RESEARCH PAPER

RNA Biology 12:7, 761–770; July 2015; Published with license by Taylor and Francis Group, LLC

Early origin and adaptive evolution of the GW182 protein family, the key component of RNA silencing in animals

Andrzej Zielezinski and Wojciech M Karlowski*

Department of Computational Biology; Institute of Molecular Biology and Biotechnology; Adam Mickiewicz University; Poznan, Poland

Keywords: Argonaute, CCR4-NOT, gene silencing, GW-repeats, GW182, microRNA, RNAi, TNRC6, WG/GW

The GW182 proteins are a key component of the miRNA-dependent post-transcriptional silencing pathway in animals. They function as scaffold proteins to mediate the interaction of Argonaute (AGO)-containing complexes with cytoplasmic poly(A)-binding proteins (PABP) and PAN2-PAN3 and CCR4-NOT deadenylases. The AGO-GW182 complexes mediate silencing of the target mRNA through induction of translational repression and/or mRNA degradation. Although the GW182 proteins are a subject of extensive experimental research in the recent years, very little is known about their origin and evolution. Here, based on complex functional annotation and phylogenetic analyses, we reveal 448 members of the GW182 protein family from the earliest animals to humans. Our results indicate that a single-copy GW182/TNRC6C progenitor gene arose with the emergence of multicellularity and it multiplied in the last common ancestor of vertebrates in 2 rounds of whole genome duplication (WGD) resulting in 3 genes. Before the divergence of vertebrates, both the AGO- and CCR4-NOT-binding regions of GW182s showed significant acceleration in the accumulation of amino acid changes, suggesting functional adaptation toward higher specificity to the molecules of the silencing complex. We conclude that the silencing ability of the GW182 proteins improves with higher position in the taxonomic classification and increasing complexity of the organism. The first reconstruction of the molecular journey of GW182 proteins from the ancestral metazoan protein to the current mammalian configuration provides new insight into development of the miRNA-dependent post-transcriptional silencing pathway in animals.

Introduction

In animals, most miRNA sequences are only partially complementary to the sequences they regulate, and catalytically active Argonaute (AGO) proteins do not cleave their targets. To mediate silencing, AGOs must interact with proteins of the GW182 family [glycine-tryptophan repeat-containing protein of 182 kDa].[1,2] Consequently, GW182 proteins play a key role in miRNA-dependent post-transcriptional silencing in animals by functioning as scaffold proteins for the assembly of effector complexes. The AGO–GW182 complexes mediate silencing of the target mRNA through induction of translational repression and/or mRNA degradation [for a review see ref. 2].

To promote mRNA degradation, the GW182 proteins recruited by AGOs interact with the cytoplasmic poly(A)-binding protein (PABP) [3-6] as well as CCR4-NOT and PAN2-PAN3 deadenylase complexes.[7-12] CCR4-NOT and PAN2-PAN3 remove the poly(A) tail and trigger mRNA decay. The CCR4-NOT complex further recruits DDX6, an RNA helicase that acts as a translation repressor and decapping activator.[11,12] Following decapping by the DCP1–DCP2 complex mRNAs are degraded by the cytoplasmic 5′-to-3′ exonuclease XRN1 [for a review see ref.[13]]. The mechanism of miRNA-mediated translation repression may involve distinct steps, including eIF4E/cap recognition and 43S ribosome or 60S ribosome joining [for a review see ref.[14]].

GW182 proteins share a common domain architecture consisting of 2 conserved structural parts: a central ubiquitin-associated (UBA-like) domain and a C-proximal RNA recognition motif (RRM). These domains are embedded in regions that are predicted to be intrinsically unstructured.[15,16] The unstructured regions include the N-terminal, Mid (M1 and M2 separated by PAM2 motif) and C-terminal tryptophan(W)-containing fragments,[8] as well as a glutamine-rich (Q-rich) region located between the UBA-like and RRM domains (**Fig. 1B**).

The W-containing regions exhibit high length diversity, sequence conservation and a number of tryptophan repeats.[17,18] Interestingly, these sequences play an essential role in miRNA-dependent silencing, in contrast to the well-defined domains

*Correspondance to: Wojciech M Karlowski; Email: wmk@amu.edu.pl
Submitted: 03/30/2015; Revised: 05/08/2015; Accepted: 05/11/2015
http://dx.doi.org/10.1080/15476286.2015.1051302

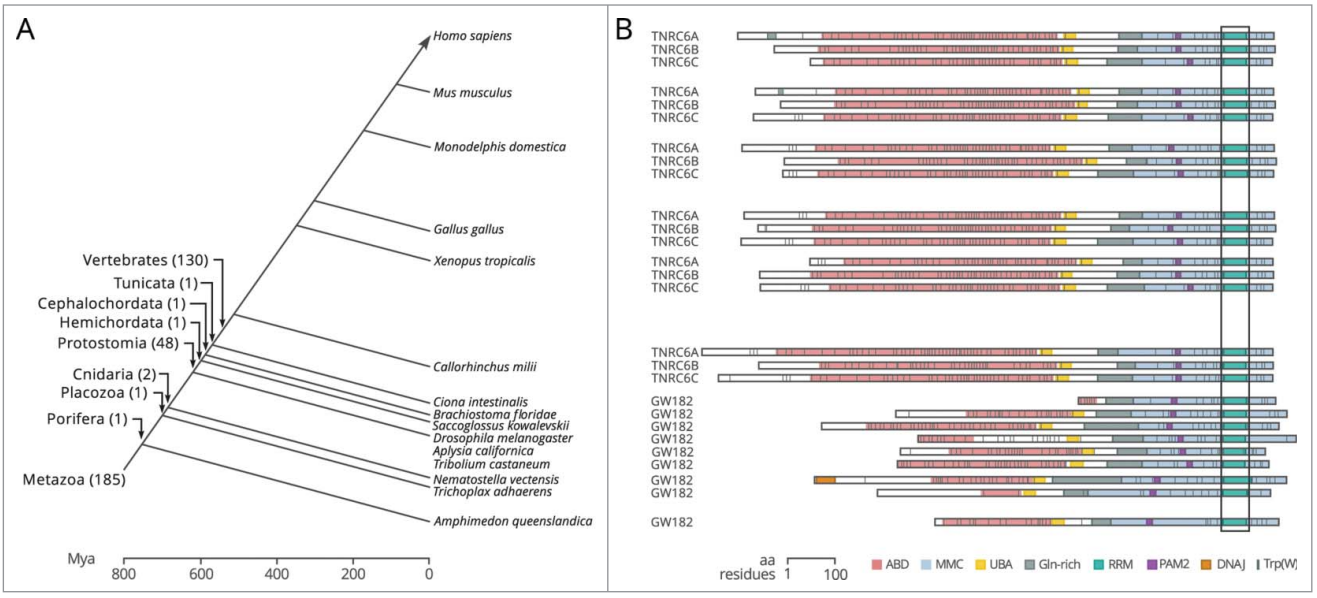

Figure 1. Metazoan evolution of GW182 proteins from sponges to human. (**A**) Evolutionary lineage of the metazoan organisms analyzed in this study based on.[69] The total number of identified metazoan species is indicated. (**B**) Domain structure of GW182 proteins from selected species representing groups shown on the tree (**A**). GW182 proteins are listed according to the alignment of RNA Recognition Motif (RRM) highlighted as a black-bordered rectangle.

