediated cleavage loading of the short RNA duplex onto an RNA-induced silencing complex (RISC) by physical association with Ago proteins, and unwinding of the guide and passenger strands from the RNA duplex (1–5). The RISC is then brought to a target sequence in the 3'-untranslated region (UTR) of a messenger RNA (mRNA), resulting in a reduction in protein levels through translational repression or mRNA de-adenylation and decay (6,7).

Mammals contain four copies of Ago genes: Ago1–Ago4, also termed eukaryotic initiation factors 2C1–2C4 (*EIF2C1*, *EIF2C2*, *EIF2C3* and *EIF2C4*). While all of these Ago genes have diverged from a single common ancestral gene (8), three (*AGO1*, *AGO3* and *AGO4*) are present as tandem copies in the same orientation on human chromosome 1p34.3, while *AGO2* is present on its own on human chromosome 8q24.3. The same pattern of chromosomal arrangement is present in mice, with the three Ago genes on mouse chromosome 4qD2.2 and one on chromosome 15qD3. Ago2 has retained the ability to cleave target mRNAs guided by small-interfering RNAs (siRNAs) in addition non-cleavage miRNA based gene silencing (9,10). Ago1, Ago3 and Ago4 most likely have lost this cleavage function and instead solely rely on a non-cleavage mechanism to induce translational repression of target mRNAs via miRNAs. Argonaute proteins are sequentially divided into an N-terminal domain, a piwi argonaute and zwiille domain that binds the 3'-end of a miRNA, a middle domain where the 5'-end of a miRNA finds its binding pocket, followed by an RNase-H-like P-element induced wimpy testis (PIWI) domain through which Ago2 is able to induce endonucleolytic cleavage (11–15).

Interestingly, several groups have reported difficulties in expressing FLAG-tagged wild-type Ago3 and

*To whom correspondence should be addressed. Tel: +1 650 498 6531; Fax: +1 650 498 6540; Email: markay@stanford.edu

Ago4, generated by the Tuschl group (16), by transient transfection (17,18). We recently noted similar observations; in controlled transfection studies, over-expression of FLAG-Ago1 and FLAG-Ago2 led to robust production of the transgene-derived protein, whereas much lower amounts of FLAG-Ago3 and FLAG-Ago4 were detected (19). However, codon optimization of Ago3 and Ago4 released the repression of Ago3 and Ago4 expression, leading to levels of FLAG-Ago3 and FLAG-Ago4 proteins that were commensurate with that of FLAG-Ago1 and FLAG-Ago2 (19). The principle of codon optimization in this context is to replace each triplet codon with one that represents the most frequently used synonymous codon in the species being studied. For example, if GTG is the most abundant codon for valine in mice, it will be selected to code for valine each time, it is present in the coding sequence. This strategy has been employed previously to create robust protein expression and to enable expression in tissues, organisms or cell types that otherwise do not express the protein [reviewed in (20)]. Codon optimization was also applied to Ago1 and Ago2; however, this did not have a notable effect on their expression levels, which remained at levels comparable with their respective wild-type sequence constructs (19). Based on this observation, we sought to determine whether there are sequence features of *Ago3* and *Ago4* that may be responsible for the deviation in expression between wild-type and optimized constructs.

MATERIALS AND METHODS

Codon and sequence optimization

Overviews of and sequence codon optimization for Ago2, Ago3 and Ago4 have been described previously (19). FLAG-tagged wild-type argonaute plasmids driven by a cytomegalovirus promoter were obtained from Addgene (plasmids 10820, 10822, 10823 and 10824) as described by Meister and colleagues (16) and were sequenced for verification. A FLAG-tagged GFP construct (Addgene construct 10825) was used as a negative control. Basically, sequences were sent to GeneART (Germany) that examined the sequences and converted each triplet codon to the codon that is most frequent in humans. The exception to this was if an AU-rich element (ARE), a cryptic splice donor or splice acceptor, or a polyadenylation binding site was present in the sequence as well; these sequence features were removed from the final coding version. Optimized Ago1 was generated for the purpose of this study as an additional basis of comparison to wild-type Ago1. Furthermore, two constructs were generated by GeneART for Ago3 which did not optimize the codons, but instead either removed only regions in the transcript that harbored potential cryptic splice sites and polyadenylation sites, or those sites in combination with all ARE sites as well.

Rare codon distribution score

To calculate the cumulative frequency of rare codons a modification of the codon adaptability index was used. Each codon was scored based on the formula $(1 - F_i/F_{ij})$

where F_i is the frequency of the codon used in the average human proteome, and F_{ij} is the frequency of the most common codon at that particular amino acid. Thus, if the most frequent codon is used at a given position, it will have a score of zero, and the slope of negative values reflects how rare the codons are at each position. The cumulative score was summed across the entire protein. Protein alignments were done by the ClustalW method.

Northern and western blots

Transfections were performed in 6-well plates with 2 μ g of each human argonaute plasmid co-transfected with 10 ng of pIRESneo GFP as a transfection control. In HEK293 cells, Lipofectamine 2000 (Invitrogen) was used as a transfection reagent. For MEF cell lines, 12-well dishes were seeded with 1.25×10^4 cells were seeded and 24 h later 500 ng of DNA were transfected using TransIT LT1 transfection reagent (Mirus). In both cases, 48 h post-transfection, total RNA was extracted using Trizol; protein was extracted using M-PER (Pierce). Mouse argonaute sequences were transfected in Huh-7 cells using 800 ng of plasmid DNA in 24-well dishes.

Western analysis was performed using antibodies against a FLAG tag (Sigma, 1:1000) and β -Actin (Santa Cruz Biotechnology, sc-57778, 1:1000). In total, 10 μ g of protein were loaded. Northern blots were performed on a 1% agarose gel using 5 μ g of Trizol-extracted RNA with a P32 end-labelled probe against the FLAG sequence, transferred to a Hybond N+ membrane (Amersham) and visualized using an autoradiography screen.

Quantitative RT-PCR

Gene-specific TaqMan probes were ordered from Applied Biosystems including Hs01084653_m1 for *EIF2C1*, Hs00293044_m1 for *EIF2C2*, Hs00227461_m1 for *EIF2C3* and Hs01059731_m1 for *EIF2C4*. Actin and Gapdh probes were used for internal normalization. Two microgram of total RNA were reverse-transcribed using Superscript II (Invitrogen) in a 20 μ l reaction volume. For qPCR reactions, 20 ng of cDNA were amplified on a 7900HT real-time PCR system (Applied Biosystems). To cross-compare among genes, a standard curve was generated for each of the argonaute genes using serial dilutions of the plasmids used for transfection. The panel of human RNA samples was obtained from Ambion.

