TOWARD A MECHANISTIC UNDERSTANDING OF THE MULTIFUNCTIONALITY OF RNA-BINDING PROTEINS:
A (GLO)BAL PERSPECTIVE

Joel Vincent Tamayo

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
MOLECULAR BIOLOGY

Advisor: Elizabeth R. Gavis

September 2018
ABSTRACT

Post-transcriptional gene regulation requires specificity in molecular recognition. RNA binding proteins (RBPs), for example, face the challenge of precisely identifying motifs that control splicing, polyadenylation, localization, and translation. This challenge is compounded by the requirement for many RBPs to interact with numerous target RNAs and function in multiple aspects of RNA metabolism. Since the precise roles of RNA binding proteins in regulating gene expression are likely to be determined by their RNA binding specificity, it is important to have a clear understanding of how these proteins interact with their RNA targets as well as how these interactions transduce a particular biological response.

The Drosophila glorund (glo) gene provides an ideal model for studying these interactions, as Glo is a member of the hnRNP F/H family of proteins, which function in nearly all aspects of RNA metabolism. To gain insight into how Glo coordinates its multiple functions, we have determined the crystal structures of each of Glo’s three RNA binding domains and have identified mutations that differentially alter Glo’s ability to recognize known RNA substrates in vitro. By evaluating the effect of these mutations on Glo function in vivo, we demonstrate that Glo uses distinct RNA-binding surfaces to differentially regulate its targets. Additionally, using a bioinformatic approach based on known targets of Glo, we identify ~800 other potential RNA targets of Glo, suggesting that the functions of Glo may be much more widespread than we previously appreciated. Our data suggest a molecular mechanism for the regulation of diverse RNA targets by Glo that may be applied to understanding the functional diversity of other RNA-binding proteins. Moreover, it illustrates a general principle that RBPs may use different combinations of RNA recognition interfaces to regulate subsets of target RNAs that correlate with particular biological functions.
TABLE OF CONTENTS

Abstract. ................................................................. iii
Table of Contents. ....................................................... iv
List of Figures and Tables. ........................................... vii
Acknowledgements. ................................................... xi
Chapter 1: Introduction. ............................................. 1
   Section 1.1: Post-Transcriptional Gene Regulation in Eukaryotes. ............ 2
   Section 1.2: The Role of hnRNPs in Post-Transcriptional Gene Regulation. ...... 8
   Section 1.3: Conserved Structures of hnRNPs. ..................................... 10
   Section 1.4: *Drosophila* Oogenesis as a System for Studying RNA Metabolism. .. 14
   Section 1.5: Thesis Overview. .................................................. 19
Chapter 2: The *Drosophila* hnRNP F/H Homolog Glorund Uses Two Distinct RNA-
     Binding Modes to Diversify Target Recognition. .................................. 20
   Section 2.1: Introduction. .................................................. 21
   Section 2.2: Crystal Structures of Glo qRRMs Predict that Glo Recognizes G-tract
     RNA. ................................................................................. 23
   Section 2.3: Crystal Structures of Glo qRRMs Suggest a Second, Distinct RNA
     Interaction Mode. .......................................................... 29
   Section 2.4: Glo Recognizes Two Different TCE Features. ............................. 30
   Section 2.5: Identification of Glo Residues that Mediate G-tract or TCE RNA Binding. . 34
   Section 2.6: Translational Repression of *nos* Requires Both Modes of RNA
     Recognition. ....................................................................... 40
   Section 2.7: TCE Function *in vivo* Requires Both Glo Binding Motifs. ............. 43
   Section 2.8: Glo Functions During Mid-oogenesis Are Mediated Solely by the G-tract
     Binding Mode. ................................................................... 46
   Section 2.9: Discussion. ....................................................... 50
Section 7.6: Immunoblotting ................................................................. 103
Section 7.7: quantitative-PCR .............................................................. 103
Section 7.8: smFISH ............................................................................ 103
Section 7.9: Analysis of Eggshell and Embryonic Phenotypes ............... 104
Section 7.10: Immunoprecipitation ...................................................... 104
Section 7.11: RNA co-immunoprecipitation ......................................... 104
Section 7.12: Viability Assay ............................................................... 105
Section 7.13: Statistical Methods ......................................................... 105
Section 7.14: Accession Numbers ....................................................... 106

Appendix .......................................................................................... 107
Section A.1: Script to Identify Potential RNA Targets of Glo on a Genome-Wide Basis ................................................................. 108

References ......................................................................................... 109
LIST OF FIGURES AND TABLES