(RRM and UBA-like) that are not strictly required for GW182 activity.[2] The N-terminal WG/GW repeat-containing region (ABD, AGO-binding domain) binds the PIWI domain of AGO proteins, while the W-rich region in the C-terminal part of the protein (SD, silencing domain) interacts with PABP through the PAM2 motif and serves as a binding platform for PAN3 and the NOT1 and NOT9 components of the PAN2–PAN3 and CCR4–NOT deadenylase complexes.[7-12]

The role of GW182 proteins in miRNA-mediated gene silencing has been experimentally explored by genetic assays in *C. elegans*, RNAi screens in *D. melanogaster*, and biochemical purifications of AGO-containing complexes from human cells [for a review see ref. 2]. Although the function of GW182 proteins is becoming clear, very little is known about their evolution because the proteins that have been predominantly studied are restricted to narrow taxonomic ranges. Three paralogs (TNRC6A, TNRC6B, TNRC6C) have been identified in vertebrates, up to 3 in insects (GW182)[2,15,17,19] and one in Cnidaria.[20] Members of the GW182 gene family in other early branching metazoan lineages have not been characterized. The *C. elegans* genome encodes 2 proteins: ALG-1-interacting protein 1 (AIN-1) and AIN-2, which interact with AGO proteins (ALG-1 and ALG-2) and are required for miRNA function.[21-24] AIN-1 and AIN-2 contain a small number of GW repeats and lack a defined glutamine-rich region, UBA domains and RRMs. Based on this observation, Eulalio et al. (2009) proposed that AIN-1 and AIN-2 are not members of the GW182 protein family but are functional analogs.[16]

In this study, we reveal the complex composition of the GW182 gene family members in numerous vertebrate and invertebrate lineages and propose a possible mechanism of their evolution from the ancestral metazoan protein to the current mammalian configuration. Our analysis involves an in-depth exploration of sequence evolution dynamics inferred from inter- and intra-protein comparisons. Finally, we apply a recently developed method for AGO- and CCR4-NOT-interacting domain prediction [18] to assess the molecular evolution of the post-transcriptional silencing capacity of the GW182 proteins across the metazoan tree of life.

Results

Early origin and distinct evolutionary pathways of GW182 proteins in animals

The application of sequence- and domain-based search strategies against available genome data sets (see the Materials and Methods section) resulted in the identification of GW182 orthologs across all animal phyla, including the most basal species represented by *Amphimedon queenslandica* (sponges) and *Trichoplax adhaerens* (placozoan) (**Fig. 1A**, **Table S1**). GW182 homologs could not be detected in non-metazoan eukaryotes, including fungi, plants and choanoflagellates (unicellular organisms constituting a sister group to metazoans), confirming that GW182s are animal-specific and arose with the origin of multicellularity.

In total, we identified 448 GW182 genes in 185 genomes representing major taxonomic groups of metazoa: Porifera (1), Placozoa (1), Cnidaria (2), Protostomia (48) and Deuterostomia (133) (**Fig. 1A**, **Table S1**). As shown in **Figure 1B**, the domain architecture of GW182 proteins is generally conserved across the analyzed organisms and exhibits consistency in the content and arrangement of functional modules. However, some lineage-

specific differences are apparent; for example, the GW182 protein from *N. vectensis* contains an additional DnaJ domain at the N-terminus of the protein (**Fig. 1B**). This domain is not present in any other identified GW182 homologues. DnaJ-containing proteins have been reported to be involved in the regulation of the ATPase activity of the Hsp70/Hsp90 chaperone machinery,[25] which is required for loading small RNA duplexes into AGO proteins.[26]

Similarly, 2 splicing isoforms of the vertebrate TNRC6A gene transcripts (TNRC6A/TNGW1/GW220 and TNRC6A/GW182) differing only in the N-terminal region have been reported in humans. The extended isoform contains an additional N-terminal 253 amino acids containing a polyglutamine (polyQ) repeat motif.[27] In our screening, these glutamine-repeats were identified exclusively in mammals (**Fig. 1**, **Fig. S1A**) and exhibit significant differences in the number and length of Q-rich repeats. In non-mammalian vertebrates, the TNRC6A protein contains an N-terminal extension without a detectable polyQ motif (**Fig. S1B**). No similar TNRC6A N-terminal extensions could be identified in invertebrate species or in the sequences of the other 2 TNRC6 homologs.

Our results indicate that vertebrate genomes typically encode 3 GW182 proteins (TNRC6A-C), with the exception of fishes, where up to 7 homologs could be identified. Conversely, 2 species considered the most closely related to vertebrates, *C. intestinalis* and *B. floridae*, contain only one copy of the GW182 gene (**Table S1**). The expansion of the GW182 protein family is consistent with the 2-rounds-of-polyploidization hypothesis of whole-genome duplications (WGDs) early in vertebrate evolution.[28-30] The single-copy gene of the invertebrate GW182 ortholog has given rise to 3 paralogs located on different chromosomes; in humans, the TNRC6A, TNRC6B and TNRC6C genes are located on chromosomes 16, 22 and 17, respectively.

Lineage-specific expansion of the GW182 gene family occurred in ray-finned (actinopterygian) fishes, where 5 to 7 copies of GW182 genes could be identified. This group includes zebrafish (*D. rerio*) with 7 copies, blind cave fish (*A. mexicanus*) containing 6 copies and 7 other species (for example, fugu fish–*T. rubripes*) with 5 copies of GW182 genes (**Table S1**, **Fig. S2**). Such multiplication of GW182 genes supports recent phylogenomics studies that propose a lineage-specific, third genome duplication in ray-finned fishes (FSGD or 3R).[31] Interestingly, only the TNRC6B and TNRC6C genes remained multiplied in currently living species. TNRC6A is a single-copy gene in all analyzed genomes in this group (**Fig. S2**). By contrast, lobe-finned fish (sarcopterygian) genomes encode up to 4 GW182 genes. For example, *Latimeria chalumna* contains an extra copy of the TNRC6B gene in addition to the 3 vertebrate-characteristic GW182 paralogs (TNRC6A, TNRC6B TNRC6C).

Our analysis of GW182 genes in insects also suggests lineage-specific amplification. The 38 species representing major taxonomic groups (Coleoptera, Diptera, Hymenoptera, Lepidoptera, Dictyoptera, Hemiptera and Phthiraptera) contain one GW182 gene (**Table S1**), including the best-studied GW182 protein, Gawky/DmGW182 from *D. melanogaster*. However, 4 mosquito species (*Anopheles darlingi, Anopheles gambiae, Aedes aegypti* and *Culex quinquefasciatus*) possess 3 copies of GW182 genes. The phylogenetic reconstruction for GW182 proteins from fly, lancelet and human indicates an expansion of the GW182 proteins in mosquitos independent of vertebrates (**Fig. S3**). In addition, the clustering on a single chromosome and arrangement (separation by no or a couple of genes) of the mosquito GW182 paralogs suggests that they emerged most likely by 2 local duplication events in the genome of their last common ancestor.

TNRC6C is a founding member of the GW182 family in vertebrates

The expansion of the GW182 gene family in vertebrates and some insects raises a question about the orthologous relationship between members of the family. The GW182 proteins exhibit high diversity in sequence length, conservation and composition. Therefore, as discussed by Moran et al. (2013), the alignment of full-length GW182 sequences is not suitable for a reliable phylogenetic reconstruction.[20] In our study, the phylogenetic relationship between the non-vertebrate GW182 and vertebrate ohnologs TNRC6A, TNRC6B and TNRC6C was reconstructed based on conserved multiple-alignment sequence blocks using 4 distinct phylogenetic methods (Materials and Methods). A tree (**Fig. 2**) was

