Crystal structure of the Drosophila Mago nashi–Y14 complex

Hang Shi1,2 and Rui-Ming Xu1,3

1W.M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; 2Department of Physics, University of Massachusetts at Amherst, Amherst, Massachusetts 01002, USA

Pre-mRNA splicing is essential for generating mature mRNA and is also important for subsequent mRNA export and quality control. The splicing history is imprinted on spliced mRNA through the deposition of a splicing-dependent multiprotein complex, the exon junction complex [EJC], at ~20 nucleotides upstream of exon–exon junctions. The EJC is a dynamic structure containing proteins functioning in the nuclear export and nonsense-mediated decay of spliced mRNAs. Mago nashi (Mago) and Y14 are core components of the EJC, and they form a stable heterodimer that strongly associates with spliced mRNA. Here we report a 1.85 Å resolution structure of the Drosophila Mago–Y14 complex. Surprisingly, the structure shows that the canonical RNA-binding surface of the Y14 RNA recognition motif [RRM] is involved in extensive protein–protein interactions with Mago. This unexpected finding provides important insights for understanding the molecular mechanisms of EJC assembly and RRM-mediated protein–protein interactions.

Received January 15, 2003; revised version accepted February 28, 2003.

Pre-mRNA splicing is coupled with subsequent cellular processes including export and nonsense-mediated decay [NMD] of spliced mRNAs [Maquat and Carmichael 2001; Dreyfuss et al. 2002; Reed and Hurt 2002; Wagner and Lykke-Andersen 2002]. Splicing deposits a multiprotein complex, known as the exon junction complex [EJC], on spliced mRNAs at a position ~20 nucleotides upstream of the exon–exon junctions. These EJC proteins have important functions in determining the fate of spliced mRNAs [Kataoka et al. 2000; Le Hir et al. 2000a,b, 2001b; Kim et al. 2001b]. To date, at least eight EJC proteins have been identified. They include Y14, Mago, DEK, RNPS1, Sm160, Upf3, UPAP56, and REF/Aly. Y14 is a putative RNA-binding protein that shuttles between the nucleus and the cytoplasm [Kataoka et al. 2000, Mingot et al. 2001]. It interacts with TAP and other EJC components REF/Aly, RNPS1, and hUpf3 [Kataoka et al. 2001; Kim et al. 2001a]. The two proteins are stably associated with mRNAs during the export from the nucleus to the cytoplasm [Kataoka et al. 2001; Kim et al. 2001b, Le Hir et al. 2001a], and this association persists until the first round of translation [Dostie and Dreyfuss 2002].

A great deal about the function of Mago and Y14 in EJC assembly, mRNA export and decay has been learned from recent studies. However, the molecular mechanisms of the Mago–Y14 interaction and their interactions with mRNA and other EJC components are still poorly understood. To understand the structural basis of Mago–Y14 interaction and the mode of protein–RNA interaction, we have solved a 1.85 Å resolution structure of a Drosophila Mago–Y14 complex. The structure reveals a novel mode of Mago–Y14 interaction through the RRM-binding surface of the Y14 RRM [RNA recognition motif], which has important implications for understanding protein–RNA and protein–protein interactions of the EJC.

Results and Discussion

Recombinant Mago–Y14 complex was produced in Escherichia coli by coexpression. Purified Mago–Y14 complex exists as a 1:1 heterodimer in solution as judged by dynamic laser scattering [data not shown]. Purified Mago–Y14 complex was subjected to endoproteinase Glu-C treatment to produce a core complex containing the full-length Mago and a truncated Y14, Y14Δ43, lacking 46 amino acids at the N terminus and 10 amino acids at the C terminus as revealed by mass spectrometric analyses [data not shown]. The crystal structure of the Mago–Y14 complex was solved by seleno-methionyl [SeMet] multiwavelength anomalous diffraction [MAD; Fig. 1C,D]. The refined structure contains all but three N-terminal residues of Mago and amino acids 61–155 of Y14, encompassing the entire RRM domain. Detailed statistics of the crystallographic analysis are shown in Table 1.
The structure of Y14