Chapter 1

Figures

Figure 1.1. Activation of CPE-mediated cytoplasmic polyadenylation. ................. 4
Figure 1.2. Schematic drawing of ASH1 mRNA localization in S cerevisiae. ............ 5
Figure 1.3. Common mechanisms for translational repression of transported mRNAs. ................................................................. 6
Figure 1.4. The functions of RBPs in post-transcriptional gene expression ............ 7
Figure 1.5. The hnRNP family. .............................................................. 9
Figure 1.6. Characteristics of the RNA Recognition Motif. .............................. 11
Figure 1.7. Characteristics of the K-Homology Motif. ..................................... 13
Figure 1.8. Overview of Drosophila oogenesis. .............................................. 16
Figure 1.9. Localized distributions of grk, bcd, osk, and nos mRNAs. ............... 17

Tables

Table 1.1. List of Drosophila hnRNPs function and conservation. ..................... 15

Chapter 2

Figures

Figure 2.1. Crystal structures of Glo qRRM1, qRRM2 and qRRM3 domains identify RNA-binding motifs. ............................................. 24
Figure 2.2. Glo qRRMs bind to a single-stranded G-tract sequence. ................. 28
Figure 2.3. Glo recognizes two elements in the nos TCE that contribute higher and lower affinity components. .................................................. 31
Figure 2.4. Effect of TCE Mutations on Glo binding. ................................. 33
Figure 2.5. Identification of Glo Residues Critical for G-tract or TCEIII Recognition. . 35
Figure 2.6. Mutation of loop residues selectively disrupts G-tract RNA binding. ...... 36
Figure 2.7. Mutation of basic RNP2 residues selectively disrupts TCEIII UA-rich motif binding. 38

Figure 2.8. Glo uses both modes of RNA recognition for translational repression of nos. 41

Figure 2.9. Expression of GFP-Glo transgenes. 42

Figure 2.10. TCE function \textit{in vivo} requires both Glo recognition sites. 44

Figure 2.11. Regulation of chromosome dispersion and dorsal-ventral patterning are mediated solely by the G-tract binding mode of Glo. 48

Figure 2.12. The Glo G-tract binding mode is required for Glo function in nurse cell chromosome dispersion. 49

Figure 2.13. Glo mutant proteins interact with Hfp and Hrp48. 52

Tables

Table 2.1. Crystallographic Summary of Glo qRRMs. 26

Table 2.2. Identification of Glo residues critical for G-tract or TCEIII recognition. 37

Chapter 3

Figures

Figure 3.1. Translational Repression of \textit{nos} Requires Binding by Two qRRMs \textit{in vivo}. 57

Figure 3.2. Expression of GFP-Glo Transgenes. 58

Figure 3.3. Model for Glo-TCE Interaction. 60

Figure 3.4. Each qRRM Confers One Component of TCE Recognition. 61

Figure 3.5. Survival to Adulthood Requires Multiple qRRMs and Both RNA Recognition Modes of Glo. 63

Figure 3.6. Glo Functions During Mid-oogenesis Require Binding of a Single qRRM to G-tract motifs. 65

Tables

Table 3.1. Summary of Glo RNA-binding Point Mutants. 68
Chapter 4

Figures

Figure 4.1. Glo Regulates *osk* During late-stages of oogenesis. ............................74

Figure 4.2. Glo Associates with *osk* mRNA in vivo. ...........................................75

Figure 4.3. Glo Recognizes Two Elements in the *osk* SOLE that Contribute Higher-
And Lower-Affinity Components. .................................................................77

Figure 4.4. Glo Uses Both Modes of RNA Recognition for Localization of *osk* to the
Posterior. ...........................................................................................................79

Chapter 5

Figures

Figure 5.1. Predicted Secondary Structure of RNAs identified by *In silico* Analysis. . .90

Figure 5.2. Gene Ontology Analysis of Potential Glo Target RNAs. .......................91

Tables

Table 5.1: Potential RNA Targets of Glo Identified by *In Silico* Analysis. ...............84
This thesis is dedicated to my family.

To my mother – who is my biggest supporter and best friend.

To my father – who is my harshest critic and greatest role model.

To my twin sister – who is my better half in every way imaginable.

And to my brother – who is proof that all you need to succeed is passion.