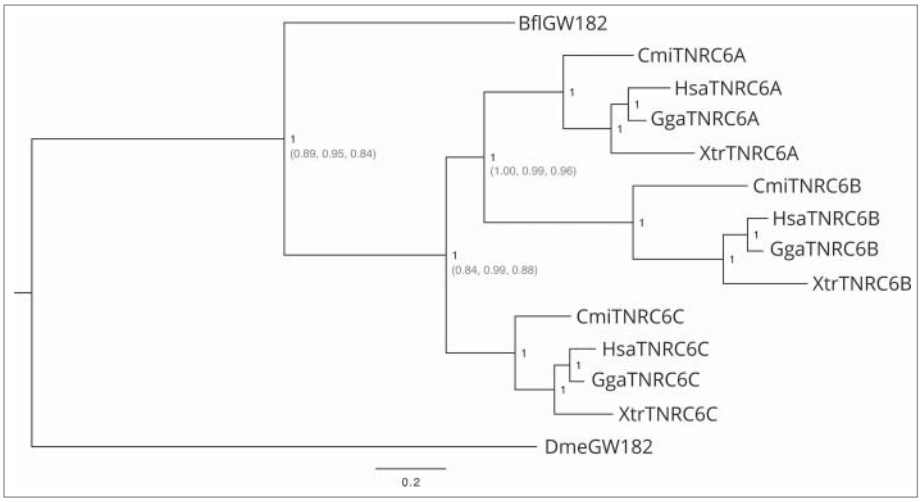

Figure 2. Bayesian inference consensus tree of GW182 paralogs (TNRC6A-C) in vertebrates using *D. melanogaster* and *B. floridae* as the out-group. Bayesian support values are given on all branches; bootstrap values found by ML (PhyML), NJ (Phylip) and MS (Phylip) approaches are in brackets. Species abbreviations: Dme (*Drosophila melanogaster*), Bfl (*Branchiostoma floridae*), Cmi (*Callorhinchus milii*), Gga (*Gallus gallus*), Xtr (*Xenopus tropicalis*) and Hsa (*Homo sapiens*).

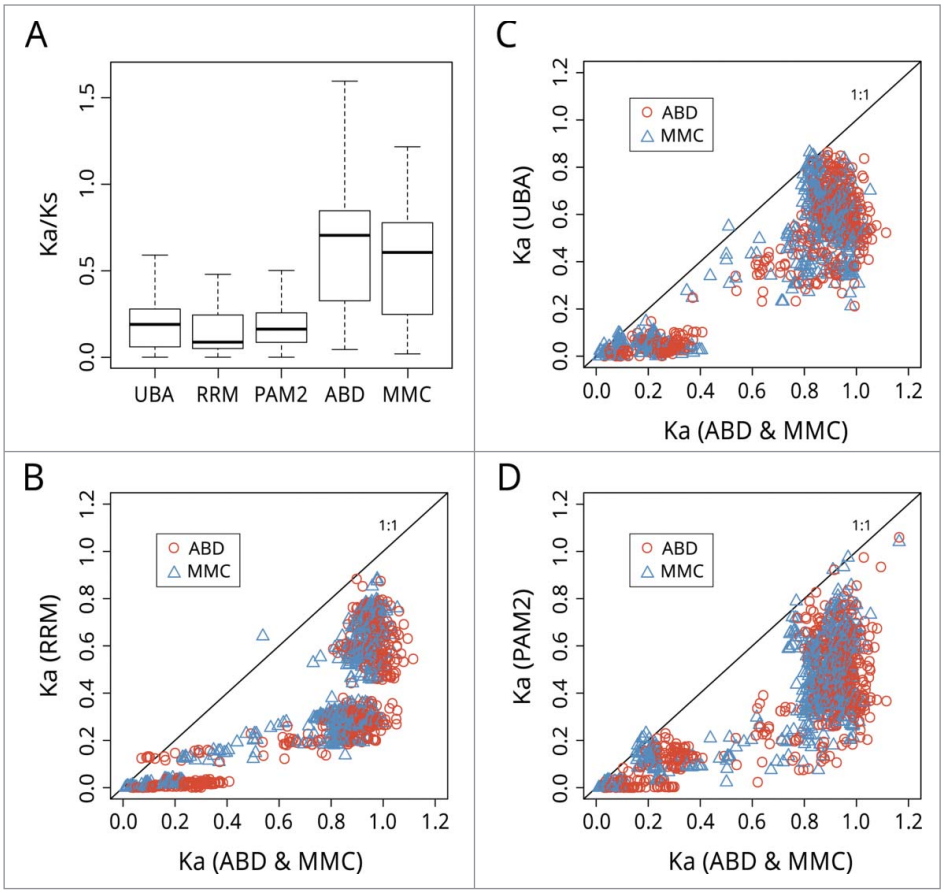

Figure 3. The *Ka/Ks* ratio values for different domains in GW182 proteins. (**A**) The *Ka/Ks* frequency distribution for the UBA, PAM2, RRM, ABD and MMC domains. The *Ka* values of the ABD and MMC domains plotted against those of the (**B**) RRM domain, (**C**) UBA domain, (**D**) and PAM2 motif of the same protein. The line indicates a one-to-one relationship between the *Ka* values of the 2 domains.

constructed from representative taxonomic groups: arthropods (*D. melanogaster*), cephalochordata (*B. floridae*), fishes (*C. milii*), amphibians (*X. tropicalis*), birds (*G. gallus*) and mammals (*H. sapiens*). This tree is supported by all used phylogenetic algorithms and suggests that TNRC6C was the founding member of the vertebrate gene family and represents the ortholog of invertebrate GW182 genes. The phylogram indicates TNRC6A and TNRC6B as second and third subfamilies that diverged from the common stem of the vertebrate GW182 evolutionary tree.

A similar pattern of divergence in phylogenetic reconstruction is exhibited by the subfamilies of GW182 paralogs from mosquitos (**Fig. S3**). Although these subfamilies are the products of independent, lineage-specific local duplication events, we propose to name them by following the scheme of divergence of vertebrate genes. In this way, the group represented by the sequences XP_001854407, XP_001652771, XP_317704, and ETN62198, which includes true orthologs of invertebrate and vertebrate TNRC6C genes, would constitute the GW182C family. Then, following the order of separation, the GW182A group is represented by the sequences XP_001862092, XP_001652770, XP_001237966, and ETN62199, and GW182B contains the sequences XP_001862090, XP_001652769, XP_001689052 and ETN64115 (**Fig. S3**). However, it must be stressed that the striking similarity in the pattern of evolution between vertebrate and mosquito GW182 genes does not imply any functional analogy between the proteins.

Regions essential for RNA silencing evolve more rapidly than other functional parts of the GW182 proteins

The GW182 proteins feature a multi-domain structure (**Fig. 1B**) with a set of well-defined fragments (i.e. RRM, UBA and PAM2) and more variable parts that play essential role in miRNA-dependent post-transcriptional silencing. The non-conserved parts of the proteins include the W-rich regions of the ABD and the M1, M2 and C-term regions (referred to as MMC) of the silencing domain (SD). The modular and multi-functional nature of the GW182 protein suggests that its components may be under different functional or selective constraints. To investigate the variability in selective pressure on different functional fragments of GW182, we calculated the synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates, a widely accepted indicator of selective pressure,[32] for the W-rich regions (ABD and MMC) and the globular domains RRM, UBA and PAM2 motif for each combination of identified orthologous pairs (**Fig. 3A**; **Fig. S4**). The frequency distribution and the mean of the *Ka/Ks* ratios for ABD (mean = 0.63, stdev = 0.33) and MMC (mean = 0.54, stdev = 0.30) are significantly different (Welch Two Sample *t*-test: p < 1e-05, F-test: p < 1e-05) from the values calculated for the RRM (mean = 0.14, stdev = 0.13), UBA (mean = 0.19, stdev = 0.14) and PAM2 (mean = 0.19, stdev = 0.17) domains (**Fig. 3A**; **Fig. S4**). This result indicates that the essential for the miRNA mediated gene silencing W-rich domains (ABD and MMC) are under significantly lower purifying selection and most likely evolve more rapidly than other parts of the GW182 proteins.