Y14 contains a central RRM flanked by highly charged N- and C-terminal regions. As expected, our crystal structure shows that Y14 contains a canonical RRM (amino acids 72–149), which consists of a four-stranded antiparallel β-sheet (β1–β4) and two helices (αA, αB) in a β1–αA–β2–β3–αB–β4 arrangement [Fig. 2A,B]. The Y14 RRM structure can be superimposed with the structure of the N-terminal RRM of U1A (Oubridge et al. 1994) with a root-mean-squared deviation of 0.62 Å using the Cα positions of 23 residues located on β1–β4 for alignment. Most notable differences lie within the loop connecting β2 and β3, and regions outside of the RRM fold. These differences are not unusual, though, as they are the most variable regions among different RRMs (Varani and Nagai 1998). The loop connecting β2 and β3 interacts with RNA in some RRMs, and conformational differences of this loop between native and RNA-bound forms have been documented. The N- and C-terminal extensions to many RRMs are normally disordered in the absence of RNA but become ordered upon nucleic acid binding. In the Mago–Y14 structure, both regions are stabilized by hydrogen bonding between residues located at the two termini. For example, the two termini interact via mainchain hydrogen bonding of Pro 65–Lys 152 and Glu 67–Phe 150. In addition, the packing of Trp 73, Pro 66, His 124, and Trp 148 provides major stabilization of the terminal regions of Y14.

The most conserved regions in the RRM are the RNP2 and RNP1 sequence motifs located on β1 and β3, respectively. In all of the RRM–RNA complex structures known to date, the β-sheet forms an RNA-binding platform, and the aromatic residues in the RNP2 and RNP1 motifs stack with RNA bases [Perez-Canadillas and Varani 2001]. Y14 has a perfectly conserved RNP2 motif, but a highly conserved aromatic residue in RNP1 [most frequently a phenylalanine] is absent. Instead, a leucine [Leu 118] occupies the position [Fig. 1B]. This deviation alone does not appear to render the RRM unable to bind RNA, as exemplified by RNA binding of the third and fourth RRMs of polypyrimidine tract-binding protein (PTB; Conte et al. 2000), in which both of the RRMs lack the highly conserved aromatic residue. Because of the presence of a canonical RRM, it is widely believed that Y14 binds mRNA directly.

Mago–Y14 interactions

Mago and Y14 form a stable heterodimer both in vitro and in vivo, and their association is not affected by RNase A treatment (Kataoka et al. 2001). The two proteins remain associated with each other even at 1M NaCl concentration in our hands. Because Y14 contains an RRM, it is generally believed that the complex binds RNA directly. However, no direct RNA binding of the Mago–Y14 complex or each individual protein has been reported. We also cannot detect any direct RNA binding of the purified Mago–Y14 complex using a gel mobility shift assay, although purified Y14 alone exhibits some RNA-binding activities [data not shown]. The structure shows that Mago and Y14A share an extensive interface [Fig. 2D]. The interaction between Mago and Y14A buries a pairwise accessible surface area of 2376 Å². Mago interacts with Y14 mainly via its two long helices, αA and αC. Surprisingly, the major interaction area on Y14 is the β-sheet surface normally involved in RNA binding in other RRMs [Fig. 3A,B]. This mode of RRM-mediated protein–protein interaction differs from all known examples involving...
Among RNP residues, Ile 74 interacts with Leu 140. Y14 is engaged in interaction with Mago (Fig. 3A, B).

Of the exposed residues in the RNP1 and RNP2 motifs of the protein–protein interface, and it interacts with Ser 136 and Leu 137 of Mago. The C-terminal loop wraps around the N-terminal end of Y14, and extends to an acidic region spanned between Ser 69 and Ser 73.