I couldn’t have done this without you.
ACKNOWLEDGMENTS

First and foremost, I must thank my advisor, Elizabeth Gavis, for her steady guidance and firm support throughout my graduate career. When I first came to Princeton, I couldn’t believe that I was going to be spending the next few years working alongside some of the greatest researchers in the world—I was, in a lot of ways, star struck. Almost immediately, Liz amazed and inspired me with her appreciation for genetics and her continued desire to do bench work. Her passion is infectious and it made it so much easier for me to deal with the perpetual intellectual uncertainty that comes with doing research. She taught me how to think in a rigorously logical fashion, but challenged me to think creatively. She taught me lessons that have left an indelible mark, and I can never thank her enough. I am still star struck and amazed by her brilliance.

Much of the work described in this thesis would not be possible if not for the efforts of our amazing collaborators at the NIEHS, Takamasa Teramoto and Traci Tanaka Hall. It has been a pleasure visiting and working so closely with you over the last few years and I am extremely proud of the work we have done. I consider myself incredibly fortunate to have been a part of this collaboration.

I am also grateful for all the members of the department that have made my scientific and personal life so fulfilling and enjoyable. I am lucky to have had an amazing thesis advisory committee and I would like to thank Alexei Korenykh and Trudi Schüpbach for all their feedback and helpful suggestions over the years. They have always been a wonderful source of scientific inspiration and guidance. Robert Leach was instrumental in helping me with the bioinformatics analysis and I couldn’t have done any of the imaging without Gary Laevsky. The administrative staff members, notably Elena Chiarchiaro, Terry Macfarland, and Matthew Montondo have been a great source of support over the years with all the work they do behind the scenes and I am forever grateful to them for helping me every step of the way.

I have had the good fortune to cross paths with many Gavis lab members during my graduate career and I couldn’t have asked for better peers. To Kristina Sinsimer, Balpreet
Bhogal, Mala Misra, Nicolette Belletier, and Kelsey Hughes: thank you for always keeping me in check and occasionally laughing at my slightly inappropriate jokes. I would also like to thank Whitby Eagle, Catherine Ruesch, and Matthew Niepielko. You have made my time here very entertaining and I appreciate you for always indulging my questions about particles/granules (they’re the same thing, right?). I have also found a good friend in Hendia Edmund, and I have enjoyed going through the program alongside her and hearing about her opinions on all things political and pop culture related. I am grateful to our lab manager, Seema Chatterjee, who works tirelessly to keep the lab running smoothly and has even helped me with cloning whenever my project proved too overwhelming. To the newest member of the Gavis lab: Anna Hakes, it was a pleasure mentoring you during your rotation and I look forward to seeing how you continue to push the frontiers of science. I cannot forget my brilliant undergraduate, Amelia Foss, who helped motivate this work and taught me far more about science and mentoring than I could have ever taught her. Finally, I want to thank my Gavis Lab Bros. Evan Abbaszadeh, Conrad Tenenbaum, and Yingshi Peng. You guys made it fun for me to come into lab every day and I consider you some of my closest friends.

My life at Princeton will always be special to me, in no small part because of my friends and classmates. I couldn’t have imagined going through this experience without my bro’s Evan, Conrad, and my best friend, Geoff Dann. Thank you guys for always being ‘bout that life and keeping it all the way 💯. From our trips to the city, to our nights at the Snakehole Lounge, and even for the PU reunions nights I’ll never remember, thank you. Aside from being great friends, you were and continue to be great role models and I can’t wait to celebrate with you when this is over. Equally important are my girls, Betsy and Michaela. You have been amazing friends and even better dance partners. To my classmates, Allison Hall and Christopher Wright, it has been amazing to watch you grow and succeed and I’m so lucky to have you in my life.

I learned so much in the last few years—about science, about myself, and about life. But the greatest discovery I made at Princeton is my girlfriend, Rebecca Alizzi. She has been a constant source of love and encouragement and I am beyond lucky to have her in my life.
She continually pushes me to be a better person and gives me both the motivation to work hard and the comfort to make it through times of stress. I am so grateful to have shared this journey with her. I’m also grateful to have had the love and support of the rest of the Alizzi clan. Donna and Tony, thank you for welcoming me into your beautiful family and for allowing me to be with your amazing daughter. You have been like 2\textsuperscript{nd} parents to me, and I appreciate that more than you’ll ever know.