To investigate the internal GW182 protein evolution dynamics between the W-rich and remaining 3 conserved domains, we compared the amino acid substitution rates of ABD and MMC with those of UBA, RRM and PAM2 within the same proteins. We found that the dynamics of non-synonymous substitutions (*Ka* values) are generally higher in the domains required for RNA silencing activity (ABD and MMC) than in other parts of

the protein (**Figs. 3B-D**). In all 3 cases, the values follow non-continuous distributions with a gap in the range of 0.4–0.6 for W-rich domain *Ka* values. This gap represents a shift in the sequence change rates of the ABD and MMC domains between vertebrates and other metazoans. Interestingly, a similar discontinuity is apparent in the *Ka* value plot for the RRM domain; the three separated clusters (**Fig. 3B**) represent sequences from basal animals, arthropods and vertebrates and indicate different evolutionary dynamics of the RRM domain among these organisms. No significant differences (χ^2 test: $p = 0.65$) were observed in the *Ka* values of ABD and MMC domains from the same protein (data not shown), which suggests that both domains evolve under similar selective constraints ($r^2 = 0.96$). The relative substitution rate (*Ka/Ks*) exhibited identical but less profound signals (**Figs. S4B–D**), confirming that W-containing domains have undergone a significant shift in the amino acid substitution rate and accumulate sequence changes at a higher frequency than other functional regions of GW182 proteins.

W-rich domains of GW182s show sequence adaptation toward higher specificity to the molecules of the silencing complex

To further explore the mechanisms involved in the evolution of the GW182 proteins, we compared the sequence change dynamics of the W-rich domains involved in RNA silencing with those of their interaction partners: the ABD versus PIWI domain of AGOs and the MMC vs. CNOT1, CNOT9 and C-terminal domain of PAN3 proteins. As shown in **Supplemental Figure 5A and B**, the relative sequence change (*Ka/Ks* ratio) values for ABD and MMC domains form 2 clusters: one with stable values of approximately 0.2 (corresponding to vertebrate sequences) and a second sharply declining over time, approximated here by *Ks* values. The decreasing trend indicates positive selection or relaxation of negative selection on both domains and is slightly more pronounced for the ABD domain. By contrast, the *Ka/Ks* values for GW182 binding partners are consistently low: PIWI (mean = 0.04, stdev = 0.04), CNOT1 (mean = 0.03, stdev = 0.03), CNOT9 (mean = 0.02, stdev = 0.02), and PAN3 (mean = 0.05, stdev = 0.30) (**Fig. S5**). As expected from the conserved function in miRNA-guided gene silencing, these low values imply strong purifying selection acting on these proteins. In addition, the distribution of the *Ka/Ks* frequency in GW182-interacting domains (**Fig. S5E**) is narrow with a sharp peak around the median and mean, confirming a homogeneous negative selective constraint acting on these domains.

The selective pressure (*Ka/Ks* rates) for the ABD and MMC fragments highly deviates from the patterns observed for other parts of the TNRC6C/GW182 proteins (**Fig. 3A**). To investigate the pattern of these differences, we plotted the *Ka/Ks* ratios of these 2 domains using all-versus-all TNRC6C/GW182 ortholog comparisons (**Fig. 4**). The highest substitution rate (0.63–1.53 for ABD and 0.21–1.16 for MMC) occurs among basal metazoans such as sponges (*A. queenslandica*), placoza (*T. adhaerens*) and cnidarians (*N. vectensis, H. vulgaris*), as well as leeches (*H. robusta*), mollusks (*C. gigas, L. gigantea, A. californica*) and worms (*S. kowalevskii*). A slight decrease in the *Ka/Ks* values for the ABD (0.14–0.83) and MMC (0.13–0.74) domains can be observed in arthropods (*D. pulex, T. castaneum, D. melanogaster, A. pisum, N. vitripennis, Z. nevadensis, I. scapularis, S. maritima*) and chordates (*S. kowlevskii, B. floridae*). The dynamics of amino acid change decrease sharply in vertebrate TNRC6Cs for both fragments (0.04–0.26 for ABD and 0.02–0.22 for MMC), suggesting purifying selection mechanisms acting on these 2 functional regions soon after TNRC6C duplication events. Interestingly, the amino acid substitution rate for all tested organisms is on average slightly higher for ABDs than deadenylase complex-binding domains, particularly in basal metazoans.

Among paralogous genes (TNRC6A-C), no significant differences in the *Ka/Ks* ratios for the ABD and MMC domains are observed, indicating that these regions are under homogeneous negative selective constraints in vertebrates. These results are consistent with biochemical studies demonstrating that GW182 paralogs in human associate with all 4 AGOs (AGO1-4) and with a common set of miRNA targets [for a review see ref. 2].

Silencing ability of the GW182 proteins improves along higher position in taxonomic classification

We recently developed a computational sequence-based method[18,33] that allows functional identification and quantitative annotation of highly variable W-containing motifs involved in RNA silencing (ABD and MMC). Application of the W-search annotation tool to the GW182 protein sequences identified in this study yields increasing prediction scores when moving from basal metazoan species toward more complex higher organisms (**Fig. 5**). The graphical presentation of the binding potential of every 20-aa-long motif of the GW182 protein shown in **Figure 5A** (AGO- and CCR4-NOT-binding score denoted by color) reveals very variable binding capability and size of the ABD and MMC/SD domains in invertebrate species. In vertebrates, both domains become evolutionary more stable in terms of size and score, with well-defined AGO-binding and CCR4-NOT-interacting regions. A schematic presentation of the diversity of the binding potential of the ABD and SD domains (**Fig. 5B**) confirms the acquired stability of the ABD domain in vertebrates. However, the SD domain seems to be more variable only when the score values defining binding potential are considered.

Based on the prediction results, we conclude that the ability of GW182 to bind AGO and CCR4-NOT proteins improves with higher position in the metazoan taxonomic classification and increasing complexity of the organism. Such directed changes of the amino acid sequence resemble classical models of adaptive evolution and may suggest that GW182 proteins were under positive selection pressure toward higher binding affinity for molecular partners. This trend in the evolution of GW182 protein function appears to have shifted to purifying selection during vertebrate-specific duplication events. Subsequently, vertebrate GW182 sequences do not exhibit significant sequence and functional variability. The changes in the GW182 protein sequences may therefore serve as an interesting example of evolution caught in action.