Construct generation

Swapped domain constructs were generated by identifying restriction sites unique in either wild-type or optimized sequence and generating primers incorporating these sites to PCR out the corresponding N-terminal or C-terminal region. For Ago3, restriction sites that were present in either wild-type or optimized constructs were used including *PasI* at nucleotide 588 and *BspEI* at 1163. To swap the region after *PasI* (in bold) the forward primer, 5'-CACCCCCTGGGAGGGGGCAG was amplified using the Ago3 wild-type construct with a reverse primer outside the coding region, 5'-GACAGCGA ATTAATTCCAGCA. This product was digested with *PasI* and *EcoRI* (restriction site present at the terminus

of the gene), and ligated with an Ago3 optimized construct digested with the same restriction enzymes. The same approach was used to generate a portion of wild-type Ago3 in an Ago3 optimized backbone; the T7 forward primer was amplified with a reverse primer incorporating a *PasI* site, 5'-CCCTCCCAGGGGGTGGTCATA, digested with *NotI* and *PasI* and ligated with the Ago3 optimized construct that was digested with the same enzymes. For *BspE1F*, the forward primer to amplify the region after nucleotide 1163 was 5'-TTGGTCCGGAGTGCAAATTAT, and the reverse primer to amplify the region before nucleotide 1163 was 5'-GCACTCCGGACCAATCTGCTA. For Ago4, the restriction sites *BstE1I* at nucleotide 780, *PmeI* at 1494 and *BstB1* at 2073 were used. *BstE1I* is present in both wild-type and optimized constructs, so these regions were digested directly from the constructs and swapped. For *PmeI*, the reverse primer 5'-TCAGGTGTTAAACATGGGCTCCACGCTGTCG was used and for *BstB1*, the forward primer 5'-GCCATTCGAAA GGCCTGCATC was used for amplification in Ago4 optimized constructs.

Ago2, Ago3 and Ago4 3'-UTRs were appended to the end of their respective wild-type and optimized constructs using a PCR-based strategy with an *EcoRI* restriction site in the forward primer and a *SacI* restriction site in the reverse primer as follows: Ago2UTR_F, 5'-CCGCTC GAGCATGTTTTAGTGTTTAGCGAT; Ago2UTR_R, 5'-TCCCCGCGGTACTGCAAACCAGATATATAT; Ago3UTR_F, 5'-GGAATTCATAGTCCAAGTATATTCTCTG; Ago3UTR_R, 5'-TCCCCGCGGTAGTTTGC CATATTTTATATT; Ago4UTR_F, 5'-GGAATTCGAG TCTCAGAAAAAGAACTCA, Ago4UTR_R, 5'-TCCC CGCGGAGAACTTGCTTTCATCCCA. The UTRs were also cloned into a *PsiCHECK-2* vector (Promega) for luciferase analysis using a similar set of primers except that the *XhoI* restriction site was used in place of *EcoRI* in the forward primer and *SpeI* was used in place of *SacI* in the reverse primer.

To generate a construct containing optimized Ago4 with 40 nt in the N-terminal sequence of wild-type Ago4, PCR primers were generated to incorporate the *NotI* restriction site (in bold), the first 40 nt of Ago4 optimized sequence followed by 10 nt of Ago4 wild-type: 5'-GCGGC CGCATGGAGGCGCTGGGACCCGGACCTCCGGC TAGCCTGTTTCAGCCCCCAG. This was amplified with a reverse primer which included a *BstE1I* site (bold) that is present in the Ago4 optimized sequence: 5'-GGTCA CCTCCACCTTCAGCCCCGATC. A similar strategy was employed to generate a wild-type Ago4 construct with 40 nt of optimized sequence at the extreme N-terminus using the forward primer 5'-GCGGCCGCATGGAAGC CCTGGGCCCTGGCCCTCCCGCCAGCCTGTTCCA GCCACCTCG, and the reverse primer which also contained a *BstE1I* restriction site (in bold): 5'-GGTCACCTC AACTTTGAGACCTCTGATT.

To generate Ago4 wild-type sequences with ARE sequences mutated out, a *QuickChange* site-directed mutagenesis kit was used (Stratagene). This converted two AREs, one starting at nucleotide 351 (ATTTA > CT TTC) using the forward primer 5'-GAGGGTAAAGACC AAACCTTCAAAGTGTCTGTTTCAGTGG (mutations

in bold) and the other at nucleotide 1370 (ATTTA > AC CTG) using the forward primer 5'-CAGAAACAATGTA GGGAAGACCTGCTAAAGAGTTTCACTGACC.

Mouse Ago1–3 sequences were PCR amplified with the following primers using cDNA from C57BL/6 mouse as template: mAgo1for, 5'-AAATATGCGGCCGCATGGA AGCGGGACCCCTCGGGAGCAGC; mAgo1rev, 5'-GT CCGGAATTCTCAAGCGAAGTACATGGTGCGTA GAG; mAgo2for, 5'-AAATATGCGGCCGCATGTACT CGGGAGCCGGCCCCGTTC; mAgo2rev, 5'-GTCCGG AATTCTCAAGCAAAGTACATGGTGCGCAGTGTG; mAgo3for, 5'-AAATATGCGGCCGCATGGAAATCG GCTCCGCAGG; mAgo3rev, 5'-GTCGCGGATCCTTA AGCGAAGTACATTGTGCGTAAG. The argonaute sequences were subsequently cloned into *NotI/EcoRI* digested *pIRES* vector. For mouse Ago4, the sequence was obtained from RZPD (IRAV p968A07162D), PCR amplified using the forward primer 5'-AAATATGCGG CCGCATGGAAATCGGCTCCGCAGG and reverse primer 5'-AAGCGCAATTGGCGATCGCTCAGGCAA AATACATAGTGTGCTGG, digested with *NotI/MfeI* and cloned into *NotI/EcoRI*-digested *pIRES* vector.

Luciferase analysis

The *PsiCHECK-2* vector system contains both *Firefly* and *Renilla* luciferase for internal normalization purposes. Luciferase measurements were performed using a dual-luciferase kit (Promega) using the supplied protocol and read on a *Modulus Microplate Luminometer* (Turner BioSystems).