Finally, I would like to acknowledge my family. Without their unconditional support and encouragement, I would not have come close to where I am today. To my mom, who is the strongest person I know and a true inspiration—I share this accomplishment with you the most. You sacrificed so much in your life so that I would have every opportunity in the world and I am forever indebted to you for that. Thank you for your love and unwavering support. To my dad, who has always selfishly encouraged me to explore my interests: Thank you for always keeping me grounded and teaching me the importance of a healthy work/life balance. I hope you feel as proud as I do to call you my dad. To my twin sister Brittney, who has literally been on this journey with me since birth (from the womb to the tomb, am I right?): Thank you for your continued interest in everything I do. It has been amazing having another scientist in the family and I am so lucky to have always had a best friend by default. And finally, to my brother Nicolas, who, despite being younger than me, taught me more about life than I could ever hope to teach him. Thank you for showing me that anything is possible as long as you work hard and do what you love.
“It is through science that we prove, but by intuition that we discover.”

- Henri Poincare
CHAPTER 1:

INTRODUCTION
1.1 Post-Transcriptional Gene Regulation in Eukaryotes

Animal and plant genomes encode hundreds of RNA-binding proteins (RBPs), predicting numerous and diverse biological roles for post-transcriptional gene regulation (Castello et al., 2012; Gamberi et al., 2006; Lehner and Fraser, 2004; Lorkovic and Barta, 2002). Indeed, mechanisms for post-transcriptional control of gene expression are widespread among eukaryotes and have proved crucial for cellular and developmental events, including the establishment and patterning of body axes, asymmetric cell fate decisions, and morphogenesis. Even before transcripts emerge from RNA polymerase II, they are covered in RBPs, which mediate various RNA processing reactions including splicing, 5’-end capping, and 3’-end cleavage and polyadenylation (Glisovic et al., 2008; Perales and Bentley, 2009). It is therefore not surprising that mutations in RBPs are often associated with cancer and numerous hereditary diseases (Castello et al., 2013; Lukong et al., 2008). This section will provide a brief overview of some of these processes as an introduction to the diverse functional requirements of RBPs during mRNA metabolism. Whenever possible, several reviews are also listed that discuss in greater detail the molecular mechanisms employed by cells during post-transcriptional gene regulation.

The first step during mRNA processing is the addition of a cap at the 5’ end, which consists of a guanosine residue methylated at the N7 position. This m7g cap is linked to the first nucleotide of the RNA via a triphosphate linkage after RNA polymerase II has synthesized about 20-30 nucleotides and provides resistance to 5’ exonuclease activity by acting as a protective group from RNases (Gu and Lima, 2005; Shatkin and Manley, 2000). The m7G cap also acts as an anchor for the recruitment of the cap-binding complex (CBP) in the nucleus and the translation initiation factor 4E (eIF4E) in the cytoplasm, which are required for the subsequent processing, export, and translation of the RNA (Cowling, 2009). The m7G cap is therefore crucial for many aspects of RNA metabolism, affecting both the stability and translation of RNA.
As RNA polymerase II continues its movement along a gene, the splicing machinery assembles on the RNA, joining exon sequences while removing the introns between them. This mechanism relies on the ability of the spliceosome to recognize consensus sequences in the RNA, as well as additional RBPs to help discriminate between the exon and intron boundaries (Saltzman et al., 2011). In many cases, these RBPs also mediate alternative splicing, resulting in the production of multiple isoforms from the same gene (House and Lynch, 2008; Roy et al., 2013). Once the splicing reaction is complete, the spliceosome directs the exon junction complex (EJC) to bind the RNA near the position formerly occupied by the intron. These proteins mark the site of a successful splicing event and also influence the subsequent fate of the RNA transcript.

Before an mRNA is transported out of the nucleus and into the cytoplasm where it can interact with the translational machinery, its 3’ end must be cleaved and polyadenylated (Colgan and Manley, 1997; Proudfoot, 2011). Similar to splicing, polyadenylation is achieved by a series of RBPs and requires specific consensus sequences within the mRNA. In particular, a polyadenylation signal (AAUAAA) in the 3’ untranslated region (3’ UTR) of the RNA signals recruitment of the cleavage and polyadenylation specificity factor (CPSF) which, in turn, triggers cleavage of the transcript downstream of the polyadenylation sequence, and recruits polyA polymerases (PAPs) that catalyze the addition of ~200 adenosine nucleotides to the 3’ end of the RNA (Proudfoot and O’Sullivan, 2002; Shatkin and Manley, 2000; Zhao et al., 1999). As the poly-A tail is synthesized, poly-A binding proteins (PABPs) bind to it and, by a poorly understood mechanism, help determine the final length of the tail and stimulate maturation of the mRNA (Mangus et al., 2003).

