

# Structural basis for the activation of the *C. elegans* noncanonical cytoplasmic poly(A)-polymerase GLD-2 by GLD-3

Katharina Nakel<sup>a</sup>, Fabien Bonneau<sup>a</sup>, Christian R. Eckmann<sup>b,1</sup>, and Elena Conti<sup>a,2</sup>

<sup>a</sup>Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany; and <sup>b</sup>Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, D-01307 Dresden, Germany

Edited by Jennifer A. Doudna, University of California, Berkeley, CA, and approved June 10, 2015 (received for review March 7, 2015)

**The *Caenorhabditis elegans* germ-line development defective (GLD)-2–GLD-3 complex up-regulates the expression of genes required for meiotic progression. GLD-2–GLD-3 acts by extending the short poly(A) tail of germ-line-specific mRNAs, switching them from a dormant state into a translationally active state. GLD-2 is a cytoplasmic noncanonical poly(A) polymerase that lacks the RNA-binding domain typical of the canonical nuclear poly(A)-polymerase Pap1. The activity of *C. elegans* GLD-2 in vivo and in vitro depends on its association with the multi-K homology (KH) domain-containing protein, GLD-3, a homolog of Bicaudal-C. We have identified a minimal polyadenylation complex that includes the conserved nucleotidyl-transferase core of GLD-2 and the N-terminal domain of GLD-3, and determined its structure at 2.3-Å resolution. The structure shows that the N-terminal domain of GLD-3 does not fold into the predicted KH domain but wraps around the catalytic domain of GLD-2. The picture that emerges from the structural and biochemical data are that GLD-3 activates GLD-2 both indirectly by stabilizing the enzyme and directly by contributing positively charged residues near the RNA-binding cleft. The RNA-binding cleft of GLD-2 has distinct structural features compared with the poly(A)-polymerases Pap1 and Trf4. Consistently, GLD-2 has distinct biochemical properties: It displays unusual specificity in vitro for single-stranded RNAs with at least one adenosine at the 3' end. GLD-2 thus appears to have evolved specialized nucleotidyl-transferase properties that match the 3' end features of dormant cytoplasmic mRNAs.**

posttranscriptional regulation | cytoplasmic poly(A)-polymerase | noncanonical poly(A)-polymerase | germ-line maturation | cytoplasmic polyadenylation

The poly(A) tail is a major regulatory determinant of eukaryotic gene expression. This string of nontemplated adenosines is added to the 3' end of the vast majority of eukaryotic mRNAs upon transcription termination by the canonical nuclear poly(A) polymerase (Pap1) (reviewed in ref. 1). The presence of an intact poly(A) tail is required for nuclear export and for cytoplasmic translation (reviewed in ref. 2). Conversely, shortening of the poly(A) tail is connected to translational repression and mRNA decay (reviewed in refs. 3–5). In metazoans, the short poly(A) tail of translationally repressed mRNAs can also be reextended by cytoplasmic noncanonical poly(A) polymerases, initiating the synthesis of the corresponding gene products (reviewed in ref. 1). This mechanism of translational regulation allows rapid protein production in physiological contexts where transcription is silenced (e.g., in oocytes and early embryos) or at a significant physical distance from the translation machinery (e.g., in neuronal dendrites) (reviewed in refs. 6–9).

The cytoplasmic poly(A) polymerase germ-line development defective 2 (GLD-2) was originally discovered in a screen for *Caenorhabditis elegans* mutants causing ectopic germ-line proliferation (10) and has since been studied in several vertebrate and invertebrate model organisms (9, 11–16). In the hermaphrodite germ line of this nematode, GLD-2 is envisioned to activate the translation of a set of mRNAs required for the transition from mitosis to meiosis and has been shown to promote mRNA

stability (10, 17). In *Drosophila*, *Xenopus*, and mice, GLD-2 orthologs are involved in the translation of maternal mRNAs during oocyte maturation and egg activation (11, 13, 15, 18). GLD-2 is highly expressed in the hippocampus in mammals, and in *Drosophila*, it activates the translation of dormant synaptic mRNAs and the formation of long-term memory (13, 19, 20). In mammals and *Drosophila*, GLD-2 orthologs have also been shown to add an adenosine tail to the 3' end of microRNAs (21–23). No ortholog of GLD-2 exists in yeast. However, related noncanonical nucleotidyl transferases are present in the *Saccharomyces cerevisiae* nucleus (Trf4/Trf5) and in the *Schizosaccharomyces pombe* cytoplasm (Cid1), where they extend the 3' end of RNAs, prompting their degradation (reviewed in refs. 24–26).

Canonical and noncanonical nucleotidyl transferases contain a similar enzymatic core composed of the catalytic and the so-called central domains, which act in concert to transfer the incoming nucleotide to the 3' end of an RNA substrate (reviewed in ref. 25). Canonical nucleotidyl transferases, like yeast Pap1, also contain an RNA recognition motif (RRM) that is key for RNA binding and activity (27, 28), but no such domain is present in the sequence of GLD-2, Trf4/Trf5, or Cid1. Different nucleotidyl transferases vary in the selection of the incoming nucleotide (ATP in the case of Pap1, GLD-2, and Trf4/Trf5, and UTP

## Significance

The efficient translation of the vast majority of eukaryotic mRNAs requires the presence of a poly(A) tail. Although the poly(A) tail was originally thought of as a stable modification, it is now clear that it is much more dynamic. Poly(A) tails can be shortened by deadenylases, leading to mRNA decay or translational repression. The short poly(A) tails of translationally inactive mRNAs can also be reextended by cytoplasmic noncanonical poly(A) polymerases, activating their translation. This mechanism of translational control is found predominantly in metazoan oocytes and neurons and is mediated by germ-line development defective (GLD)-2. Here, we report the molecular mechanism with which *Caenorhabditis elegans* GLD-2 is activated by GLD-3, a homologue of Bicaudal-C, and identify the unusual substrate specificity of this class of noncanonical poly(A) polymerases.

Author contributions: K.N. and E.C. designed research; K.N. and F.B. performed research; K.N., F.B., and C.R.E. analyzed data; and K.N. and E.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID code 4ZRL).

<sup>1</sup>Present address: BioCenter, Department of Genetics, Martin-Luther-University of Halle-Wittenberg, Weinbergweg 22, 06120 Halle (Saale), Germany.