Table 1. Summary of crystallographic analysis

| Data sets wavelength | Native (0.9 Å) | SeMet MAD |
|----------------------|---------------|-----------|
| Resolution (Å)       | 1.85  | 2.3  | 2.2  | 2.5  |
| Measured reflections | 155,205 | 64,822 | 72,167 | 57,058 |
| Unique reflections   | 23,222 | 11,928 | 13,787 | 10,302 |
| Completeness (%)     | 97.5 (81.4) | 98.2 (99.9) | 98.7 (98.4) | 99.3 (100) |
| Average I/σ          | 18.0  | 22.9  | 20.2  | 20.5  |
| Rmerge<sup>b</sup>   | 0.051 (0.346) | 0.048 (0.143) | 0.054 (0.227) | 0.054 (0.149) |

SOLVE Phasing
- Resolution: 2.8 Å
- Overall figure of merit: 0.65
- Overall Z-score: 37.1

Refinement
- Resolution range (Å): 50–1.85
- R-factor<sup>c</sup>/Rfree: 0.215/0.261 (0.29/0.36)
- Overall Z-score: 37.1
- Overall figure of merit: 0.65
- Resolution range: 50–1.85
- R-factor: 0.051 (0.143) 0.054 (0.227) 0.054 (0.149)

RRMs. For example, the RRM of the spliceosomal protein U2<sup>B</sup> interacts with the leucine-rich-repeat region of U2A<sup>B</sup> via the surface opposite to the RNA-binding β-sheet surface (Price et al. 1998), and the atypical RRM of U2AF65 interacts with a U2AF65 peptide, also via α-helices located on the back of the presumed RNA-binding β-sheet surface (Kielkopf et al. 2001). Thus, the structure of the Mago–Y14 complex reveals a unique protein–protein interaction mode involving the conserved RNP motifs.

A detailed examination of the structure shows that all of the exposed residues in the RNP1 and RNP2 motifs of Y14 are engaged in interaction with Mago (Fig. 3A, B). Among RNP2 residues, Ile 74 interacts with Leu 140 of Mago located at the C-terminal end of αC (Mago and Y14 residues will respectively be labeled with suffixes M and Y hereafter). Phe 76 of Mago is surrounded by Ser 56 of Y14, Val 57 of Y14, and Glu 60 of Y14 located in αA, and Leu 135 of Y14 located in αC. Thr 78 of Y14 is at the edge of the protein–protein interface, and it interacts with Ser 56 of Mago and Glu 60 of Mago via water-mediated and van der Waals interactions. Among RNP1 residues, Lys 114 of Y14 interacts with Asp 129 of Mago via charge interaction. Tyr 116 of Mago interacts with Leu 133 of Y14 and Thr 117 of Y14 via hydrophobic and van der Waals interactions, and Leu 137 of Mago interacts with Lys 143 of Y14 via hydrophobic interaction. In summary, all RNP residues previously known to directly interact with RNA in other RRM motifs are involved in interactions between Mago and Y14.

In addition to the RNP1 and RNP2 residues on β1 and β3, Asn 101 of Y14, His 103 of Y14, Asn 105 of Y14, Arg 108 of Y14, and Arg 109 of Y14, located on β2 or a region immediately C-terminal to β2 of Y14, are involved in extensive interactions with residues located on helix αC of Mago. In contrast, only one β4 residue, Asp 147 of Y14, is within the van der Waals radii of two Mago residues (Fig. 3A). Nevertheless, the region C-terminal to β4 (amino acids 148–155) of Y14 interacts with Mago extensively (Fig. 3A, B). The Y14 C-terminal loop wraps around the N-terminal end of αA and extends to an acidic region spanned between αA and the β-sheet (Fig. 4A, B) of Mago. The interactions include mainchain hydrogen bonding of Cys 149 of Y14 with Asp 148 of Mago, and Stabilizing the conformation of the N-terminal loop of Y14, which is crucial for Ser 69 of Y14, Val 70 of Y14, and Glu 71 of Y14, in this loop to interact with the C-terminal end of helix αC of Mago.