Once mRNAs have been processed, they are exported through nuclear pores to the cytoplasm where many undergo additional regulation by RBPs before they are translated. In some cases, transcripts must first be translationally activated by cytoplasmic PABPs, which interact with sequences in the 3’ UTRs of transcripts and promote elongation of their poly-A tails (Cui et al., 2013; Radford et al., 2008). For example, in *Xenopus* oocytes, many mRNAs
are stored with short poly-A tails and, as a result, remain dormant. Upon fertilization however, these mRNAs are polyadenylated and become translationally activated. This polyadenylation is conferred by a cytoplasmic polyadenylation element binding protein (CPEB), which binds discrete cytoplasmic polyadenylation elements (CPEs) in the 3’ UTRs of transcripts (Mendez and Richter, 2001; Richter, 2007) (See also Figure 1.1). Upon fertilization, the phosphorylation of CPEB signals the recruitment of CPSF and triggers polymerization of the poly-A tail by PAPs. The elongated poly-A tail is thought to be stimulatory to translation through the interaction of PABPs, which can recruit the translational machinery (Gorgoni and Gray, 2004; Mendez and Richter, 2001; Radford et al., 2008).

**Figure 1.1. Activation of CPE-mediated cytoplasmic polyadenylation.** mRNAs containing a CPE in their 3’ UTR are translationally repressed in developing oocytes due to the interaction between CPEB and Maskin. Maskin, in turn, interacts directly with the cap-binding protein eIF4E, preventing its association with the translational machinery. During oocyte maturation, phosphorylation of CPEB causes a rearrangement of the complex, which leads to the polymerization of the poly-A tail and recruits the translational machinery by means of PABP. This figure is adapted from Mendez and Richter, 2001.

In other instances, mRNAs must first be localized to specific regions of the cell by complexes consisting of RBPs and motor proteins (Becalska and Gavis, 2009; Jansen, 2001; Martin and Ephrussi, 2009). This localization typically requires the ability of RBPs to recognize specific sequence and/or structural motifs that are frequently located in the 3’ UTRs of
transcripts. In *S. cerevisiae*, for example, the ASH1 mRNA is localized to the bud of the daughter cell by its association with myosin and actin (Bobola et al., 1996). This interaction requires the ability of two proteins, She2 and She3, to recognize localization elements located partly in the coding region and in the 3′ UTR of the ASH1 mRNA (Bohl et al., 2000; Gonsalvez et al., 2005). This interaction enhances the affinity of She2 for She3, which itself binds myosin. Thus, the localization of ASH1 mRNA is modulated by RBPs bound to cis-acting sequences (Figure 1.2).

**Figure 1.2. Schematic drawing of ASH1 mRNA localization in *S. cerevisiae*.** In the nucleus of the mother cell (dashed box, left) ASH1 mRNA interacts with She2p. This complex is joined by Puf6p and Khd1p and is then exported into the cytoplasm. There, a ternary interaction of She2p, She3p, and RNA achieves the stable and specific assembly of mature transport complexes. Once the complex has reached the bud tip (right), the mRNA is anchored and locally translated, resulting in local Ash1p production. This figure is taken from Niessing, 2011.

Localization of mRNA is often coupled to translational control, allowing cells to target the synthesis of specific proteins to their cytoplasmic site of function (Becalska and Gavis, 2009; Besse and Ephrussi, 2008). Similar to the requirements for localization, such transcript-specific translational control is mediated by RBPs that bind specific regulatory motifs in the 3′
Indeed, a single RBP may mediate both the localization and translational regulation of a given transcript, ensuring precise translational control (Huttelmaier et al., 2005; Ross et al., 1997). RBPs can inhibit translation by a variety of mechanisms, but most frequently control the step that initiates translation (Gebauer and Hentze, 2004; Hershey et al., 2012; Jackson et al., 2010). Still, it should be noted that there are examples of regulation that occur post-initiation (Onouchi et al., 2005) and, in fact, some mRNAs are regulated by multiple mechanisms that target different steps of the translation pathway (Friend et al., 2012; Gebauer et al., 2012; Nelson et al., 2004).

**Figure 1.3. Common mechanisms for translational repression of transported mRNAs.**

Translational repressors (red) can interfere with formation of the eukaryotic translation initiation factor (eIF4F) complex when bound to localizing mRNAs by recruiting eIF4E-binding protein (eIF4E-BP), thus blocking the eIF4E–eIF4G interaction (left panel). Repressors can also block 60S ribosomal subunit joining (right panel), or decrease poly-A tail length through recruitment of deadenylation complexes (lower panel). This figure is taken from Besse and Ephrussi, 2008.
In principle, every step in the pathway from RNA to protein can be controlled. Indeed, most genes require multiple layers of regulation, and the interplay of translational control with other post-transcriptional regulatory mechanisms such as mRNA transport allow cells to precisely control the expression of a particular protein. Central to virtually all aspects of post-transcriptional gene regulation – from RNA processing to mRNA localization and translation – is the interaction of RNA with RBPs (Figure 1.4). The identification of RBPs that function in multiple aspects of post-transcriptional gene regulation therefore presents exciting challenges for understanding how these diverse networks are coordinated in the cell.