<sup>2</sup>To whom correspondence should be addressed. Email: [conti@biochem.mpg.de](mailto:conti@biochem.mpg.de).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504648112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504648112/-DCSupplemental).

in the case of Cid1) and in the number of consecutive reactions they perform on a given substrate (reviewed in refs. 24–26). *C. elegans* GLD-2 has weak activity in isolation, but is converted into an active poly(A) polymerase upon binding to GLD-3, a nematode protein with similarity to Bicaudal-C (Bic-C) (10, 29). In addition to GLD-3, several other GLD-2-interacting proteins have been identified in various species (e.g., CPEB, RNP-8, Rbm9, Musashi, and Ago-1) and typically contain domains found in RNA-binding proteins (11, 30–32). In the case of GLD-3, the GLD-2-binding region features a pattern of hydrophobic residues that matches the signature sequence of K homology (KH) domains (33, 34). These observations have led to the concept that GLD-2-binding partners might endow the non-canonical poly(A) polymerase with an RNA-binding fold in the context of the heterodimer (10). The conundrum, however, is that the closest homolog of GLD-2, Cid1, is an efficient nucleotidyl transferase even in the absence of additional proteins. To obtain insights into how GLD-2 functions at the molecular level, we set out to dissect the molecular mechanisms with which GLD-3 stimulates the poly(A)-polymerase activity of GLD-2.

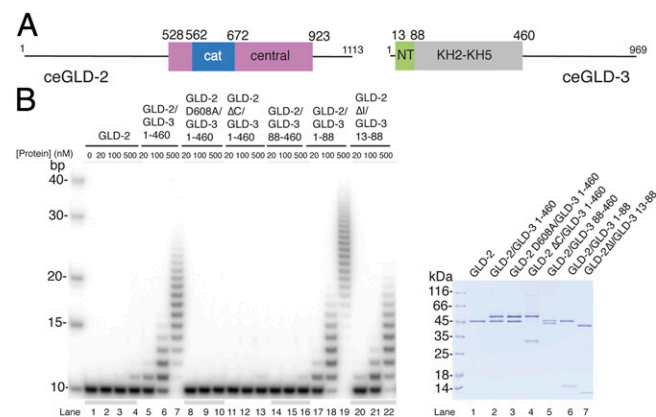
## Results and Discussion

**Identification of a Minimal poly(A) Polymerase Core of *C. elegans* GLD-2–GLD-3.** In *C. elegans* (*C.e.*) GLD-2 (1,113 residues), N- and C-terminal low-complexity sequences flank the catalytic and central domains of the nucleotidyl-transferase region (10) (Fig. 1A). *C.e.* GLD-3 (969 residues) is predicted to contain 5 KH domains followed by a C-terminal low-complexity sequence (29, 33, 34). In a previous study, we have shown that a proteolytically resistant core of GLD-3 folds into a compact globular unit formed by the KH2–KH5 domains (residues 110–460) and a C-terminal segment of the putative KH1 domain (residues 88–110) (33, 34) (Fig. 1A). An N-terminal segment of the putative KH1 domain (residues 33–81) has been shown to interact with the GLD-2 nucleotidyl-transferase region by yeast two-hybrid assays (29, 33, 34). To map the minimal interacting regions experimentally, we purified two different complexes consisting of either GLD-2<sub>528–1042</sub> and GLD-3<sub>1–460</sub> (as described in ref. 34) or GLD-2<sub>528–808</sub> and GLD-3<sub>1–106</sub> and subjected

them to limited proteolysis experiments with different proteases (elastase and GluC). Using mass spectrometry and sequence analyses, we mapped the domain boundaries of a smaller core complex to GLD-2<sub>528–923</sub> and GLD-3<sub>1–88</sub>, and also identified the presence of proteolytically sensitive regions at GLD-3 residues 1–12 and at GLD-2 residues 814–856 (Fig. S1A). These residues of GLD-2 correspond to an insertion that is present in *C. elegans* but not in other species (see below).

We tested the activity of the complex in *in vitro* polyadenylation assays by using a 10-mer poly(A) RNA substrate. When in isolation, the activity of GLD-2<sub>528–923</sub> (hereafter defined as GLD-2) was negligible (Fig. 1B, lanes 2–4). Addition of GLD-3<sub>1–460</sub> resulted in robust polyadenylation (Fig. 1B, lanes 5–7). As control, the GLD-2–GLD-3<sub>1–460</sub> activity was abrogated by introducing a single substitution of one of the aspartic acids in the catalytic site of GLD-2 (Asp608Ala) (Fig. 1B, lanes 8–10). Deletion of residues 815–923 in the central domain (GLD-2<sub>ΔC</sub>) did not impair the interaction with GLD-3 (Fig. S1B) but impaired the polyadenylation properties of the complex (Fig. 1B, lanes 11–13). Incubating GLD-2 with a GLD-3 protein consisting only of the N-terminal GLD-2-interacting region (GLD-3<sub>1–88</sub>) was sufficient to restore robust polyadenylation (Fig. 1B, lanes 17–19). Consistently, a GLD-3 protein where the N-terminal GLD-2-interacting region had been deleted (GLD-3<sub>88–460</sub>) was unable to increase polyadenylation above the level detected for GLD-2 in isolation (Fig. 1B, lanes 14–16). We concluded that GLD-2–GLD-3<sub>1–88</sub> is a catalytically active core of the polyadenylation complex.