RESULTS

To determine the overall expression level of Ago mRNA transcripts, we performed quantitative RT-PCR measurements of all four human Ago mRNA transcripts in various tissues using *TaqMan* probes and primers sets (Figure 1A). These experiments demonstrated that on average, expression of *Ago1* or *Ago2* was highest, particularly in skeletal muscle, brain and placenta. Conversely, *Ago3* and *Ago4* consistently had a low level of expression in all tissues examined. We investigated whether this reduced expression level could result from higher amounts of rare codons present in Ago3 and Ago4. This analysis revealed that indeed Ago3 and Ago4 have a similar and more frequent usage of rare codons compared with Ago1 and Ago2 (Figure 1B). Notably, the distribution of rare codons is relatively evenly spaced throughout the Ago3 and Ago4 proteins, while the frequency for Ago2 is markedly bi-modal with more rare codons from approximately amino acids 1–400 and more frequent codons thereafter. A line of best fit for codon score for Ago2 yielded slopes of -0.174 for the first portion and -0.118 for the second portion (Figure 1B), though the reason for this bi-modal pattern of codon usage is presently unclear. Upon examining the first 100 amino acids of each Ago protein for rare codons, it is interesting that Ago3 and Ago4 initially have the fewest rare codons relative to Ago1 and Ago2, and in particular the codon score for Ago3 only drops below that of Ago2

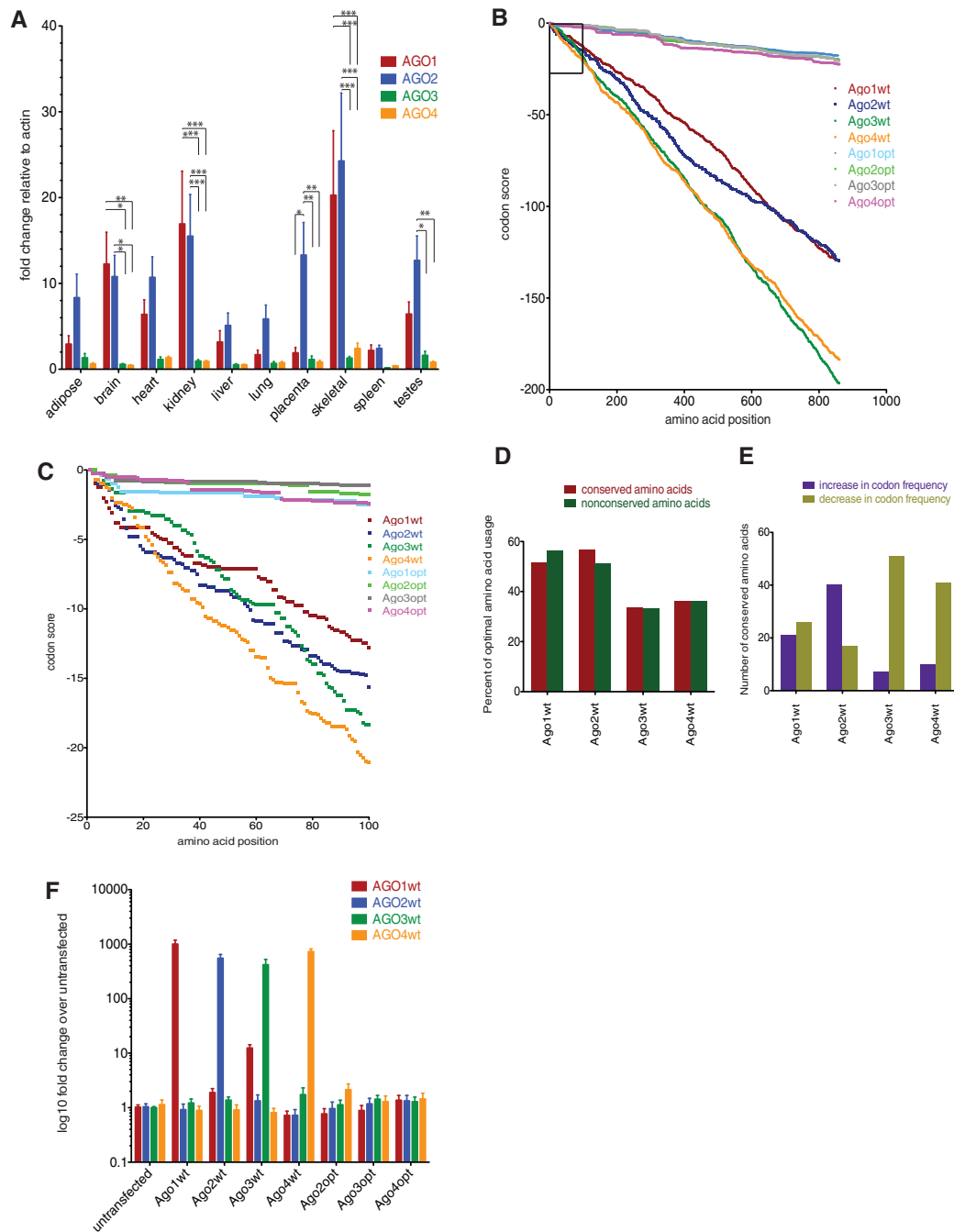


Figure 1. Ago3 and Ago4 consistently have a lower level of expression across various tissues, but a higher cumulative abundance of rare codons. **(A)** Relative levels of argonaute mRNA expression in various human tissues. Data represent mean plus standard error of the mean and data points are normalized to actin using TaqMan probes. A standard curve of argonaute DNA plasmids was constructed for cross-comparison purposes. Two-way ANOVA significance tests with a Bonferroni post-hoc test were performed to compare expression levels between argonaute genes for each tissue; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **(B)** The rare codon distribution score across each Ago gene. A rare codon score was calculated from 0 to -1 at each amino acid position and the cumulative score was plotted for wild-type and optimized Ago1–4. Thus, the lower the score, the more rare codons are present. **(C)** The score for the first 100 amino acids, boxed in B is shown. **(D)** The four human argonaute proteins were aligned by the ClustalW method. Locations where each argonaute used the most frequently used codon for a given amino acid was tabulated in instances where all four argonautes were identical for a given amino acid (conserved, $n = 612$) and where they differed (non-conserved, $n = 245$ – 249). **(E)** Regions where the four argonautes share the same amino acid, but three use one codon for this and the fourth uses a different codon were identified, and scored based on whether codon present in the unique argonaute had an increased codon frequency or a decreased codon frequency. **(F)** The effect of transient transfection of individual wild-type (wt) and optimized (opt) Ago constructs on the expression of other argonaute genes using probes that specifically recognize only wild-type argonaute genes. The fold change of each argonaute gene increased in a comparable manner, though Ago3 and Ago4 had lower basal (and thus final) absolute levels of expression.