**Figure 1.4. The functions of RBPs in post-transcriptional gene expression.** Once pre-mRNAs are transcribed by RNA polymerase II in the nucleus, they undergo many different processing steps mediated by RBPs, which help determine the fate of the transcript. In the cytoplasm, other RBPs can also affect the fate of the mRNA. Some modify the length of the poly-A tail whereas others bind to the 3'UTR and repress translation while directing the subcellular localization of the mRNA. Other mechanisms of post-transcriptional gene regulation are not shown.
1.2 The Role of hnRNPs in Post-Transcriptional Gene Regulation

Of all the RBPs that assemble on mRNA, the most abundant are the heterogeneous nuclear ribonucleoproteins. Indeed, hnRNPs are amongst the most abundant proteins in the nucleus, rivaling histones, and have been estimated to contain ~100 million mRNA transcripts per cell (Dreyfuss et al., 2002b; Kamma et al., 1995). The hnRNPs were first defined as a complex of proteins involved in the general packaging and processing of nascent pre-mRNA transcripts (Dreyfuss et al., 1993). However, more recent findings have shown that the functions of these proteins extend far beyond the packaging of nascent RNA, and they are now recognized as having important roles in all aspects of post-transcriptional gene regulation.

In humans, over 20 hnRNP proteins have been identified, and are designated hnRNP A through hnRNP U (Geuens et al., 2016) (See also Figure 1.5). These proteins share several structural features, and these similarities may reflect their ability to function in multiple aspects of mRNA metabolism. For example, with the exception of hnRNP U, all hnRNPs contain multiple RNA binding domains (RBDs), which can be arranged in various ways to satisfy their diverse functional requirements (Dreyfuss et al., 2002b). Additionally, many hnRNPs contain RGG motifs, which are intrinsically disordered, and thus do not adopt a single, stable structure in the absence of RNA (Rajyaguru and Parker, 2012; Thandapani et al., 2013). This makes these domains conformationally flexible and allows them to bind a variety of RNAs. Finally, the functional diversity of hnRNPs is further achieved by combining RBDs with auxiliary domains that regulate protein-protein interactions or subcellular localization. The domain composition of the hnRNPs is thus highly modular and creates structural variation in terms of domain combinations and arrangements. This modularity makes hnRNPs highly versatile proteins and allows them to participate in a wide range of biological functions.
Figure 1.5. The hnRNP family. The hnRNPs are named alphabetically from hnRNP A to hnRNP U. The hnRNPs are composed of RNA binding domains, including the RNA recognition motif (RRM) and K homology domain (KH), as well as auxiliary domains like the glycine-rich domain (Gly-rich), which mediate protein-protein interactions or subcellular localization.

Another key characteristic of the hnRNPs is that they undergo nucleocytoplasmic shuttling (Pinol-Roma and Dreyfuss, 1992). This property makes hnRNPs especially suited to regulate all aspects of post-transcriptional gene regulation, from RNA processing to mRNA localization and translation, and also provides a link between them. Indeed, some hnRNPs exhibit distinct localization patterns, and this localization is of functional importance. For example, hnRNP A2/B1 and hnRNP Q are both present in mRNA granules and have been shown to regulate mRNA transport (Bannai et al., 2004; Shan et al., 2003), whereas hnRNP P is localized to cytoplasmic stress granules where it participates in the cellular stress response (Andersson et al., 2008). Subcellular localization can thus provide an additional mechanism for regulating RBP function.
Many of the biological functions ascribed to the hnRNPs involve dynamic and cooperative interactions within large complexes, and it is likely that the different binding affinities of the hnRNPs target them to distinct RNAs within these assemblages. For example, although hnRNP A2 functions as a general splicing factor in the nucleus (Martinez-Contreras et al., 2007; Martinez-Contreras et al., 2006), it has also been shown to selectively bind the cytoplasmic transport sequence of myelin basic protein and direct its localization to oligodendrocytes (Hoek et al., 1998). Thus, hnRNPs can have both general roles in the processing of RNA, as well as specialized roles that are dependent on specific RNA-protein or protein-protein interactions.