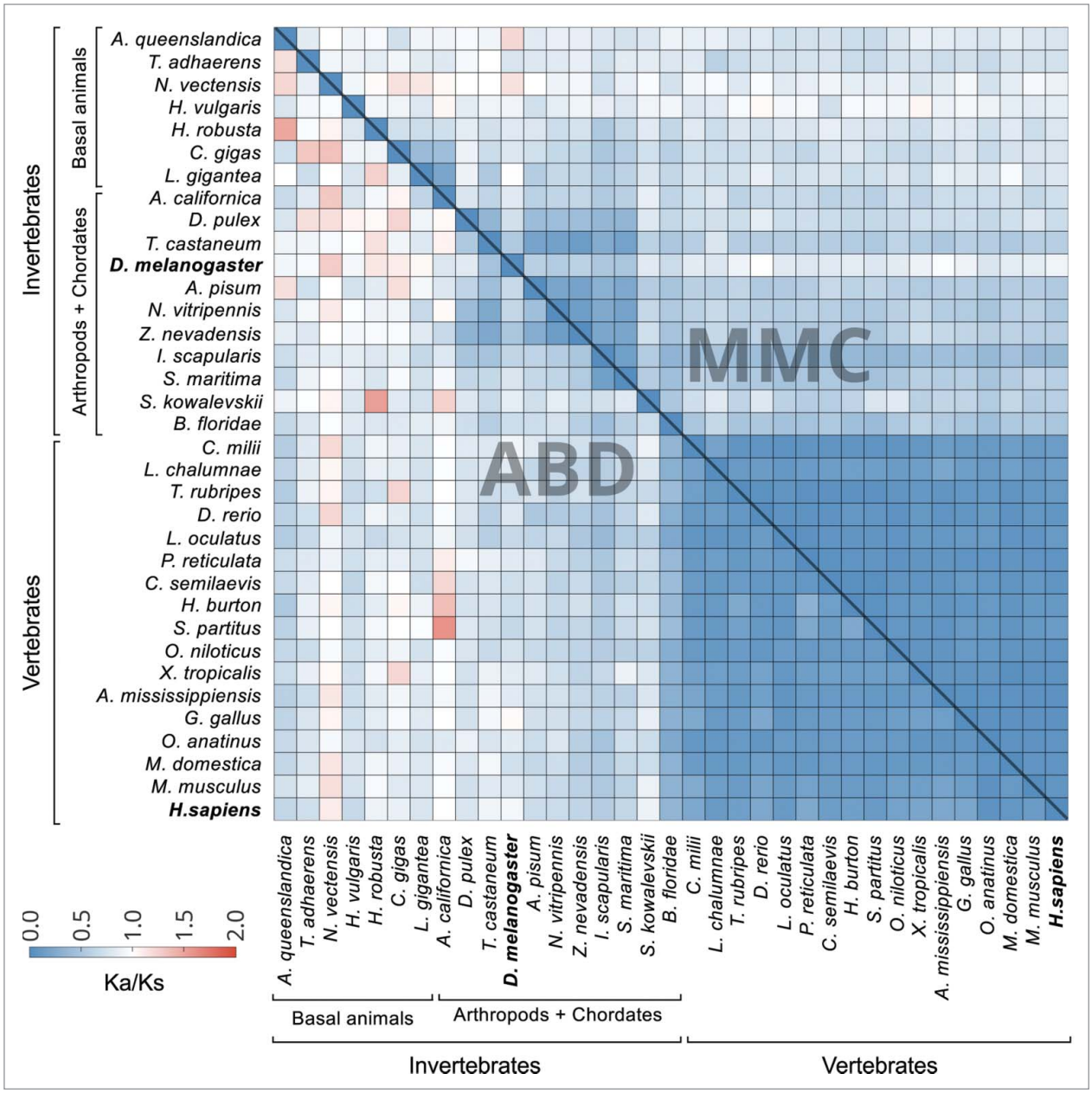

Figure 4. The *Ka/Ks* ratios in all-vs.-all species comparisons. The diagonal line divides the results for the Argonaute-binding domain (ABD) and the W-rich component of the Silencing Domain (MMC).

Discussion

The GW182 proteins are the best-characterized AGO partners in animal cells and have been intensively investigated because of their primary function in RNA silencing.[2,3,4] However, most of this research has been performed using a limited number of sequences from very few organisms: human, fly and worm. In this study, we characterized the full complement of GW182-related genes from the earliest metazoans to human. Our findings address the origin, evolution and functional specialization of this very important for RNA metabolism group of proteins.

In general, the GW182 proteins evolved by speciation forming singleton gene families. The family expanded to include 3 members only in mosquitos and vertebrates. However, the expansion in these 2 groups was independent and a result of distinct mechanisms. In the case of insects, the gene was amplified by tandem duplications. By contrast, in early vertebrates, the multiplication of the GW182 gene family was a result of 2 WGD events. The various mechanisms of GW182 expansion could have influenced the evolutionary dynamics of their family members. It has been reported that ohnologs, defined as paralogs derived from a WGD event, are essential in more biological processes than paralogs, which are a product of

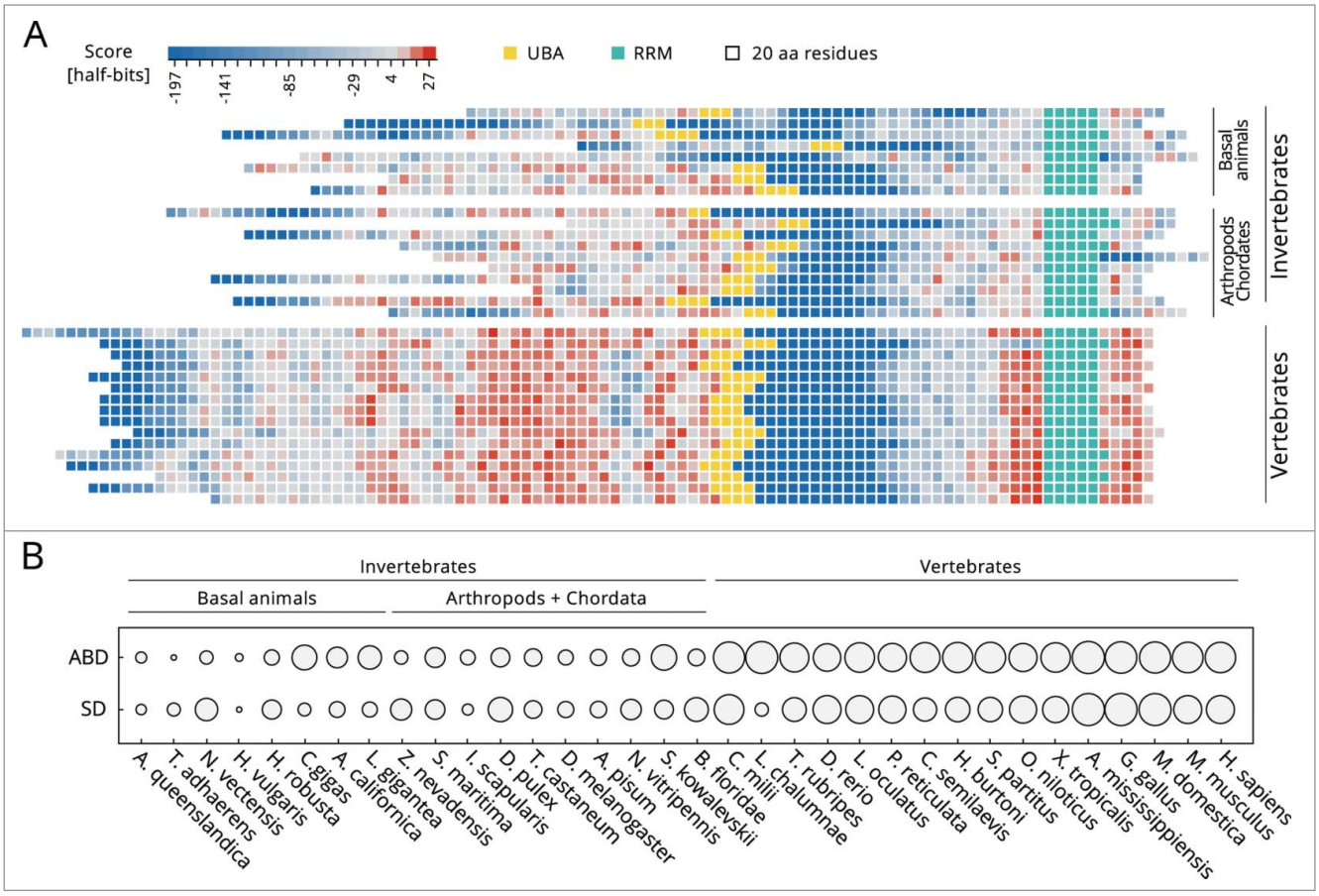

Figure 5. Evaluation of the AGO- and CCR4-NOT-binding capacity of GW182 proteins based on the Wsearch predictions at the level of (**A**) 20-aa-long motifs (color denoting score) and (**B**) domains (circle size denoting score and length).

small-scale duplications.[29,35] In addition, ohnologs likely evolve more slowly than locally duplicated paralogs.[36]

Phylogenetic reconstruction of the identified GW182 homologs allowed us to identify the founding members of the insect and vertebrate families and establish their orthology relationships with all other sequences. We postulate that TNRC6C is the ortholog of the invertebrate genes, including the *D. melanogaster* GW182 protein. Similarly, we have determined the divergence times for the mosquito gene family members and, following the naming schema for vertebrate GW182 proteins, propose appropriate nomenclature (GW182A, GW182B and GW182C). However, whether the presence of GW182 proteins as singletons or 3-member families is a coincidence remains an open question.