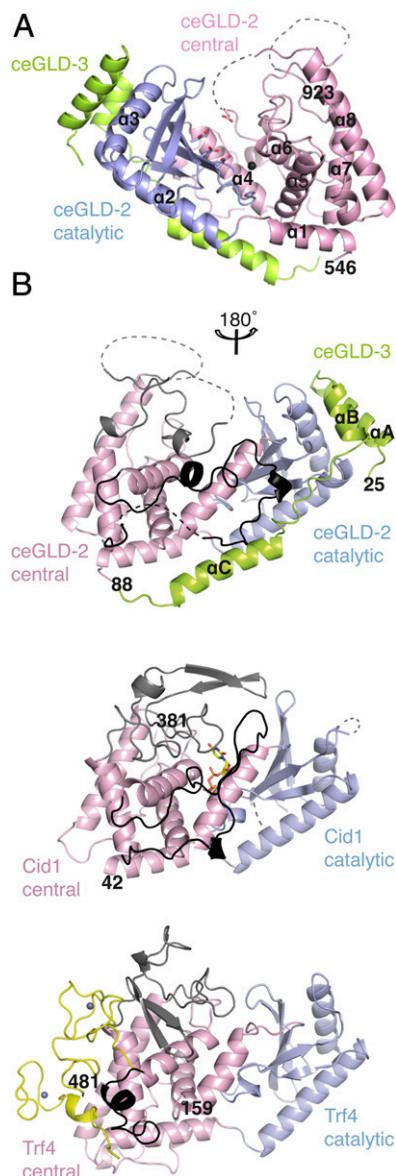
Attempts to crystallize the GLD-2–GLD-3<sub>1–88</sub> complex failed. To obtain crystals, we trimmed the proteolytically sensitive regions further and formed a complex between GLD-3<sub>13–88</sub> (hereby defined as GLD-3<sub>NT</sub>) and a GLD-2 protein where the insertion between residues 813–846 had been deleted (GLD-2<sub>ΔI</sub>). This complex was capable of polyadenylation activity, albeit to a slightly lower extent than GLD-2–GLD-3<sub>1–460</sub> (Fig. 1B, lanes 20–22). The structure of GLD-2<sub>ΔI</sub>–GLD-3<sub>NT</sub> was solved by using selenomethionine-based single anomalous dispersion (SAD) and refined at 2.3-Å resolution to  $R_{\text{free}}$  of 23.0%,  $R$  factor of 18.7%, and good stereochemistry (Table S1). The final model includes residues 546–923 of GLD-2<sub>ΔI</sub> (with the exception of disordered loops between residues 766 and 773, 804 and 812, 847 and 854, and 876 and 881), residues 25–88 of GLD-3<sub>NT</sub>, and a chloride ion.



**Fig. 1.** Poly(A)-polymerase core of the GLD-2–GLD-3 complex. (A) Schematic domain organization of *C.e.* GLD-2 and GLD-3. Folded domains are shown in rectangles and low-complexity sequences as lines. The portions of the molecule included in the structure reported in this work are in blue and pink [for the catalytic (cat) and central domains of GLD-2] and in green (the N-terminal NT domain of GLD-3). In gray is the folded region formed by the KH2–KH5 domains (34). (B) Poly(A)-polymerase assays with different *C.e.* GLD-2 and GLD-3 protein fragments and mutants. In the assay, 20, 100, or 500 nM for proteins were incubated with 100 nM 5'-<sup>32</sup>P-labeled A<sub>10</sub> RNA and 0.5 mM ATP. Reactions were run on a 10% polyacrylamide/7 M urea gel and visualized by phosphorimaging. The purified proteins used in the assay are shown at Right in a Coomassie-stained 12% SDS/PAGE gel.

**GLD-2 Features Large Insertions in the poly(A) Polymerase Core.** The structure of GLD-2<sub>ΔI</sub> shows the characteristic architecture of an RNA-dependent nucleotidyl transferase (Fig. 2A) (25, 28). Briefly, the catalytic domain is centered at a five-stranded β-sheet flanked by two α-helices (α2 and α3). Helix α4 connects the catalytic domain to the central domain. The central domain is an α-helical bundle formed by two noncontiguous stretches (helix α1 and helices α5–α8, Figs. 1A and 2A). The two domains of GLD-2<sub>ΔI</sub> are separated by a pronounced cleft. The sidewalls of the cleft are lined by the β-sheet of the catalytic domain and by helices α5–α6 of the central domain, whereas helix α4 lines the bottom. The cleft harbors the active site: Strands β2 and β5 of the catalytic domain provide the three conserved aspartic acids that mediate the chemical reaction characteristic of nucleotidyl transferases (25, 35) (Fig. 2A).

Comparisons with previously determined structures in the Protein Data Bank (PDB) using the program Dali (36) shows that GLD-2<sub>ΔI</sub> resembles most closely the *S. pombe* poly(U)-polymerase Cid1 [root mean square deviation (rmsd) of 1.7 Å over 75% of all Cα atoms; 30% sequence identity] and to a lesser extent the *S. cerevisiae* poly(A)-polymerase Trf4 (rmsd of 3.0 Å over 70% of all Cα atoms; 21% sequence identity) (37–40). All three noncanonical nucleotidyl transferases were crystallized in the absence of RNA and show a similar open conformation of the catalytic and central domains (Fig. 2B). In the structure of the canonical poly(A)-polymerase Pap1 determined with ATP



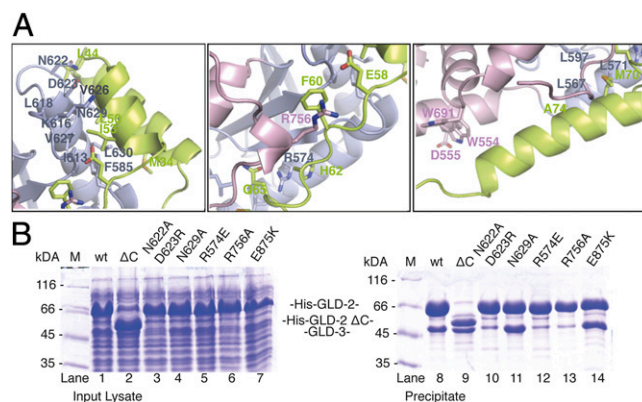
**Fig. 2.** Structure of the GLD-2-GLD-3 core complex. (A) The structure of GLD-2 $\Delta$ I-GLD-3 $_{NT}$  is shown in cartoon representation, colored as in the schematics in Fig. 1A. The N- and C-terminal residues and the  $\alpha$ -helices of GLD-2 are indicated. Disordered regions are shown with dotted lines. The region deleted for crystallization (residues 813–846) resides in the disordered loop region between  $\alpha$ -helices 7 and 8. In stick representations are the three active-site aspartic acids and the Glu875 residue identified in the *C. elegans* *gld-2* mutant allele *h292* (42). The chloride ion is shown as a black sphere. This and all structure figures were made with the PyMOL program (57). (B) *Upper* shows a cartoon representation of GLD-2 $\Delta$ I-GLD-3 $_{NT}$ , oriented with a 180° rotation around a vertical axis with respect to the view in A. The N- and C-terminal residues and the  $\alpha$ -helices of GLD-3 are labeled. The  $\alpha$ 6- $\alpha$ 7 and  $\alpha$ 7- $\alpha$ 8 insertions of the central domain are highlighted in black and gray, respectively. *Center* and *Lower* show the structures of Cid1 [PDB ID code 4E80 (39); UTP molecule shown as stick representation] and Trf4-Air2 [PDB ID code 3NYB (37); Air2 in yellow, Zn ions shown as gray spheres] with the same color coding and in the same orientation as GLD-2 $\Delta$ I-GLD-3 $_{NT}$  in *Upper*, after optimal superposition. N- and C-terminal residues are labeled. Disordered loop regions are highlighted with dotted lines.

and an A<sub>5</sub> RNA, the interdomain cleft closes upon substrate binding (27) (Fig. S24).