Previous deletion and pull-down experiments showed that the N-terminal region of human Y14 (amino acids 1–73) interacts with Magoh, while no interactions involving the middle (amino acids 74–139) and C-terminal regions (amino acids 140–174) of Y14 were detected (Kataoka et al. 2001). The crystal structure shows that the minimal RRM core structure, from the first residue of β1 to the last residue of β4, encompasses amino acids 73–148. Thus, the absence of interactions between the middle and C-terminal regions of Y14 with Magoh in the GST-pull down experiment is most likely due to disruption of the RRM fold by deletions. The N-terminal frag-
ment of Y14 (amino acids 1–73) may contain an additional Magoh interacting region, as this fragment interacts with Magoh in yeast two-hybrid and GST-pull down experiments (Kataoka et al. 2001). However, the N-terminal 63 residues of Y14 is not present in the structure, and thus we cannot determine whether these residues form additional Magoh interaction regions from a structural standpoint.

Structural implications

Mago and Y14 are present in the fully assembled spliceosome during pre-mRNA splicing and remain stably bound to spliced mRNA (Kataoka et al. 2000, 2001; Reichert et al. 2002). Because Y14 contains an RRM, it is thought that the heterodimer binds RNA via the RRM domain of Y14. However, the crystal structure presented here shows that the presumed RNA-binding surface of Y14 is masked by the protein–protein interactions between Mago and Y14. Therefore, new models of protein–RNA interaction involving the Mago–Y14 complex must be considered. We discuss several possible scenarios of mRNA binding by the Mago–Y14 complex below.

In one scenario, the Mago–Y14 complex may bind to spliced mRNA indirectly. Although Mago and Y14 have been shown to associate with mRNA in numerous studies, no evidence of direct interaction has been documented. In fact, Y14 failed to chemically cross-link to mRNA (Reichert et al. 2002). However, no suitable protein candidates that can mediate the interaction between the Mago–Y14 complex and mRNA have been identified. Most known EJC components cannot serve this role, as they dissociate from spliced mRNA during or soon after mRNA export, whereas Mago and Y14 remain associated with mRNA. RNPS1 and Upf3 appear to associate with mRNA until a late stage, and they both interact with Y14. However, Upf3 is unlikely to mediate the association of Mago and Y14 with mRNA, as it joins the EJC after Mago and Y14 (Reichert et al. 2002). This leaves RNPS1 as the only known candidate that can potentially mediate the interaction between Mago–Y14 and mRNA. It remains to be tested whether RNPS1 or other as yet unidentified EJC proteins can mediate a stable association of Mago and Y14 with mRNA.

In another scenario, the Mago–Y14 complex may directly bind RNA via the RRM of Y14, but a major conformational change of the heterodimer is required to expose the RNA-binding surface of the Y14 RRM. Because the RRM surface interacts strongly with Mago, a great deal of energy is required to open up the heterodimer. The process of pre-mRNA splicing may provide means and energy to open up the heterodimer for binding RNA, which is consistent with the observation that splicing is required for the association of Mago–Y14 with RNA [Kataoka et al. 2001; Le Hir et al. 2001a]. In this scenario, it is possible that the heterodimer can remain associated via the N and C termini of Y14 whereas the RRM β-sheet surface and the helical side of Mago can open like a book to allow access of RNA (Fig. 4C). A related scenario would be that Y14 first binds to mRNA as a monomer, and then Mago joins the complex. However, no evidence for the existence of monomeric Mago or Y14 has been reported. Mago and Y14 appear to be imported into the nucleus as a heterodimer [Mingot et al. 2001], and re-
proteins. The structure shows that the exposed surface of Y14 helix αA, located on the opposite side of the β-sheet surface, has a distinct patch of negatively charged residues (Glu 86Y, Asp 87Y, Glu 88Y, and Glu 91Y; Fig. 4A,B). This negatively charged region of Y14 may interact with positively charged EJC partners. Mago also interacts with in vitro translated TAP in GST-pull down assays. This interaction is most likely direct, as RNase A treatment had no effect on this interaction (Kataoka et al. 2001). The structure of Mago, together with the crystal structure of a TAP–p15 complex (Fribourg et al. 2001), provides an interesting molecular model of Mago–TAP interaction.