In our quest for GW182 protein homologs, one exception must be noted. As reported in previous studies,[16] we could not identify orthologous sequences in *C. elegans* and other nematodes. Instead, the genome of *C. elegans* encodes 2 functional analogs of the GW182 proteins, AIN-1 and AIN-2. However, the absence of canonical GW182 proteins is not an exception; more than half of nematode genes are unique.[37] Nevertheless, using the AGO-binding prediction tool,[18] we identified as many as 10 nematode-specific W-containing proteins, including AIN-1 and AIN-2, that may be involved in RNA silencing. Together, these results suggest that AGO- and CCR4-NOT-binding capabilities evolved independently at least twice in metazoans, confirming its importance in the RNA silencing process.

The domain architecture of the GW182 homologs is, in general, conserved across all investigated organisms. However, the *N. vectensis* DnaJ domain and TNRC6A isoform-encoded N-terminal extension indicate that the variety of possible domain combinations is not restricted to sequences located in the core of GW182 proteins. The glutamine-rich tract found in mammalian TNRC6A proteins merits special attention. Analysis of the W-rich domain-containing proteins involved in RNA silencing and deposited in the Whub portal[18] indicates that such motifs are also present at the C-terminal end of NRPE1 in Arabidopsis as well as in the AIN-1 and AIN-2 proteins of *C. elegans*. The observed conservation between such distantly related organisms strongly suggests the functional importance of these Q-rich tracks in the RNA silencing process.

Our in-depth analysis of the dynamics of evolution of GW182 protein domains allowed a clear distinction between amino acid substitution rates in invertebrates and vertebrates, which is particularly evident in the case of W-rich parts of the Silencing Domain (ABD and MMC). By contrast, the well-

defined motifs (UBA and PAM2) seem to have uniform rates of divergence. However, in the case of the RRM domain, we distinguished 3 groups that were defined by different dynamics of amino acid substitution rates. These groups represent vertebrates, arthropods along with non-vertebrate chordata and basal invertebrates. It seems that evolution follows a different path in the case of the RRM motif compared to other conserved domains. However, the function and importance of the RRM motif in GW182 proteins from fly and human are still not well studied. This domain exhibits no detectable RNA-binding affinity *in vitro* and lacks RNA-binding features.[17] Similarly, the RRM domains identified in the GW182 homologs in this study also lack the characteristic aromatic residues in the 2 conserved sequence signatures RNP1 and RNP2.[38] This suggests that the RRM motif in the ancestral GW182/TNRC6C proteins did not bind RNA and was most likely involved in protein-protein interactions.

The diversity of the AGO-binding motifs appears to be one of the main characteristics of the GW182's ABD domain. For example, in plants,[33] the sequences of AGO-binding proteins (e.g., NRPE1, SPT5, GTB1) exhibit very limited sequence similarity even between closely related species. The multiple tryptophan residues in AGO-binding proteins are embedded in a hyper-variable, low complexity sequences that form locally disordered regions with low overall hydropathy and high net charge. This pattern may reveal an evolutionary compensatory mechanism, in which elevated hydrophobicity brought by tryptophan residues had to be compensated by an increased net charge of small polar amino acids.[18] As a result, the Trp(W)-containing motifs (W-motifs) acquired features characteristic for Short Linear Motif (SLiM) class of sequences that are defined by a few affinity- and specificity-determining residues and short length (from 3 to 11 amino acids).[39] Because of the limited number of binding determinants in such motifs, novel SLiMs can readily evolve *de novo*, adding new functionalities to proteins. The results of our analysis of GW182 in metazoa indicated high evolutionary dynamics of short linear W-motifs in invertebrates. Interestingly, in parallel to plants, the RNAi pathway is a major anti-viral defense system in this group of organisms.[40,41] Viruses, in turn, have evolved a number of adaptations to suppress and evade RNAi, for example, by encoding proteins, often containing SLiMs that mimic and block defense-related protein-protein interactions. For instance, some plant viruses encode proteins that resemble WG/GW motifs and target host AGO family members.[42-44] For viruses infecting invertebrates, suppressor proteins interact with AGOs to inhibit their activity or to induce degradation [45-47]; however, it is not yet clear whether the interaction is mediated by W-motifs. We therefore speculate that the accelerated rate of amino acid change in ABD and MMC domains observed in invertebrate GW182 proteins may, to some extent, reflect the arms race between hosts and pathogens. Accordingly, other studies have indicated that the components of the RNA silencing pathway in invertebrates are among the fastest evolving immune-related genes.[48-50] Our results support the view that the diversification of the W-motifs, which are responsible for AGO/CCR4-NOT recruitment, would be advantageous for adaptive evolution and a successful response to various continuously changing pathogens.

Vertebrate-infecting viruses possess suppressor activity against components of the silencing pathway. The recently described HIV NEF protein binds human AGO2 through its conserved GW motifs. NEF also inhibits the slicing activity of AGO2 and disturbs the sorting of GW182 into exosomes, resulting in the suppression of miRNA-induced silencing.[51] This mechanism, which has not been observed previously in any other virus that infects vertebrates, opens new opportunities for studies of co-evolution between animal GW182 proteins and WG/GW-bearing viral proteins.

Materials and Methods

Identification of GW182 family members

A dual approach was used to identify GW182 proteins. First, the full-length GW182 protein sequences from *H. sapiens* and *D. melanogaster* were used as queries in BLASTp[52] searches against the non-redundant protein databases of UniProtKB[53] and National Center for Biotechnology Information (NCBI). BLAST hits with E-value < 1e-05 were further used in Pfam annotations (version 27)[54] to identify and extract the RRM and UBA protein domains. In the second step, the alignments of both domains were used to build HMM profiles using HMMER3 software [55] and to re-screen the protein databases. The complete set of identified proteins (E-value < 1e-05) was annotated for conserved domains by scanning against the InterPro,[56] SMART,[57] PFAM[54] and PROSITE[58] domain databases. Finally, the results of both approaches were merged to yield the full list of candidate GW182 orthologous proteins, which were further evaluated for synteny in shared domain order.

Phylogenetic analyses

Multiple sequence alignments (MSAs) were generated by the M-COFFEE program,[59] which computes a consensus alignment from several MSA programs (ClustalW, Mafft, PCMA, Dialign, Muscle, Probsons and T-Coffee). The sequence blocks that were in perfect agreement across all alignment programs were used for phylogenetic reconstruction employing 4 independent approaches: Neighbor-Joining (NJ), Maximum Parsimony (MP), Bayesian and Maximum Likelihood (ML). Both NJ and MP analyses with 1,000 bootstraps were performed using PHYLIP v.3.695.[60] ML analysis was conducted using the PhyML v3.0[61] with the LG model recommended by ProtTest v2.2.[62] The frequencies of amino acids were estimated from the data set and statistical support for the different internal branches using the approximate Likelihood-ratio test.[63] Bayesian trees with posterior probabilities were constructed in MrBayes 3.2.2[64,65] with mixed amino acid models (to reduce assumptions prior to analysis), a gamma distribution for rate variation among sites, and a proportion of invariable sites. Two independent runs were launched (4 chains for each run) with one million generations of Markov Chain Monte Carlo (MCMC) analyses sampled every 1000 generations and 25% of trees discarded as burn-in.