The major differences in the structures of different nucleotidyl transferases lie in the central domain (Fig. 2B and Fig. S3).

GLD-2 features particularly long and convoluted segments in the linkers connecting helix  $\alpha$ 6 to  $\alpha$ 7 and helix  $\alpha$ 7 to  $\alpha$ 8 (Fig. 2B, *Upper*). The  $\alpha$ 7- $\alpha$ 8 linker spans 90 residues (794–886), and includes the so-called nucleotide recognition motif (NRM, residues 871–881) and the *C. elegans*-specific insertion that was truncated in GLD-2 $\Delta$ I. The  $\alpha$ 7- $\alpha$ 8 loop interacts with helices  $\alpha$ 4 and  $\alpha$ 6 and contributes to line the entry of the interdomain cleft (suggesting the NRM and possibly part of the truncated insertion might contribute to substrate binding). The  $\alpha$ 6- $\alpha$ 7 linker of GLD-2 $\Delta$ I is 50 residues long (732–781). It zig-zags around helix  $\alpha$ 4 and contributes to form the binding site for GLD-3 $_{NT}$ . In comparison, the  $\alpha$ 6- $\alpha$ 7 and  $\alpha$ 7- $\alpha$ 8 linkers of Trf4 are shorter and both form the binding site for the zinc knuckle protein Air2 (37). A remarkable difference between the two noncanonical poly(A) polymerases is that they recognize binding partners on diametrically opposite surfaces: Air2 binds exclusively at the central domain of Trf4 (37), whereas GLD-3 binds mainly at the catalytic domain of GLD-2 (Fig. 2B).

**The N Terminus of GLD-3 Wraps Around the GLD-2 poly(A)-Polymerase Domain.** GLD-3 $_{NT}$  folds into three helices (referred to as  $\alpha$ A,  $\alpha$ B, and  $\alpha$ C) (Fig. 2A) and does not resemble a KH domain fold. GLD-3 $_{NT}$  stretches on the surface of GLD-2, extending for more than 90 Å and burying a surface area of approximately 1,900 Å<sup>2</sup> (calculated for GLD-2 by using the PISA program; refs. 27 and 41). Indeed, in isolation, GLD-3 $_{NT}$  appears to be predominantly unstructured, as judged by circular dichroism on GLD-3 $_{1-110}$  (Fig. S44). The catalytic domain of GLD-2 mediates more than 70% of the total GLD-3- $\alpha$  interaction surface. A helix-turn-helix formed by the  $\alpha$ A and  $\alpha$ B helices of GLD-3 $_{NT}$  (residues 25–57) binds the GLD-2 catalytic domain at helix  $\alpha$ 3 and at the preceding loop (Fig. 3A, *Left*). Here, a hydrophobic surface of GLD-2 (Phe585, Ile613, Leu618, Val626, Val627, Leu630) recognizes apolar residues of GLD3. The interaction is reinforced by polar contacts (via GLD-2 Asn622, Asp623, Lys616, Asn629). GLD-3 $_{NT}$  then continues with an extended segment (residue 58–66), which packs on the side of the GLD-2  $\beta$ -sheet and wedges in between helix  $\alpha$ 4 and the  $\alpha$ 6- $\alpha$ 7 insertion (Fig. 3A, *Middle*). Here, the interaction is dominated by two arginine residues of GLD-2 (Arg756 and Arg757). Finally, the last helix of GLD-3 $_{NT}$  (residues 67–83) binds a composite surface of GLD-2 formed by the catalytic and central domains. The GLD-3 interaction is mediated both by hydrophobic contacts (with GLD-2



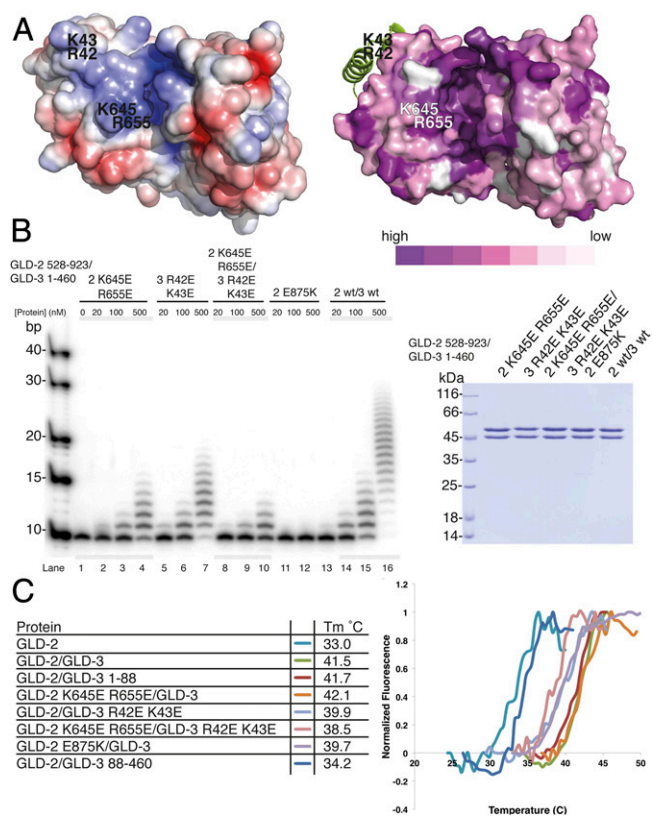
**Fig. 3.** Hydrophobic and polar interactions anchor GLD-3 on the GLD-2 surface. (A) Three zoom-in views showing a representative set of interactions between GLD-2 $\Delta$ I and GLD-3 $_{NT}$ . The molecules are oriented as in Fig. 2B. (B) Coomassie-stained 12% SDS/PAGE of His-pull-down experiments of coexpressed wild-type (wt) and mutant GLD-2 $_{528-923}$  with GLD-3 $_{1-460}$ . Proteins were coexpressed in 4 mL of TB overnight at 20 °C. Pull-down assays were carried out by using a Tecan robot, using 50  $\mu$ L of total lysate and 200  $\mu$ L of Ni-NTA resin, and a buffer containing 250 mM NaCl. Total lysate control is shown at *Left*; pulled-down protein precipitate is shown at *Right*.