The C-terminal half of TAP interacts with the nuclear pore complex and contains an NTF2-like domain. The NTF2-like domain of TAP interacts with p15. Both the TAP NTF2-like domain and p15 contain a six-strand antiparallel β-sheet [Fribourg et al. 2001], and the two proteins interact extensively via their β-sheet surfaces [Fig. 4D]. It is conceivable that Mago may interact with TAP also via its β-sheet surface in a manner similar to that between TAP and p15. The β-sheet surface of Mago is capable of mediating protein–protein interactions, as shown in the packing interaction between two symmetrically related Mago molecules in the crystal lattice [Fig. 4E]. Thus, replacing one Mago β-sheet in Figure 4E with one from TAP would generate an intriguing model of Mago–TAP interaction. In this model, Mago occupies the p15-binding site in TAP, and p15 will not be able to bind TAP. It is possible that Mago can substitute the function of p15 in stabilizing the NTF2-like domain of TAP necessary for interacting with nucleoporin (Braun et al. 2002). Alternative models of Mago–TAP interactions are also possible, and the crystal structure of Mago should be helpful in facilitating biochemical dissection of the molecular basis of Mago–TAP interactions.

Materials and methods

Protein expression, purification, and crystallization

The cDNA fragments encoding full-length Mago and Y14 were amplified by PCR from a Drosophila total cDNA preparation. The Mago cDNA was cloned into a modified pGEX-KG vector to produce GST-fused Mago. The Y14 cDNA was first cloned into a pET28a (Novagen) vector and then transferred into a pMR101 vector (ATCC) to produce polyhistidine-tagged Y14. The two vectors, pGEX-KG and pMR101, have compatible replication and different antibiotic selection markers, and thus are suitable for coexpression of Mago and Y14 in E. coli. Recombinant Mago–Y14 complex was first purified by glutathione-sepharose column. The GST tag and His tag were then removed by in-column thrombin digestion, and the eluted Mago–Y14 complex was further purified by HiTrap-Q and Superdex-75 (Pharmacia) column chromatography. The purified full-length Mago–Y14 complex was then treated with endoproteinase Glu-C overnight at room temperature. Mass spectrometry shows that the Glu-C treatment removes 46 residues at the N terminus and 10 residues at the C terminus of Y14 [Y14Δ] while Mago remains intact. The Mago–Y14A complex was further purified by HiTrap-Q and Superdex-75 column chromatography. Purified Mago–Y14A complex was then concentrated to ~25–30 mg/mL in a buffer containing 10 mM Hepes at pH 8.0, 5% glycerol, and 500 mM NaCl for crystallization. Best diffracting crystals were grown in conditions containing 100 mM sodium acetate at pH 5.8, 30% 2-methyl-2,5-pentanediol (MPD), and 10 mM SrCl2. SeMet substituted Mago–Y14Δ complex was produced using E. coli strain DL4[DE3] in a defined medium containing 30 mg/L of SeMet. Purification and crystallization of SeMet Mago–Y14Δ were similar to that of the native proteins.