The ability of hnRNP proteins to have both generic and specific functions during RNA metabolism can be at least partially attributed to their RNA binding properties (as discussed below). In vitro studies have shown that hnRNPs bind to a wide variety of RNA sequences. However, these sequences are relatively short and/or degenerate, and do not contain enough information to predict the binding sites of specific mRNA targets through simple sequence analysis. Thus, in order to understand how hnRNPs and other RBPs carry out their varied functions in vivo, it is essential to elucidate their RNA recognition properties as well as how they use these properties to regulate diverse RNA networks.

**1.3 Conserved Structures of hnRNPs**

The discovery of the heterogeneous nuclear ribonucleoproteins hnRNPs led to the identification of the first amino acid motifs and functional domains that confer binding to RNA (Burd and Dreyfuss, 1994). Since then, three unique RBDs have been identified in hnRNP proteins: the RNA recognition motif (RRM), the K-homology (KH) domain, and the glycine-rich domain constituting the RGG box (Figure 1.5, Figure 1.6).

The most prevalent RBD among hnRNPs is the well-studied RRM (Venter et al., 2001). The canonical RRM is characterized by the packing of a four-stranded β-sheet against two α-helices and typically interacts with RNA using aromatic residues within two highly conserved
RNP consensus sequences, designated RNP1 and RNP2, which are present on the surface of the β-sheet (Figure 1.6A,B). This exposed binding platform, in combination with the hydrophobic interactions formed by the aromatic residues, allows the RRM to bind single-stranded RNA in a non-sequence specific manner. Instead, specificity is achieved using residues outside of the two conserved RNP motifs (Maris et al., 2005) as shown in the case of RRM2 in hnRNPA1 (Figure 1.6B). RRsMs therefore have the capacity to participate in both general and specific interactions with nucleic acids.
Figure 1.6. Characteristics of the RNA-Recognition Motif. (A) Amino acid sequences of the indicated (left) RRM containing RBPs are aligned, and conserved residues are highlighted in gray, while residues required for RNA binding are in green. The RNP1 and RNP2 consensus sequences are indicated at the top. The consensus RNP Motif, as well as an alignment of the secondary structural elements that constitute each RBD are also shown. (B) Structure of the second RRM of hnRNP A1 in complex with 5'-UCAGUU-3' RNA (PDB code 5MPL) (Beusch et al., 2017). Important residues for the interaction are highlighted in green as in (A) and RNA is in magenta.

Other hnRNPs contain KH domains, which were initially identified in hnRNP K (Siomi et al., 1993). The KH domain is similar to the RRM in that it is also composed of multiple α-helices and antiparallel β-strands. However, the two types of domains differ in the precise interactions that mediate target recognition. The KH domain can be separated into two subfamilies on the basis of their topology but both are characterized by a three-stranded β-sheet packed against three α-helices (type I has βααββα; type II has αβαααβ topology) and typically interacts with RNA using residues in a Gly-X-X-Gly loop that is located between the first two α-helices (Figure 1.7). Unlike the RRM, the RNA binding surface of the KH domain is devoid of aromatic residues and instead makes base-specific contacts using an aliphatic surface. Nevertheless, KH domains can bind a large panel of sequences and are therefore well suited to meet the diverse functional requirements of hnRNPs (Valverde et al., 2008).
**Figure 1.7. Characteristics of the K-Homology motif.** (A) Amino acid sequences of the indicated (left) KH motif containing RBPs are aligned, and conserved residues are highlighted in gray. The consensus KH motif is also shown. (B) Structure of the third KH motif of Nova in complex with 5'-GAUCA-3' RNA, shown in magenta (PDB code 1EC6) (Lewis et al., 2000).

Finally, some hnRNPs use RGG motifs to bind RNA. Although originally described in hnRNP U, the RGG box has been identified in various other RBPs and is typically found in combination with other RBDs (Kiledjian and Dreyfuss, 1992). The RGG motif is characterized by closely spaced clusters of Arg-Gly-Gly tripeptide repeats with interspersed aromatic residues, although both the number and spacing of the RGG repeats vary for different proteins. This high degree of variation as well as the high density of glycine suggest that the RGG box is not a rigid protein structure. However, spectroscopic analysis and molecular modeling of the RGG repeats in Nucleolin predict a helical β-spiral structure (Ghisolfi et al., 1992).
The diversity of functions of RBPs would suggest a correspondingly large diversity in the structures that are responsible for RNA recognition. However, many RBPs are built from the three conserved RBDs described here. Instead, it is generally thought that by combining these motifs in various structural arrangements, hnRNPs can recognize RNA with the affinity and selectivity that is required to bind cognate sequences, while retaining the versatility required to regulate numerous RNAs. Alternatively, the types of RNA that can be recognized by RBDs can be increased by expanding the canonical RNA-binding surface through additional secondary structures or loops (Lunde et al., 2007; Maris et al., 2005; Wu et al., 2004). Indeed, recent structural and biochemical studies have revealed variations on the mechanism of RNA recognition for several RBDs, suggesting that this mechanism may be even more common than was previously assumed (Clery et al., 2008; Daubner et al., 2013) (See also Chapter 2).