Leu567, Leu571, Leu597, Trp554, and Trp691) and polar contacts with GLD-2 Asp555 (Fig. 3A, *Right*). The portion of the central domain spanning the  $\alpha 7$ - $\alpha 8$  linker and helix  $\alpha 8$  is not involved in GLD-3 binding, rationalizing why its deletion in the GLD-2 $\Delta C$  mutant still supported the interaction (Fig. 3B, lanes 2 and 9 and Fig. S1B). GLD-3, which is highly conserved in all *Caenorhabditis* species (Fig. S4B), shares no obvious sequence similarity with other GLD-2-binding factors (e.g., CPEB, RNP-8, Rbm9, Musashi, and Ago1) (11, 30–32) and the similarity with Bic-C is too low to draw conclusions based on the GLD-3 $_{NT}$  interactions. However, a subset of residues lining the GLD-3-binding surface of GLD-2 is conserved across species (Fig. 3A and Figs. S3 and S4C). GLD-2 orthologs might therefore use similar surface hotspots to mediate macromolecular interactions.

The interpretation of previous yeast-two hybrid results (33) in light of these structural data suggests that the helix-turn-helix of GLD-3 $_{NT}$  is essential for GLD-2 binding. Based on the structural analysis, we engineered specific mutations in GLD-2 and tested their ability to interact with GLD-3 $_{1-460}$  in coexpression and pull-down assays. Binding was impaired with the GLD-2 N622A and D623R mutant, which is expected to affect the interaction with the turn between the GLD-3 helices  $\alpha A$  and  $\alpha B$  (Fig. 3B, lanes 3 and 10). Interfering with a single polar contact between GLD-2 (N629A) and GLD-3 helix  $\alpha A$  did not noticeably affect complex formation (Fig. 3B, lanes 4 and 11). The helix-turn-helix of GLD-3 $_{NT}$  is, however, not sufficient for the interaction, because mutations of GLD-2 (either R574E or R756A) predicted to destabilize binding to the extended segment of GLD-3 $_{NT}$  impaired complex formation (Fig. 3B, lanes 5 and 12 and lanes 6 and 13). As a note, Arg574 and Arg756 are much better conserved than Asn629 in GLD-2 orthologs and are not present in Trf4. Finally, we tested the effect of the E875K substitution on complex formation. This substitution was originally identified in the *C. elegans* mutant *gld-2(h292)* to result in the failure of normal gametes formation and ectopic proliferation in the proximal germ line (42). In the mutant, the interaction with GLD-3 was reduced, although the protein levels were comparable to those of the wild type (10). However, this conserved residue (Fig. S3) points inside the interdomain cleft and is far from GLD-3 $_{NT}$  (Fig. 2A). Consistently, the pull-down assays showed that GLD-2 E875K was able to coprecipitate GLD-3 $_{1-460}$  (Fig. 3B, lanes 7 and 14).

**GLD-3 Contributes Both Directly and Indirectly to the Enzymatic Activity of GLD-2.** The finding that GLD-3 $_{NT}$  is positioned on the outer surface of GLD-2 and is far from the active site (Figs. 2A and 3A) raises the question of how it boosts the poly(A)-polymerase activity. In the case of the canonical poly(A)-polymerase Pap1, RNA binds at a positively charged surface that spans from the RRM domain (at the entrance of the cleft) to the active site (27) (Fig. S2A and B). When analyzing the surface properties of GLD-2 $\Delta I$ -GLD-3 $_{NT}$  based on electrostatic potential and evolutionary conservation (Fig. 4A), we noticed that the conserved positively charged surface of the active site is extended at the entrance of the cleft, with residues from GLD-2 (Lys645 and Arg655) and from the helix-turn-helix of GLD-3 $_{NT}$  (Arg42 and Lys43). We purified recombinant GLD-2-GLD-3 $_{1-460}$  mutant complexes with reverse-charged substitutions at these residues and tested them in polyadenylation assays (Fig. 4B).

Complexes containing either the GLD-2 K645E,R655E or the GLD-3 R42E,K43E mutants showed decreased activity with respect to the wild-type sample (Fig. 4B, lanes 2–4, 5–7, and 14–16, respectively). Combining both sets of protein mutants had an additive effect on the extent of polyadenylation (Fig. 4B, lanes 8–10), but did not significantly decrease the stability of the complex, as assessed by measuring the corresponding melting temperatures using thermofluor assays (Fig. 4C). We concluded that GLD-3 has a direct effect on GLD-2 poly(A)-polymerase activity, probably by contributing to RNA binding at the entry of the active site cleft. However, we found that the melting temperatures



**Fig. 4.** Molecular basis for the stimulation of GLD-2 activity by GLD-3. (A) Structure of GLD-2 $\Delta I$ -GLD-3 $_{NT}$  shown in surface representation; colored at *Left* according to electrostatic surface potential over the range from  $-5$  kT/e (red) to  $+5$  kT/e (blue) and at *Right* according to evolutionary conservation, from dark violet (conserved) to white (variable). The surface conservation was calculated with the ConSurf Server (58) by using the alignment of the four GLD-2 orthologs in Fig. 2C. For clarity, GLD-3 is shown in a ribbon representation in green. Residues targeted for mutagenesis are indicated. (B, *Left*) Polyadenylation assay of GLD-2-GLD-3 wild type (wt) and mutants (0, 20, 100, 500 nM) in the presence of a 5'- $^{32}$ P-labeled 10-mer poly(A) RNA substrate (100 nM). (B, *Right*) Coomassie-stained 12% SDS/PAGE of GLD-2-GLD-3 protein complexes used for polyadenylation assays. (C) Protein stability of different GLD-2-GLD-3 complexes as determined by thermofluor experiments. Proteins were assayed in the presence of 150 mM NaCl and 10% glycerol. The table at *Left* shows the melting temperature ( $T_m$ ), and *Right* shows the corresponding normalized curves.

of GLD-2 either in isolation ( $T_m$  33 °C) and or supplemented with GLD-3 $_{88-460}$  ( $T_m$  34.2 °C) were markedly lower than when in the presence of GLD-3 $_{1-460}$  or GLD-3 $_{NT}$  ( $T_m$   $> 41$  °C) (Fig. 4C). Thus, GLD-3 also appears to have an indirect effect in stabilizing the fold of GLD-2. In the structure, GLD-3 covers hydrophobic surface patches of GLD-2, shielding them from solvent. With hindsight, the finding that GLD-2 in isolation is capable of adding only a few adenosines (10, 23, 43) (Fig. 1B) might be explained by the rapid inactivation of the protein in the time and conditions of the assays.