after ~80 amino acids (Figure 1C). This may be advantageous to expression of all the argonautes as a ramp of rare codons at the first 50 amino acids has been shown to help position ribosomes evenly across the beginning of the transcript, thus enabling smooth translation (21). While the N-terminus of the protein is relatively more divergent between argonautes in amino acid conservation, we found that the choice of optimal versus non-optimal codons did not depend on amino acid conservation (Figure 1D). Overall, 612 amino acids were identical between all four Ago proteins and in these locations, Ago1 used the most frequent codon 51.6% of the time and Ago2 56.7% of the time. Conversely, Ago3 and Ago4 used the most frequent codon only 33.1 and 36.3% of the time, respectively. We further focused on 213 fully conserved amino acids where three argonautes use one codon and the last argonaute uses a different one. These would represent instances where a mutation has likely occurred solely in the one unique argonaute, or between Ago2 and the ancestral gene common to Ago1, Ago3 and Ago4 (8). We would expect that the mutations would have an equal probability of changing to a codon that is more frequently expressed compared with a codon that is less frequent. This is what is observed for Ago1 with 21 places where the change in Ago1 led to use of a codon that is more optimal relative to Ago2–4, and 26 places where the change is less optimal (Figure 1E). However, for Ago2, there is a 2.35-fold bias towards more frequent codons while Ago3 and Ago4 have a 7.29- and 4.10-fold bias toward less optimal codons, respectively.

To confirm the notion that particular coding sequence features were responsible for this differential level of expression of Ago proteins, we generated codon and sequence optimized constructs of all four human Ago genes. These constructs were devoid of any RNA-instability motifs, potential internal cryptic splice sites, poly-A signals and any other sequence feature that had the potential to reduce levels of mRNA and protein expression. In all, 464, 454, 646 and 596 nt were altered in Ago1–4, respectively (18–26% of total nucleotides changed) (Table 1), leading to sufficient sequence divergence from the wild-type versions that their expression could no longer be detected with the TaqMan probes designed for the wild-type Ago mRNAs (Figure 1F).

Northern blot analysis on RNA isolated from HEK293 cells transfected with the individual Ago constructs that contain the same anti-FLAG probe for each set of

transfected cells confirmed that the level of Ago3 and Ago4 had a lower level of expression relative to Ago1 and Ago2 (Figure 2A), and that this difference was alleviated in the codon-optimized constructs. Indeed, this result was true both for mRNA levels (Figure 2A) and protein levels (Figure 2B). Transient transfection of murine FLAG-Ago cDNAs similarly exhibited reduced expression of mouse Ago3 and Ago4 relative to Ago1 and Ago2 (Figure 2C). The overall distribution of codons for Ago3 and Ago4 compared with Ago1 and Ago2 is shown in Supplementary Table S1.

Regulation by one Ago gene on the remaining argonautes could be envisioned as a mechanism to maintain a certain window of expression of the various genes, or to ensure they are expressed only in a particular temporal or spatial manner. Thus, we explored the possibility that an Ago2-dependent cleavage event was responsible for the low levels of Ago3 and Ago4. We examined the coding regions of Ago3 and Ago4 for potential miRNA target sites that may act in an Ago2-dependent cleavage manner, which generally requires exact complementarity with nucleotides 1–11 or 2–12 nucleotide seed region of a miRNA. We predicted target sites for several miRNAs in the coding region of Ago3 and Ago4 (Supplementary Figure S1). This includes hsa-miR-1913, hsa-miR-497* and hsa-let-7a-2* in Ago3 and hsa-miR-1913, hsa-miR-29a*, hsa-hsa-miR-17*, hsa-miR-18b* and hsa-mir-769-5p in Ago4. Target sites for hsa-miR-497* and hsa-miR-33a* were additionally predicted in Ago2. The genomic interval for hsa-miR-1913 is not present in the mouse and the mature hsa-miR-769 is not conserved in the mouse, thus these miRNAs were excluded as having a potential cleavage role. However, if Ago2 is responsible for cleavage-based repression of Ago3 and Ago4, transcript and protein levels of Ago3 and Ago4 should be elevated in Ago2^{-/-} cells, since Ago2 is the only mammalian Ago that has the potential to cleave target mRNAs. When constructs expressing the Ago genes were transfected into mouse embryonic fibroblast (MEF) cells that had a disrupted and inactive Ago2 transcript (9), the level of expression of wild-type Ago3 and Ago4 remained low to non-detectable, while codon-optimized Ago3 and Ago4 levels were also unchanged (Figure 2D). This indicates that putative cleavage-mediated regulation of the transcripts is not responsible for the low levels of Ago3 and Ago4, though it is still possible that miRNA sites exist for

Table 1. Sequence features in wild-type (wt) versus optimized (opt) argonaute constructs

	Ago1wt	Ago1opt	Ago2wt	Ago2opt	Ago3wt	Ago3opt	Ago4wt	Ago4opt
Nucleotide changes	464		454		646		597	
Nucleotides changed (%)	18		18		26		24	
Codon adaptability index	0.80	0.98	0.80	0.97	0.71	0.97	0.73	0.97
GC content (%)	54	62	56	64	45	64	47	64
Prokaryotic inhibitory motifs	nd	nd	5	0	7	0	2	0
Polyadenylation sites	0	0	1	0	1	0	2	0
Consensus splice donors	nd	nd	2	0	3	0	2	0
ARE RNA-instability motifs	1	0	1	0	6	0	6	0

nd, not determined.