Substitution rate estimations

The multiple alignments of the amino acid sequences of the GW182 domains (ABD, UBA, MMC, RRM) and motifs (PAM2) of each member were converted into a codon alignment by PAL2NAL,[66] and the corresponding nucleotide sequences of the domains were then extracted. The ratios of non-synonymous (*Ka*) and synonymous (*Ks*) nucleotide substitutions were calculated for each pair of orthologs using maximum likelihood-based model-averaged methods implemented in KaKs_Calculator.[67] As a complementary approach, the distance-based Nei-Gojobori estimation of *Ka/Ks* was calculated using the yn00 program in the PAML package.[68]

Prediction of ABD and MMC domains

The W-containing domains of GW182 proteins–ABD and MMC–were predicted and scored using the Wsearch program.[18] Wsearch uses PSSM matrices and permits the detection of potentially functional single W-motifs and the determination of their boundaries, as well as statistical quantifications of predicted sequences.[18]

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We would like to thank Maciej Szymanski for helpful discussions and consultations on metazoan taxonomy.

Funding

This work was supported by grants from the National Center of Science (UMO-2011/03/B/NZ2/01416 to W.M.K and UMO-2011/03/N/NZ2/01440 to A.Z) and the KNOW RNA Research Center in Poznan (No. 01/KNOW2/2014). A.Z. was a scholarship holder within the project "Scholarship support for PhD students specializing in majors strategic for Wielkopolska's development," Sub-measure 8.2.2 Human Capital.

Supplemental Material

Supplemental data for this article can be accessed on the publisher's website.

References

1. Eystathioy T, Chan EKL, Tenenbaum SA, Keene JD, Griffith K, Fritzler MJ. A phosphorylated cytoplasmic autoantigen, GW182, associates with a unique population of human mRNAs within novel cytoplasmic speckles. Mol Biol Cell 2002; 13:1338-51; PMID:11950943; http://dx.doi.org/10.1091/mbc.01-11-0544
2. Braun JE, Huntzinger E, Izaurralde E. The role of GW182 proteins in miRNA-mediated gene silencing. Adv Exp Med Biol 2013; 768:147-63; PMID:23224969
3. Fabian MR, Mathonnet G, Sundermeier T, Mathys H, Zipprich JT, Svitkin YV, Rivas F, Jinek M, Wohlschlegel J, Doudna JA, et al. Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell 2009; 35:868-80; PMID:19716330; http://dx.doi.org/10.1016/j.molcel.2009.08.004
4. Zekri L, Huntzinger E, Heimstädt S, Izaurralde E. The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. Mol Cell Biol 2009; 29:6220-31; PMID:19797087; http://dx.doi.org/10.1128/MCB.01081-09
5. Huntzinger E, Braun JE, Heimstädt S, Zekri L, Izaurralde E. Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. EMBO J 2010; 29:4146-60; PMID:21063388; http://dx.doi.org/10.1038/emboj.2010.274
6. Jinek M, Fabian MR, Coyle SM, Sonenberg N, Doudna JA. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. Nat Struct Mol Biol 2010; 17:238-40; PMID:20098421; http://dx.doi.org/10.1038/nsmb.1768
7. Braun JE, Huntzinger E, Fauser M, Izaurralde E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell 2011; 44:120-33; PMID:21981923; http://dx.doi.org/10.1016/j.molcel.2011.09.007
8. Chekulaeva M, Mathys H, Zipprich JT, Attig J, Colic M, Parker R, Filipowicz W. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol 2011; 18:1218-26; PMID:21984184; http://dx.doi.org/10.1038/nsmb.2166
9. Fabian MR, Cieplak MK, Frank F, Morita M, Green J, Srikumar T, Nagar B, Yamamoto T, Raught B, Duchaine TF, et al. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol 2011; 18:1211-7; PMID:21984185; http://dx.doi.org/10.1038/nsmb.2149
10. Christie M, Boland A, Huntzinger E, Weichenrieder O, Izaurralde E. Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol Cell 2013; 51:360-73; PMID:23932717; http://dx.doi.org/10.1016/j.molcel.2013.07.011
11. Chen Y, Boland A, Kuzuoğlu-Öztürk D, Bawankar P, Loh B, Chang C-T, Weichenrieder O, Izaurralde E. A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. Mol Cell 2014; 54:737-50; PMID:24768540; http://dx.doi.org/10.1016/j.molcel.2014.03.034
12. Mathys H, Basquin J, Ozgur S, Czarnocki-Cieciura M, Bonneau F, Aartse A, Dziembowski A, Nowotny M, Conti E, Filipowicz W. Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. Mol Cell 2014; 54:751-65; PMID:24768538; http://dx.doi.org/10.1016/j.molcel.2014.03.036
13. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 2011; 12:99-110; PMID:21245828; http://dx.doi.org/10.1038/nrg2936
14. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem 2010; 79:351-79; PMID:20533884; http://dx.doi.org/10.1146/annurev-biochem-060308-103103
15. Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev 2006; 20:1885-98; PMID:16815998; http://dx.doi.org/10.1101/gad.1424106
16. Eulalio A, Tritschler F, Izaurralde E. The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA 2009; 15:1433-42; PMID:19535464; http://dx.doi.org/10.1261/rna.1703809
17. Eulalio A, Tritschler F, Büttner R, Weichenrieder O, Izaurralde E, Truffault V. The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. Nucleic Acids Res 2009; 37:2974-83; PMID:19295135; http://dx.doi.org/10.1093/nar/gkp173
18. Zielezinski A, Karlowski WM. Integrative data analysis indicates an intrinsic disordered domain character of Argonaute-binding motifs. Bioinformatics 2015; 31:332-9; PMID:25304778; http://dx.doi.org/10.1093/bioinformatics/btu666
19. Till S, Lejeune E, Thermann R, Bortfeld M, Hothorn M, Enderle D, Heinrich C, Hentze MW, Ladurner AG. A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. Nat Struct Mol Biol 2007; 14:897-903; PMID:17891150; http://dx.doi.org/10.1038/nsmb1302
20. Moran Y, Praher D, Fredman D, Technau U. The evolution of microRNA pathway protein components in Cnidaria. Mol Biol Evol 2013; 30:2541-52; PMID:24030553; http://dx.doi.org/10.1093/molbev/mst159
21. Ding L, Spencer A, Morita K, Han M. The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in C. elegans. Mol Cell 2005; 19:437-47; PMID:16109369; http://dx.doi.org/10.1016/j.molcel.2005.07.013
22. Zhang L, Ding L, Cheung TH, Dong M-Q, Chen J, Sewell AK, Liu X, Yates JR, Han M. Systematic identification of C. elegans miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. Mol Cell 2007; 28:598-613; PMID:18042455; http://dx.doi.org/10.1016/j.molcel.2007.09.014
23. Ding XC, Grosshans H. Repression of C. elegans microRNA targets at the initiation level of translation requires GW182 proteins. EMBO J 2009; 28:213-22; PMID:19131968; http://dx.doi.org/10.1038/emboj.2008.275
24. Kuzuoglu-Öztürk D, Huntzinger E, Schmidt S, Izaurralde E. The Caenorhabditis elegans GW182 protein