**Specific Substrate-Binding Properties of the GLD-2 Active Site.** The most severe effect on the polyadenylation properties of GLD-2-GLD-3 $_{1-460}$  was obtained with the reverse-charged substitution of Glu875 corresponding to *gld-2(h292)* (42). The Glu875Lys mutation did not destabilize the GLD-2-GLD-3 $_{1-460}$  complex in thermofluor assays ( $T_m$   $\sim 40$  °C; Fig. 4C), yet it essentially abolished enzymatic activity (Fig. 4B, lanes 11–13). Because Glu875 points into the interdomain cleft, we superposed the catalytic and central domains of GLD-2 separately to the equivalent domains of Pap1 in the ATP- $A_5$ -bound structure (27) and compared the residues in the active site (Fig. 5A).

The structural analysis shows that the chloride ion in the GLD-2 $\Delta$ –GLD-3 $\Delta$  structure occupies the position between the  $\beta$ - and  $\gamma$ -phosphates of ATP. Functionally important residues are present at conserved positions, including those expected to mediate electrostatic and polar interactions with the magnesium ions and with the phosphate and ribose moieties of the nucleotides, and those expected to mediate hydrophobic interactions with the bases (Fig. 5A and Fig. S5). In the case of Cid1, the specificity for UTP depends on His336 (38, 40, 44, 45). In the case of Pap1, the base of ATP is recognized indirectly by stacking interactions with the base of the nucleotide at the 3' end of the RNA substrate (nucleotide  $-1$ ) (27). With the possible exception of Pap1 Asn226, no other polar residue of the canonical poly(A) polymerase approaches the pyrimidines of ATP and nucleotide  $-1$  (27). In GLD-2, however, this site is lined by several polar residues that are strictly conserved across species (Figs. S3 and S4). In particular, Glu875 from the NRM segment is expected to approach both adenines and is flanked by Asn673 from helix  $\alpha$ 4 (Fig. 5A), suggesting an involvement in substrate binding and specificity. Therefore, we assessed the substrate specificity of GLD-2 in polyadenylation assays by using different homopolymeric substrates in vitro. We found that GLD-2–GLD-3 $\Delta$  is highly selective for a poly(A) RNA substrate and does not polyadenylate poly(U) or poly(C) RNAs (Fig. 5B and Fig. S2C). Remarkably, addition of a single adenine nucleotide at the 3' end of a poly(U) RNA was sufficient to restore GLD-2-mediated polyadenylation (Fig. 5B).

## Conclusions

At the mechanistic level, GLD-3 increases the poly(A)-polymerase activity of GLD-2 and directs it to specific substrates (10, 33). Although it was expected that GLD-3 would provide an RNA-binding KH domain to boost the polyadenylation properties of this noncanonical poly(A) polymerase, the KH2–KH5 unit of GLD-3 has no detectable RNA-binding properties in vitro (34), and we now find the so-called KH1 domain does not adopt the structure of a KH fold: It wraps around the surface of the poly(A) polymerase rather than docking onto it as observed for the globular RRM of Pap1 (27, 46–48). The mechanism of GLD-3-mediated activation of GLD-2 is conceptually rather different, and involves a direct contribution of residues lining the RNA-binding site and an indirect contribution due to increased

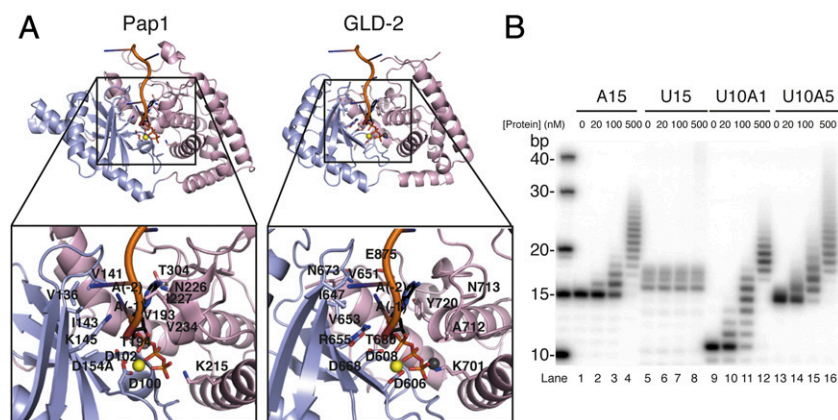
stability. In this context, the observation that the closely related nucleotidyl-transferase Cid1 is active without binding partners can be rationalized on the basis of its different surface properties (38–40, 49).

GLD-2 has distinct structural features in the active site and distinct biochemical properties, with specificity for an adenine at the 3' end of the RNA substrate. This specificity is unusual, because Pap1 and Trf4 polyadenylate a relatively broad range of RNA substrates (27, 37, 50, 51). It is tempting to speculate that this property might be tailored to the majority of GLD-2 substrates, namely dormant mRNAs with shortened poly(A) tail at their 3' end.

## Experimental Procedures

**Protein Expression and Purification.** *C. elegans* GLD-2 and GLD-3 proteins and mutants were subcloned in expression vectors with an N-terminal His-tag cleavable by tobacco etch virus (TEV) protease (SI Experimental Procedures). They were expressed (in isolation or as a complex) in *Escherichia coli* BL21 Gold pLys cells (Stratagene) by using Terrific Broth (TB) medium and overnight induction at 18 °C. Expression of selenomethionine-derivatized proteins was carried out in minimal medium in the presence of 50 mg/L selenomethionine. All proteins were purified with a similar protocol (see SI Experimental Procedures for details). In short, cells were lysed in a buffer containing 500 mM NaCl and protease inhibitors (Roche), and the lysate was loaded on a Ni-NTA affinity column (His60; GE Healthcare). After elution and tag cleavage, the proteins were further purified by ion exchange chromatography (Q Sepharose; GE Healthcare) and size exclusion chromatography (Superdex; GE Healthcare). They were concentrated in 20 mM Tris pH 7.5, 150 mM NaCl, and 4 mM DTT. The stability of the purified samples was assessed by thermofluor experiments (SI Experimental Procedures). Coexpression and pull-down assays were carried out by using a Tecan robot, using 50  $\mu$ L of total lysate and 200  $\mu$ L of Ni-NTA resin, and a buffer containing 250 mM NaCl (SI Experimental Procedures).