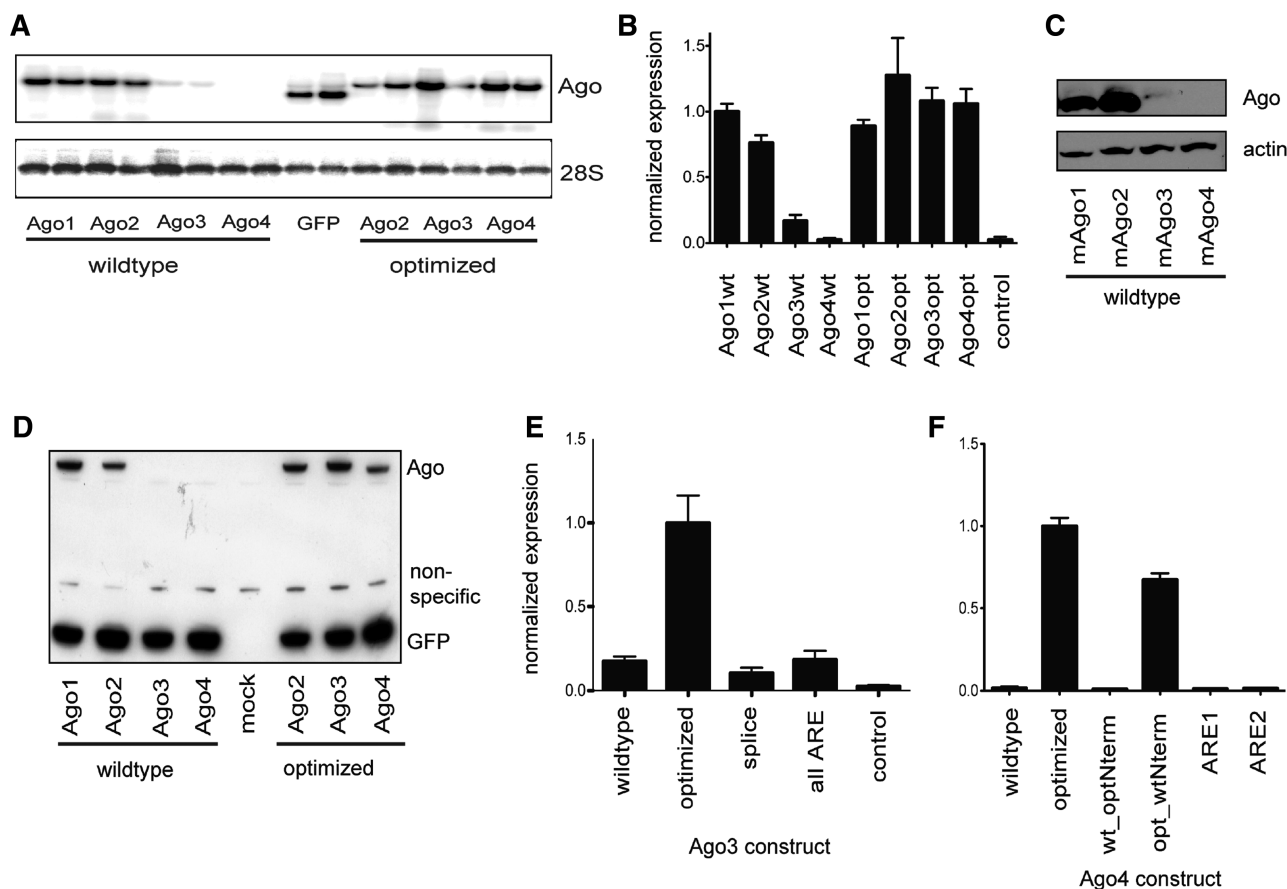


Figure 2. Ago3 and Ago4 mRNA and protein expression is decreased dramatically relative to Ago1 and Ago2 while this expression difference is alleviated by codon and sequence optimization. (A) Northern blot of argonaute wild-type or codon-optimized Trizol-extracted RNA samples from 293 cells loaded in duplicate shows reduced expression of Ago3 and Ago4. The blot was stripped and re-probed for 28S as a loading control. (B) Western blot analysis from FLAG-tagged wild-type and optimized human Ago sequences transiently transfected in HEK293 cells shows that protein levels correlate with the mRNA levels in A. Data represent mean \pm SEM, quantified relative to β -actin and normalized to Ago1wt. (C) Western blot of FLAG-tagged murine Ago1–4 cDNA constructs transiently transfected in Huh7 cells displayed a similar profile of reduced level of expression of Ago3 and Ago4 relative to Ago1 and Ago2. (D) Absence of Ago2 does not influence expression of Ago3 and Ago4. Ago2 null MEF cells were transfected with the same wild-type or optimized plasmids, co-transfected with a GFP plasmid. (E) Western blot quantification of Ago3 constructs that have an abolition of internal potential splice site sequences and all AREs does not account for the reduced protein expression of wild-type relative to optimized sequences. Bands are quantified relative to β -actin and normalized to Ago3opt levels. (F) Western blot quantification of Ago4 constructs normalized to Ago4opt expression. Wt_optNterm represents modification of only the first 40 nt of Ago4wt to corresponding Ago4opt sequences. Opt_wtNterm contains a wild-type sequence of Ago4 for the first 40 nt and an optimized sequence thereafter. ARE1 and ARE2 represent site-directed mutagenesis of two individual AREs in wild-type Ago4 sequence at nucleotide positions 351 and 1370, respectively.

non-cleavage-based regulation of the mRNA transcripts. We have shown that miRNAs are ineffective at targeting coding regions unless their binding sites are immediately downstream of rare codons (22). While certain infrequent codons are present more often in Ago3 and Ago4, most notably GTA encoding Valine (Supplementary Table S1), a preponderance of miRNA sites following these rare codons was not observed. In addition, transient transfection of FLAG-HA tagged human Ago expression cassettes into HEK293 cells resulted in an increase in the transcript level of the expressed Ago and did not have an effect on the endogenous level of any of the Ago family members, apart from a slight increase in Ago1 detection levels after transfection of Ago3 (Figure 1D).