AIN-1 interacts with PAB-1 and subunits of the PAN2-PAN3 and CCR4-NOT deadenylase complexes. Nucleic Acids Res 2012; 40:5651-65; http://dx.doi.org/10.1093/nar/gks218
25. Qiu X-B, Shao Y-M, Miao S, Wang L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. Cell Mol Life Sci 2006; 63:2560-70; PMID:16952052; http://dx.doi.org/10.1007/s00018-006-6192-6
26. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell 2010; 39:292-9; PMID:20605501; http://dx.doi.org/10.1016/j.molcel.2010.05.015
27. Li S, Lian SL, Moser JJ, Fritzler ML, Fritzler MJ, Satoh M, Chan EKL. Identification of GW182 and its novel isoform TNGW1 as translational repressors in Ago2-mediated silencing. J Cell Sci 2008; 121:4134-44; PMID:19056672; http://dx.doi.org/10.1242/jcs.036905
28. Huminiecki L, Heldin CH. 2R and remodeling of vertebrate signal transduction engine. BMC Biol 2010; 8:146; PMID:21144020; http://dx.doi.org/10.1186/1741-7007-8-146
29. Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc Natl Acad Sci U S A 2010; 107:9270-4; PMID:20439718; http://dx.doi.org/10.1073/pnas.0914697107
30. Makino T, McLysaght A, Kawata M. Genome-wide deserts for copy number variation in vertebrates. Nat Commun 2013; 4:2283; PMID:23917329; http://dx.doi.org/10.1038/ncomms3283
31. Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). BioEssays 2005; 27:937-45; PMID:16108068; http://dx.doi.org/10.1002/bies.20293
32. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 2002; 18:486-7; PMID:12175810; http://dx.doi.org/10.1016/S0168-9525(02)02722-1
33. Karlowski WM, Zielezinski A, Carrère J, Pontier D, Lagrange T, Cooke R. Genome-wide computational identification of WG/GW Argonaute-binding proteins in Arabidopsis. Nucleic Acids Res 2010; 38:4231-45; PMID:20338883; http://dx.doi.org/10.1093/nar/gkq162
34. Fabian MR, Sonenberg N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol 2012; 19:586-93; PMID:22664986; http://dx.doi.org/10.1038/nsmb.2296
35. Makino T, Hokamp K, McLysaght A. The complex relationship of gene duplication and essentiality. Trends Genet 2009; 25:152-5; PMID:19285746; http://dx.doi.org/10.1016/j.tig.2009.03.001
36. Wang Y. Locally duplicated ohnologs evolve faster than nonlocally duplicated ohnologs in Arabidopsis and rice. Genome Biol Evol 2013; 5:362-9; PMID:23362157; http://dx.doi.org/10.1093/gbe/evt016
37. Parkinson J, Mitreva M, Whitton C, Thomson M, Daub J, Martin J, Schmid R, Hall N, Barrell B, Waterston RH, et al. A transcriptomic analysis of the phylum Nematoda. Nat Genet 2004; 36:1259-67; PMID:15543149; http://dx.doi.org/10.1038/ng1472
38. Maris C, Dominguez C, Allain FH-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 2005; 272:2118-31; PMID:15853797; http://dx.doi.org/10.1111/j.1742-4658.2005.04653.x
39. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Krüger D, Grebnev G, Kuban M, et al. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res 2014; 42:D259-66; PMID:24214962; http://dx.doi.org/10.1093/nar/gkt1047
40. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci 2009; 364:99-115; PMID:18926973; http://dx.doi.org/10.1098/rstb.2008.0168
41. Ding S-W. RNA-based antiviral immunity. Nat Rev Immunol 2010; 10:632-44; PMID:20706278; http://dx.doi.org/10.1038/nri2824
42. Azevedo J, Garcia D, Pontier D, Ohnesorge S, Yu A, Garcia S, Braun L, Bergdoll M, Hakimi MA, Lagrange T, et al. Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. Genes Dev 2010; 24:904-15; PMID:20439431; http://dx.doi.org/10.1101/gad.1908710
43. Giner A, Lakatos L, García-Chapa M, López-Moya JJ, Burgyán J. Viral protein inhibits RISC activity by argonaute binding through conserved WG/GW motifs. PLoS Pathog 2010; 6:e1000996; PMID:20657820; http://dx.doi.org/10.1371/journal.ppat.1000996
44. De Ronde D, Pasquier A, Ying S, Butterbach P, Lohuis D, Kormelink R. Analysis of Tomato spotted wilt virus NSs protein indicates the importance of the N-terminal domain for avirulence and RNA silencing suppression. Mol Plant Pathol 2014; 15:185-95; PMID:24103150; http://dx.doi.org/10.1111/mpp.12082
45. Van Rij RP, Saleh M-C, Berry B, Foo C, Houk A, Antoniewski C, Andino R. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev 2006; 20:2985-95; PMID:17079687; http://dx.doi.org/10.1101/gad.1482006
46. Nayak A, Tassetto M, Kunitomi M, Andino R. Intrinsic Immunity. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013.
47. Van Mierlo JT, Overheul GJ, Obadia B, van Cleef KWR, Webster CL, Saleh M-C, Obbard DJ, van Rij RP. Novel Drosophila viruses encode host-specific suppressors of RNAi. PLoS Pathog 2014; 10:e1004256; PMID:25032815; http://dx.doi.org/10.1371/journal.ppat.1004256
48. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol 2006; 16:580-5; PMID:16546082; http://dx.doi.org/10.1016/j.cub.2006.01.065
49. Kolaczkowski B, Hupalo DN, Kern AD. Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol Biol Evol 2011; 28:1033-42; PMID:20971974; http://dx.doi.org/10.1093/molbev/msq284
50. Obbard DJ, Jiggins FM, Bradshaw NJ, Little TJ. Recent and recurrent selective sweeps of the antiviral RNAi gene Argonaute-2 in three species of Drosophila. Mol Biol Evol 2011; 28:1043-56; PMID:20978039; http://dx.doi.org/10.1093/molbev/msq280
51. Aqil M, Naqvi AR, Bano AS, Jameel S. The HIV-1 Nef Protein Binds Argonaute-2 and Functions as a Viral Suppressor of RNA Interference. PLoS One 2013; 8:e74472; PMID:24023945; http://dx.doi.org/10.1371/journal.pone.0074472
52. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-402; PMID:9254694; http://dx.doi.org/10.1093/nar/25.17.3389
53. Consortium U. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 2014; 42:D191-8; PMID:24253303; http://dx.doi.org/10.1093/nar/gkt1140
54. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. Nucleic Acids Res 2014; 42:D222-30; PMID:24288371; http://dx.doi.org/10.1093/nar/gkt1223
55. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol 2011; 7:e1002195; PMID:22039361; http://dx.doi.org/10.1371/journal.pcbi.1002195
56. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, et al. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res 2012; 40:D306-12; PMID:22096229; http://dx.doi.org/10.1093/nar/gkr948
57. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 2012; 40:D302-5; PMID:22053084; http://dx.doi.org/10.1093/nar/gkr931
58. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE. Nucleic Acids Res 2013; 41:D344-7; PMID:23161676; http://dx.doi.org/10.1093/nar/gks1067
59. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res 2006; 34:1692-9; PMID:16556910; http://dx.doi.org/10.1093/nar/gkl091
60. Felsenstein J. PHYLIP -Phylogeny inference package (Version 3.2). Cladistics 1989; 5:164-66.
61. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010; 59:307-21; PMID:20525638; http://dx.doi.org/10.1093/sysbio/syq010
62. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 2005; 21:2104-5; PMID:15647292; http://dx.doi.org/10.1093/bioinformatics/bti263
63. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol 2006; 55:539-52; PMID:16785212; http://dx.doi.org/10.1080/10635150600755453
64. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. Bayesian inference of phylogeny and its impact on evolutionary biology. Science (80-) 2001; 294:2310-4; http://dx.doi.org/10.1126/science.1065889
65. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003; 19:1572-4; PMID:12912839; http://dx.doi.org/10.1093/bioinformatics/btg180
66. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 2006; 34:W609-12; PMID:16845082; http://dx.doi.org/10.1093/nar/gkl315
67. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 2006; 4:259-63; PMID:17531802; http://dx.doi.org/10.1016/S1672-0229(07)60007-2
68. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci CABIOS 1997; 13:555-6.
69. Sidow A. Sequence first. Ask questions later. Cell 2002; 111:13-6; PMID:12372296; http://dx.doi.org/10.1016/S0092-8674(02)01003-6