**Crystallization and Structure Determination.** Crystals of the selenomethionine-derivatized GLD-2 $\Delta$ –GLD-3 $\Delta$  grew at 20 °C in a sitting-drop vapor diffusion setup by using as reservoir solution 18% (vol/vol) PEG MME 550, 50 mM potassium nitrate, 60 mM magnesium nitrate, and 30 mM Hepes pH 7.0. Single crystals appeared in a few days. They were transferred to a cryoprotectant solution containing 15% (vol/vol) ethylene glycol and flash-cooled in liquid nitrogen. Single-anomalous diffraction data were collected at the beamline PXII at the Swiss Light Source (Switzerland) at 100K and processed with X-ray detector software (52). The crystals belong to an orthorhombic spacegroup with one molecule per asymmetric unit. The HKL2MAP program (53) located 11 of 14 selenium sites. The initial phases were calculated with



**Fig. 5.** Idiosyncratic biochemical and structural features of the GLD-2 poly(A)-polymerase site. (A) Cartoon representations of Pap1 (Left; PDB ID code 2Q66; ref. 27) and GLD-2 $\Delta$ –GLD-3 $\Delta$  (Right; GLD-3 not displayed, RNA and ATP molecules are modeled according to the superposition with Pap1; ref. 27) after optimal, separate superposition of the catalytic and central domains. The zoom-in views show residues outlined in the text involved in interaction with the RNA and ATP molecules as seen in Pap1 and at the same or similar position in GLD-3-complexed GLD-2. Molecules are rotated by 20° around a horizontal axis with respect to the view in Fig. 2A and colored accordingly. The RNA and ATP molecules are shown as cartoon and stick representations, and the magnesium and chloride ions are shown as yellow and gray spheres, respectively. (B) Polyadenylation assay of GLD-2–GLD-3 wild type (0, 20, 100, 500 nM) in the presence of 5'- $^{32}$ P-labeled poly(U):poly(A) heterooligomers (100 nM).

PHENIX software (54). Model building was first carried out with PHENIX (54) and BUCCANEER (55). Further manual model building and iterative refinement were performed by using COOT (56) and PHENIX (54). Detailed data collection and refinement statistics are summarized in Table S1.

**Polyadenylation Assays.** Polyadenylation assays were carried out in a 10- $\mu$ L reaction volume containing 25 mM Tris pH 8.0, 20 mM KCl, 5 mM MgCl<sub>2</sub>, 10% (vol/vol) glycerol, 0.02% (vol/vol) Nonidet P-40, 1 mM DTT, and 0.05 mg/mL BSA. Final concentrations were either 20, 100, or 500 nM for proteins and 0.5 mM for ATP. Relevant 5'-<sup>32</sup>P-labeled RNA (synthetic oligos from biomers.net;  $\gamma$ -[<sup>32</sup>P]ATP 3000 Ci/mmol from PerkinElmer) was added to a final concentration of 100 nM to start the reaction. Reaction mixtures were incubated at 30 °C for 10 min and quenched by adding 10  $\mu$ L of a buffer containing 50 mM EDTA, 0.1% SDS, and 2 mg/mL Proteinase K (New England

Biolabs). Samples were incubated 10 min at 37 °C before diluting 1:3 in 95% (vol/vol) formamide, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanole, and loading 2  $\mu$ L of each reaction on a 10% (wt/vol) polyacrylamide/7M urea gel. Gels were exposed overnight at -80 °C to Fuji image plates and visualized by using a Typhoon FLA 7000 phosphorimager (GE Healthcare).

**ACKNOWLEDGMENTS.** We thank the Max Planck Institute of Biochemistry Core Facility and Crystallization Facility, the staff members at the beamlines PXII and PXIII of the Swiss Light Source, and members of our laboratories for useful discussions and critical reading of the manuscript. This study was supported by the Max Planck Gesellschaft; the European Commission from European Research Council Advanced Investigator Grant 294371, Marie Curie ITN RNPnet; and Deutsche Forschungsgemeinschaft Grants SFB646, SFB1035, GRK1721, FOR1680, CIPSM (to E.C.), FOR855 and EC369/3-1 (to C.R.E.), and BU2451/1-2 (to K.N.).