Another means by which particular mRNAs can be rapidly degraded and turned over is through the presence of RNA-instability motifs within their sequence. While these

elements are typically concentrated in the 3'-UTRs of genes, they theoretically could have the same effect if present within the coding region (23,24). For instance, one miRNA that is involved in this process is miR-16 which, by virtue of its AU-rich sequence and facilitated by tristetraprolin (TTP) is able to recruit the RISC complex to AU-rich sequence elements and induce rapid mRNA turnover (25). Notably, when the coding sequences of Ago3 and Ago4 were examined, a preponderance of AREs were present, six each in Ago3 and Ago4, relative to only one each present in Ago1 and Ago2 (Table 1). These became key targets of our investigation. In order to address the potential influence of these AREs on Ago expression levels, additional Ago3 constructs were engineered in which only the regions where the AREs were present were converted to the codon-optimized sequence and the remainder of the DNA

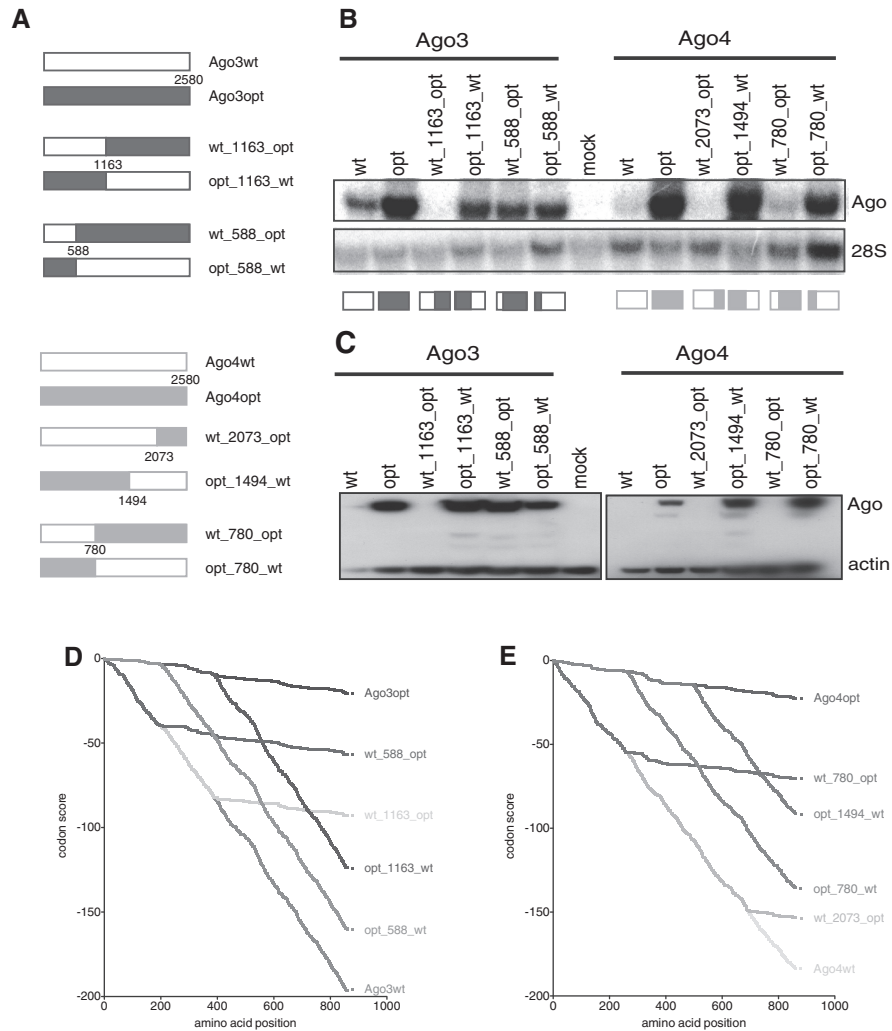


Figure 3. A swap of Ago wild-type (wt) and optimized (opt) sequences reveals the N-terminus of both Ago3 and Ago4 are critical for differential expression. (A) Regions of wild-type and optimized Ago3 and Ago4 constructs were swapped to create chimera proteins as depicted (e.g. 3wt1163opt contains the first 1163 nt of wild-type Ago3 followed by the remaining C-terminal optimized Ago3 sequence). Northern (B) and western (C) results indicate that a sequence motif in the N-terminal portion of the protein is responsible for this differential expression. (D) Rare codon cumulative scores for Ago3 wild-type, optimized and swapped domain constructs were compared as in Figure 1C. Samples with high levels of expression were dark gray while those with low levels of expression were light gray; overall the rare codon burden did not correlate with expression of the construct. (E) A similar rare codon cumulative frequency score was plotted for Ago4 wild-type, optimized and swapped domain constructs, with high-expressing samples in dark gray and low-expressing samples in light gray.

sequence was identical to wild-type Ago3. As an additional test, both cryptic splice donors and acceptors as well as weak polyadenylation binding protein sites were also converted to optimized sequence to prevent the potential generation of truncated transcripts or proteins which could subsequently be degraded and no longer detected. This resulted in changes in 13 sequence motifs totaling 37 nt. It is important to point out here that this new Ago3 contained considerably fewer changes than the total of 646 that were modified in the original codon-optimized version (see above). Remarkably however, as shown in Figure 2E, the elimination of these AREs and potential splice or polyadenylation sites had no effect on the expression of the Ago3 construct as compared to wtAgo3. Site-directed mutagenesis of two of the AREs in the Ago4 construct similarly had no effect on its

expression relative to Ago4 (Figure 2F). Thus we were able to exclude RNA-instability motifs as the cause of the differential expression among the Ago proteins.

We next asked whether a particular domain of the Ago3 or Ago4 transcript was responsible for their differential levels of expression relative to codon-optimized versions, or was a global distribution of a particular feature, such as rare codons, responsible. To address this question, we generated chimeric constructs whereby sequences of the wild-type Ago3 were swapped with sequences of optimized Ago3 (Figure 3A). Similar domain swapping was also performed for Ago4, and the resulting effect on mRNA and protein level was analysed (Figure 3B). As seen in Figure 3, the region that was most responsible for Ago3 expression was contained within nucleotides 1–1163 at the N-terminal half of the gene. However, when this region

was further narrowed down to include only nucleotides 1–588 of either wild-type or optimized Ago3 combined with nucleotides 589–2580 of the other construct, the enhanced expression effect was ameliorated, as the overall mRNA and protein levels became comparable to those observed with the wild-type sequence. Thus, though the precise region involved in controlling expression levels cannot be definitively elucidated, it does appear to involve features that are spread throughout the first half of the gene. For Ago4, again the critical region that determines expression was present in the first 780 nt of the gene. Both mRNA and protein levels were repressed when the first 780 nt were from Ago4 wild-type sequence and the remainder were codon and sequence optimized, while high mRNA and protein levels were present in the reciprocal situation when nucleotides 1–780 were codon and sequence optimized and the rest of the gene remained wild-type (Figure 3B and C). The rare codon score was also calculated for these chimeric constructs and compared with wild-type and optimized Agos. In both Ago3 (Figure 3D) and Ago4 (Figure 3E), the final codon score did not correlate with expression levels of each construct. This indicated that it was not the overall codon burden that was implicated but instead rare codons or related sequence features within specific regions. Sequences at the extreme 5'-terminus of the coding region have also been implicated in translational dynamics, aiding in mRNA stability and ribosomal acceleration through the transcript (26,27). Notably, we found that the extreme N-terminus of each coding region was not responsible as swapping the initial 40 nt of wild-type Ago4 with the optimized sequence and optimized Ago4 with an initial wild-type 40 nt did not alter expression (Figure 2F). In addition, the overall secondary structure at the translation initiation site was not dramatically altered when sequences were optimized (Supplementary Figure S2). This correlated with the observation that rare codon distribution was not markedly different for Ago1–4 at the N-terminus of each gene (Figure 1D).