Crystallographic analysis

All diffraction data were collected at 100 K using a CCD detector (ADSC) at beamline X26C of National Synchrotron Light Source, Brookhaven National Laboratory. The HKL software package (Otwinowski and Minor 1997) was used for data processing. The Mago–Y14A complex crystallizes in the crystal lattice suggests a model of interaction with TAP.
in space group C2 with cell dimensions of 105.5 Å × 52.4 Å × 52.9 Å and β = 104.9°. The structure was solved by SeMet MAD. Three MAD data sets, with resolutions ranging from 2.2 to 2.5 Å, were collected at wavelengths of 0.9786 Å, 0.9780 Å, and 0.9500 Å, corresponding to inflection point, peak, and remote wavelengths of the Se K edge. A 1.8 Å resolution native data set was collected at a ϕ = 0.85 Å. Detailed data statistics are shown in Table 1. Three ordered Se positions were identified in anomalous and isomorphous Patterson maps, and phasing and solvent flattening were carried out using SOLVE (Terwilliger and Berendzen 1999). Iterative cycles of model building and refinement were carried out using O (Jones et al. 1991), CNS (Brünger et al. 1998), and CCP4 programs. Ten percent of the data was used for cross-validation in CNS. Refinement statistics are shown in Table 1. Figures were prepared using MOLSCRIPT (Kraulis 1991), Raster3D (Merritt and Bacon 1997), RIBBONS (Carson 1997), and GRASP (Nicholls et al. 1991) programs.

The PDB accession code is 1O00.

Acknowledgments

We thank Frances Hannan for Drosophila cDNA, Annie Heroux and Dieter Schneider for technical help during data collection, Michael Myers for help in protein mass analyses, and Michelle Hastings, Lisa Manche, and Zuo Zhang for comments on the manuscript. The work is supported in part by the W.M. Keck Foundation and an NIH grant (GM56874).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Braun, I.C., Herold, A., Rode, M., and Izaualdelle, E. 2002. Nuclear export of RIBBONS (Carson 1997), and GRASP (Nicholls et al. 1991) programs.

Refinement statistics are shown in Table 1. Three ordered Se positions were identified in Table 1. Figures were prepared using O (Jones et al. 1991), CNS (Brünger et al. 1998), and CCP4 programs. Ten percent of the data was used for cross-validation in CNS. Refinement statistics are shown in Table 1. Figures were prepared using MOLSCRIPT (Kraulis 1991), Raster3D (Merritt and Bacon 1997), RIBBONS (Carson 1997), and GRASP (Nicholls et al. 1991) programs.

The PDB accession code is 1O00.

Acknowledgments

We thank Frances Hannan for Drosophila cDNA, Annie Heroux and Dieter Schneider for technical help during data collection, Michael Myers for help in protein mass analyses, and Michelle Hastings, Lisa Manche, and Zuo Zhang for comments on the manuscript. The work is supported in part by the W.M. Keck Foundation and an NIH grant (GM56874).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Braun, I.C., Herold, A., Rode, M., and Izaualdelle, E. 2002. Nuclear export of RIBBONS (Carson 1997), and GRASP (Nicholls et al. 1991) programs.

Refinement statistics are shown in Table 1. Three ordered Se positions were identified in Table 1. Figures were prepared using O (Jones et al. 1991), CNS (Brünger et al. 1998), and CCP4 programs. Ten percent of the data was used for cross-validation in CNS. Refinement statistics are shown in Table 1. Figures were prepared using MOLSCRIPT (Kraulis 1991), Raster3D (Merritt and Bacon 1997), RIBBONS (Carson 1997), and GRASP (Nicholls et al. 1991) programs.

The PDB accession code is 1O00.
Crystal structure of the *Drosophila* Mago nashi –Y14 complex

Hang Shi and Rui-Ming Xu

*Genes Dev.* 2003. 17:
Access the most recent version at doi:10.1101/gad.260403

---

**References**
This article cites 39 articles, 16 of which can be accessed free at: [http://genesdev.cshlp.org/content/17/8/971.full.html#ref-list-1](http://genesdev.cshlp.org/content/17/8/971.full.html#ref-list-1)

**License**

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](http://genesdev.cshlp.org/content/17/8/971.full.html#ref-list-1).