- Eckmann CR, Rammelt C, Wahle E (2011) Control of poly(A) tail length. *Wiley Interdiscip Rev RNA* 2(3):348–361.
- Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136(4):688–700.
- Goldstrohm AC, Wickens M (2008) Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* 9(4):337–344.
- Chen CY, Shyu AB (2011) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* 2(2):167–183.
- Norbury CJ (2013) Cytoplasmic RNA: A case of the tail wagging the dog. *Nat Rev Mol Cell Biol* 14(10):643–653.
- Tadros W, Lipshitz HD (2005) Setting the stage for development: mRNA translation and stability during oocyte maturation and egg activation in *Drosophila*. *Dev Dyn* 232(3):593–608.
- Radford HE, Meijer HA, de Moor CH (2008) Translational control by cytoplasmic polyadenylation in *Xenopus* oocytes. *Biochim Biophys Acta* 1779(4):217–229.
- Richter JD, Klann E (2009) Making synaptic plasticity and memory last: Mechanisms of translational regulation. *Genes Dev* 23(1):1–11.
- Barckmann B, Simonelig M (2013) Control of maternal mRNA stability in germ cells and early embryos. *Biochim Biophys Acta* 1829(6-7):714–724.
- Wang L, Eckmann CR, Kadyk LC, Wickens M, Kimble J (2002) A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature* 419(6904):312–316.
- Barnard DC, Ryan K, Manley JL, Richter JD (2004) Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell* 119(5):641–651.
- Kwak JE, Wang L, Ballantyne S, Kimble J, Wickens M (2004) Mammalian GLD-2 homologs are poly(A) polymerases. *Proc Natl Acad Sci USA* 101(13):4407–4412.
- Rouhana L, et al. (2005) Vertebrate GLD2 poly(A) polymerases in the germline and the brain. *RNA* 11(7):1117–1130.
- Nakanishi T, et al. (2006) Possible role of mouse poly(A) polymerase mGLD-2 during oocyte maturation. *Dev Biol* 289(1):115–126.
- Cui J, Sackton KL, Horner VL, Kumar KE, Wolfner MF (2008) Wispy, the *Drosophila* homolog of GLD-2, is required during oogenesis and egg activation. *Genetics* 178(4):2017–2029.
- Sartain CV, Cui J, Meisel RP, Wolfner MF (2011) The poly(A) polymerase GLD2 is required for spermatogenesis in *Drosophila melanogaster*. *Development* 138(8):1619–1629.
- Suh N, Jedamzik B, Eckmann CR, Wickens M, Kimble J (2006) The GLD-2 poly(A) polymerase activates gld-1 mRNA in the *Caenorhabditis elegans* germ line. *Proc Natl Acad Sci USA* 103(41):15108–15112.
- Cui J, Sartain CV, Pleiss JA, Wolfner MF (2013) Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in *Drosophila*. *Dev Biol* 383(1):121–131.
- Kwak JE, et al. (2008) GLD2 poly(A) polymerase is required for long-term memory. *Proc Natl Acad Sci USA* 105(38):14644–14649.
- Udagawa T, et al. (2012) Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. *Mol Cell* 47(2):253–266.
- Katoh T, et al. (2009) Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev* 23(4):433–438.
- D'Ambrogio A, Gu W, Udagawa T, Mello CC, Richter JD (2012) Specific miRNA stabilization by Gld2-catalyzed monoadenylation. *Cell Reports* 2(6):1537–1545.
- Lee M, et al. (2014) Adenylation of maternally inherited microRNAs by Wispy. *Mol Cell* 56(5):696–707.
- Kwak JE, Wickens M (2007) A family of poly(U) polymerases. *RNA* 13(6):860–867.
- Martin G, Keller W (2007) RNA-specific ribonucleotidyl transferases. *RNA* 13(11):1834–1849.
- Schmidt MJ, Norbury CJ (2010) Polyadenylation and beyond: Emerging roles for noncanonical poly(A) polymerases. *Wiley Interdiscip Rev RNA* 1(1):142–151.
- Balbo PB, Bohm A (2007) Mechanism of poly(A) polymerase: Structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure* 15(9):1117–1131.
- Aravind L, Koonin EV (1999) DNA polymerase beta-like nucleotidyltransferase superfamily: Identification of three new families, classification and evolutionary history. *Nucleic Acids Res* 27(7):1609–1618.
- Eckmann CR, Kraemer B, Wickens M, Kimble J (2002) GLD-3, a bicaudal-C homolog that inhibits FBF to control germline sex determination in *C. elegans*. *Dev Cell* 3(5):697–710.
- Papin C, Rouget C, Mandart E (2008) *Xenopus* Rbm9 is a novel interactor of XGld2 in the cytoplasmic polyadenylation complex. *FEBS J* 275(3):490–503.
- Kim KW, et al. (2009) Antagonism between GLD-2 binding partners controls gamete sex. *Dev Cell* 16(5):723–733.
- Cragle C, MacNicol AM (2014) Musashi protein-directed translational activation of target mRNAs is mediated by the poly(A) polymerase, germ line development defective-2. *J Biol Chem* 289(20):14239–14251.
- Eckmann CR, Crittenden SL, Suh N, Kimble J (2004) GLD-3 and control of the mitosis/meiosis decision in the germline of *Caenorhabditis elegans*. *Genetics* 168(1):147–160.
- Nakel K, Hartung SA, Bonneau F, Eckmann CR, Conti E (2010) Four KH domains of the *C. elegans* Bicaudal-C ortholog GLD-3 form a globular structural platform. *RNA* 16(11):2058–2067.
- Martin G, Doublé S, Keller W (2008) Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. *Biochim Biophys Acta* 1779(4):206–216.
- Holm L, Rosenstrom P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545–W549.
- Hamill S, Wolin SL, Reinisch KM (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc Natl Acad Sci USA* 107(34):15045–15050.
- Yates LA, et al. (2012) Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat Struct Mol Biol* 19(8):782–787.
- Munoz-Tello P, Gabus C, Thore S (2012) Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure* 20(6):977–986.
- Lunde BM, Magler I, Meinhart A (2012) Crystal structures of the Cid1 poly(U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res* 40(19):9815–9824.
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797.
- Kadyk LC, Kimble J (1998) Genetic regulation of entry into meiosis in *Caenorhabditis elegans*. *Development* 125(10):1803–1813.
- Minasaki R, Rudel D, Eckmann CR (2014) Increased sensitivity and accuracy of a single-stranded DNA splint-mediated ligation assay (sPLAT) reveals poly(A) tail length dynamics of developmentally regulated mRNAs. *RNA* 11(2):111–123.
- Munoz-Tello P, Gabus C, Thore S (2014) A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure. *Nucleic Acids Res* 42(5):3372–338.
- Yates LA, et al. (2015) Structural plasticity of Cid1 provides a basis for its distributive RNA terminal uridylyl transferase activity. *Nucleic Acids Res* 43(5):2968–2979.
- Bard J, et al. (2000) Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science* 289(5483):1346–1349.
- Martin G, Keller W, Doublé S (2000) Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J* 19(16):4193–4203.
- Martin G, Möglich A, Keller W, Doublé S (2004) Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase. *J Mol Biol* 341(4):911–925.
- Rissland OS, Mikulasova A, Norbury CJ (2007) Efficient RNA polyuridylation by non-canonical poly(A) polymerases. *Mol Cell Biol* 27(10):3612–3624.
- Butler JS, Platt T (1988) RNA processing generates the mature 3' end of yeast CYC1 messenger RNA in vitro. *Science* 242(4883):1270–1274.
- Butler JS, Sathale PP, Platt T (1990) RNA processing in vitro produces mature 3' ends of a variety of *Saccharomyces cerevisiae* mRNAs. *Mol Cell Biol* 10(6):2599–2605.
- Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
- Pape T, Schneider TR (2004) HKL2MAP: A graphical user interface for phasing with SHELX programs. *J Appl Cryst* 37(Pt 5):843–844.
- Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
- Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62(Pt 9):1002–1011.
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
- Schrodinger LLC (2010) The PyMOL Molecular Graphics System, Version 1.5.0.4.
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38(Web Server issue):W529–W533.