Each of these constructs that were generated in this study contained only the coding region of the Ago genes, raising the possibility that sequence elements in

the 3'-UTR of each gene could influence Ago3 and Ago4 expression. To help address this, we examined the effect of appending the 3'-UTR of Ago2, Ago3 and Ago4 after each gene's respective coding sequence. However, this did not have an effect on protein expression (Figure 4A). In addition, the effect of each 3'-UTR on luciferase expression in a firefly/renilla dual luciferase reporter system was similarly unchanged from a control sequence (Figure 4B), indicating that adding back the 3'-UTR does not over-ride the signals observed that are specific to the coding region.

Finally, another factor that could potentially influence efficiency of translation and gene expression is RNA secondary structure (28). We used a recently published RNA folding algorithm (29), which considers the entire ensemble of possible RNA structures for a given sequence, to assess differences in predicted RNA structure between the wild-type and optimized versions of each Ago. We found that the overall RNA structure is slightly less correlated between wild-type and optimized sequences for Ago3 (correlational coefficient, CC = 0.15) and Ago4 (CC = 0.21) than for Ago1 (CC = 0.22) and Ago2 (CC = 0.24). This result suggests that while secondary structure may be a contributor to the expression differences, and as such merits further investigation outside of the scope of this study, it is likely not the sole factor.

DISCUSSION

We explored several coding region features that may account for the differential gene and protein expression of Ago3 and Ago4 compared to Ago1 and Ago2. Of these features, we experimentally excluded the possibility that AREs in the coding region, despite being more prevalent in Ago3 and Ago4 relative to Ago1 and Ago2, were responsible for the observed differences. In addition, the overall codon distribution across the genes did not seem to be critical for expression. Instead, codon distribution or other related sequence features in the N-terminal regions (but not the direct N-terminus itself) appear to have the greatest impact on sustained levels of Ago3 and Ago4 expression. The intermediate level of expression observed

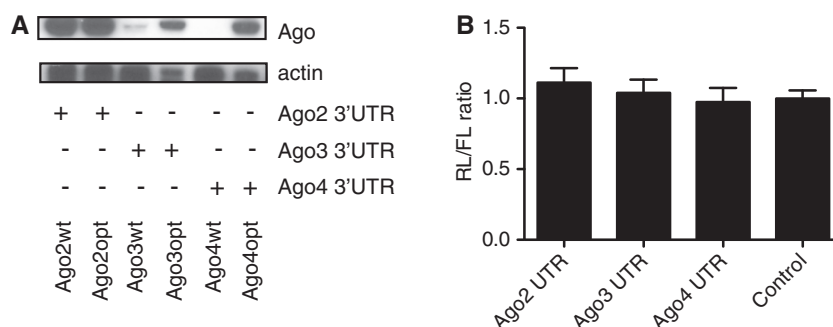


Figure 4. Addition of argonaute 3'-UTRs does not influence or rescue a difference in expression between wild-type and optimized constructs. (A) The 3'-UTR of Ago2 was appended to the coding sequence of wild-type and optimized Ago2, Ago3 or Ago4. No influence on expression levels was observed after western analysis. FLAG-tagged argonaute constructs co-transfected with FLAG-GFP were detected using an anti-FLAG antibody and also probed for actin for normalization purposes. (B) Expression of FLAG-GFP incorporating the 3'-UTRs of Ago2, 3, 4 or none (control). Addition of the UTR after the *Renilla* luciferase reporter gene had no influence on expression as measured by relative levels of renilla to firefly luciferase in 293 cells.

after domain swapping indicates that more than just one sequence motif is responsible. This makes it prohibitive to determine the relevant contribution of each of the individual sequence motifs. However, the finding does exclude the possibility that rare codons across the whole coding region of the gene are responsible. Examination of the coding region of these genes did reveal that certain rare codons, most notably those for valine (GTA), were present in greater frequencies in Ago3 and Ago4 relative to Ago1 and Ago2 (Supplementary Table S1); however, the fairly uniform distribution of these rare codons throughout Ago3 and Ago4 is inconsistent with our results, which suggested that the N-terminal region was the most critical for determining expression. Another important distinction relates to the level of RNA expression versus protein abundance. We have observed that both the RNA and protein levels of Ago3 and Ago4 are diminished. The level of rare codons, while predominantly influencing protein expression can also impact RNA stability. Pausing of a ribosome at rare codon clusters can increase the accessibility of RNA-instability elements to degrade mRNA (30,31). Because the cDNA constructs were cloned into the same expression vector, DNA features in promoters and introns such as transcription factor binding sites, histone and DNA methylation marks, insulators and enhancers do not play a role in this change of expression. Further, pre-mRNA splice factors would not be influential in this system.

A recent study performed HITS-CLIP analysis and identified miRNA target sites in the P13 mouse brain (32), including miR-9, miR-153 and let-7 sites in the Ago2 coding region and a let-7 and miR-9 site in the coding region of Ago3. However, Ago1 and Ago4 did not have annotated miRNA target sites. While this targeting may represent a tissue-specific phenomenon, particularly for miR-9 an abundant miRNA in the brain, the presence of potential target sites in the coding region of these argonaute transcripts is still intriguing. The fact that no miRNA target sites were present in Ago4, and the observation that the only miRNAs that displayed evidence of binding in Ago3 were also found in Ago2 argues against miRNA targeting of the coding region as the primary mechanism for differential regulation of expression of Ago3 and Ago4 relative to Ago1 and Ago2. We noted variable expression of Ago2 depending on the tissue examined (Figure 1A). Since, we observed no change in Ago2 protein levels upon its codon and sequence optimization, it suggests that additional factors apart from rare codons and RNA sequence features are still responsible for the expression of this gene.

Maintaining a high level of Ago3 and Ago4 through these codon-optimized constructs could help indicate a distinct role for these genes with respect to the other argonautes. However, the possibility remains that the two Ago genes are being evolutionarily phased out. While a formal analysis of the nucleotide substitutions that have taken place over the evolution of the genes could help address this issue, the lower level of expression of both Ago3 and Ago4 could represent a decreased dependency for three mammalian argonaute genes. Alternatively, sequence elements and rare codons may

have accumulated in Ago3 and Ago4 to dampen their expression in certain contexts. The high frequency of rare codons in Ago3 and Ago4 is conserved across species, which is suggestive of evolutionary constraint due to a heretofore unknown functional importance.

Another issue to address is the overall benefits and perils of codon and sequence optimization. There is no guarantee that an optimized version of a gene will have a sustained level of expression in an endogenous scenario, or that such an effect will actually be beneficial for a cell. Indeed, certain sequence elements encoded in the mRNA may positively or negatively regulate Ago3 and Ago4 expression. These sequence motifs likely would be abrogated when generating codon-optimized constructs. In addition, by constantly using the most frequent codon at each position, the anticodon-containing tRNAs that recognize this codon could become limiting particularly in repetitive stretches of identical codons. Our laboratory has been interested in maintaining high levels of argonaute expression to prevent saturation of RNAi components which can cause lethality in mice (33). While we have evidence that maintaining high levels of Ago2 is crucial for preventing this toxicity (19), the levels of other argonaute proteins may also aid in circumventing this toxicity. In this context, administering optimized Ago3 or Ago4 in conjunction with short-hairpin RNAs (shRNAs) could release the burden of a saturated RISC complex and enable both exogenous Ago2-mediated cleavage of target mRNAs while maintaining the ability for endogenous miRNA pathways to operate unperturbed. Another possibility is that the relative ratios of the various Argos may be important for maintaining a balance of various miRNAs which are RISC loaded (active) versus unloaded (inactive) under different physiological cellular conditions. The reasons as to why Ago3 and Ago4 have low levels of expression are presently unclear; however, we can conclude that certain sequence features of Ago3 and Ago4, present in the first half of the mRNA are responsible for their endogenous expression levels. Further studies will likely reveal more insights into the biological role of the non-cleaving Ago proteins in mammalian systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

We would like to thank Leszek Lisowski and Dan Cao for critical reading of this article and discussions. We also thank Matt Halvorsen and Alain Laederach (UNC) for their RNA folding algorithm, SNPfold.

FUNDING

The National Institutes of Health [DK78424]; a Canadian Institutes of Health Research Bisby Fellowship (to

P.N.V.); the National Institutes of Diabetes and Digestive and Kidney Diseases/National Institutes of Health [K99 Grant No. 1K99DK091318-01 (to P.S.)]; and NIH DK078424 (to M.A.K.). Funding for open access charge: NIH NIDDK.

Conflict of interest statement. None declared.

REFERENCES

- Czech, B. and Hannon, G.J. (2011) Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.*, **12**, 19–31.
- Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. and Siomi, M.C. (2005) Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.*, **19**, 2837–2848.
- Rand, T.A., Petersen, S., Du, F. and Wang, X. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, **123**, 621–629.
- Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.*, **79**, 351–379.
- Guo, H., Ingolia, N.T., Weissman, J.S. and Bartel, D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
- Sasaki, T., Shiohama, A., Minoshima, S. and Shimizu, N. (2003) Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics*, **82**, 323–330.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, **305**, 1437–1441.
- Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, **304**, 594–596.
- Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T. and Patel, D.J. (2005) Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, **434**, 666–670.
- Parker, J.S., Roe, S.M. and Barford, D. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, **434**, 663–666.
- Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, **305**, 1434–1437.
- Wang, Y., Juraneck, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T. and Patel, D.J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, **461**, 754–761.
- Wang, Y., Sheng, G., Juraneck, S., Tuschl, T. and Patel, D.J. (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature*, **456**, 209–213.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, **15**, 185–197.
- Nelson, P.T., De Planell-Saguer, M., Lamprinak, S., Kiriakidou, M., Zhang, P., O'Doherty, U. and Mourelatos, Z. (2007) A novel monoclonal antibody against human Argonaute proteins reveals unexpected characteristics of miRNAs in human blood cells. *RNA*, **13**, 1787–1792.
- Chu, Y., Yue, X., Younger, S.T., Janowski, B.A. and Corey, D.R. (2010) Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Res.*, **38**, 7736–7748.
- Grimm, D., Wang, L., Lee, J.S., Schurmann, N., Gu, S., Borner, K., Storm, T.A. and Kay, M.A. (2010) Argonaute proteins are key determinants of RNAi efficacy, toxicity, and persistence in the adult mouse liver. *J. Clin. Invest.*, **120**, 3106–3119.
- Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**, 32–42.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I. and Pilpel, Y. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, **141**, 344–354.
- Gu, S., Jin, L., Zhang, F., Sarnow, P. and Kay, M.A. (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.*, **16**, 144–150.
- Bernstein, P.L., Herrick, D.J., Prokopcak, R.D. and Ross, J. (1992) Control of c-myc mRNA half-life in vitro by a protein capable of binding to a coding region stability determinant. *Genes Dev.*, **6**, 642–654.
- Marshall, B., Isidro, G. and Boavida, M.G. (1996) Naturally occurring splicing variants of the hMSH2 gene containing nonsense codons identify possible mRNA instability motifs within the gene coding region. *Biochim. Biophys. Acta*, **1308**, 88–92.
- Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.C., Gram, H. and Han, J. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell*, **120**, 623–634.
- Gu, W., Zhou, T. and Wilke, C.O. (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.*, **6**, e1000664.
- Kudla, G., Murray, A.W., Tollervey, D. and Plotkin, J.B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, **324**, 255–258.
- Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Rupp, E. and Ziv-Ukelson, M. (2011) Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.*, **12**, R110.
- Halvorsen, M., Martin, J.S., Broadway, S. and Laederach, A. (2010) Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.*, **6**, e1001074.
- Sauna, Z.E. and Kimchi-Sarfaty, C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.*, **12**, 683–691.
- Tsai, C.J., Sauna, Z.E., Kimchi-Sarfaty, C., Ambudkar, S.V., Gottesman, M.M. and Nussinov, R. (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J. Mol. Biol.*, **383**, 281–291.
- Chi, S.W., Zang, J.B., Mele, A. and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
- Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F. and Kay, M.A. (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature*, **441**, 537–541.