hnRNP proteins also contain auxiliary domains that mediate important biological functions (Dreyfuss et al., 2002b). Contrary to RBDs, these domains are often divergent in protein sequence and many are also unstructured, making it difficult to group them into discrete classes. The most frequently found and best-characterized auxiliary domains are the glycine-rich (gly-rich) domains found in the hnRNP A/B proteins (Weighardt et al., 1996). In hnRNP A1, this gly-rich domain contains sites for post-translational modification and is important not only for its subcellular localization (Allemand et al., 2005; Siomi and Dreyfuss, 1995) but also for its ability to form protein-protein interactions (Weighardt et al., 1996). It is therefore important to understand how hnRNPs specify biological function in the context of both RNA-binding and auxiliary domains.

1.4 *Drosophila* oogenesis as a system for studying RNA metabolism

The development of the *Drosophila* oocyte provides a valuable system for studying post-transcriptional gene regulation because zygotic gene expression is delayed after fertilization and the earliest events in embryogenesis are controlled by proteins synthesized from maternal mRNAs. Since these mRNAs are present in the oocyte prior to fertilization, their
expression during development must be controlled post-transcriptionally. The oocyte can thus be considered a microcosm of RNA metabolic pathways, providing a mini laboratory to study how mRNAs are regulated by hnRNPs during post-transcriptional gene regulation. Indeed, at least 14 hnRNPs with structural and functional homologs in mammals have been identified in *Drosophila*, and many of them have multiple functions during oogenesis (Table 1.1).

**Table 1.1 List of *Drosophila* hnRNPs function and conservation.**

<table>
<thead>
<tr>
<th>hnRNP</th>
<th><em>Drosophila</em> homolog</th>
<th>Biological Processes in <em>Drosophila</em></th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Hrb87F</td>
<td>oocyte development, gonad development, mitotic cell cycle</td>
<td>Singh and Lakotia, 2012; Pisa et al., 2009</td>
</tr>
<tr>
<td>A/B</td>
<td>Hrb96DE</td>
<td>oocyte polarity, germline stem cell maintenance</td>
<td>Ji and Tulin, 2012</td>
</tr>
<tr>
<td>A/B</td>
<td>Squid</td>
<td>oocyte polarity, neuromuscular development</td>
<td>Krecic and Swanson, 1999</td>
</tr>
<tr>
<td>A2/B1</td>
<td>Hrp48</td>
<td>oocyte polarity, border cell migration, axon guidance</td>
<td>Huynh et al., 2004; Mathieu et al., 2007; Berger et al., 2008</td>
</tr>
<tr>
<td>F/H</td>
<td>Glorund</td>
<td>oocyte polarity, chromatin organization</td>
<td>Kalina et al., 2006; Kalina et al., 2009</td>
</tr>
<tr>
<td>I</td>
<td>Hephaestus</td>
<td>oocyte polarity, spermatid development, imaginal disc development</td>
<td>Besse et al., 2009; Robida and Singh 2003; Dansereau et al., 2002</td>
</tr>
<tr>
<td>K</td>
<td>Hrb57A</td>
<td>cell proliferation, nucleoplasmic compartment organization</td>
<td>Bornsztyk et al., 2004</td>
</tr>
<tr>
<td>L</td>
<td>Smooth</td>
<td>neuromuscular development</td>
<td>Layalle et al., 2005</td>
</tr>
<tr>
<td>M</td>
<td>Rumpelstiltskin</td>
<td>oocyte polarity, mitotic cell cycle</td>
<td>Sinsimer et al., 2011; Ducat et al., 2008</td>
</tr>
<tr>
<td>P</td>
<td>Cabeza</td>
<td>neuromuscular development</td>
<td>Zinszer et al., 1997; Wang et al., 2011</td>
</tr>
<tr>
<td>Q</td>
<td>Syncrip</td>
<td>oocyte development, neuromuscular development</td>
<td>McDermott et al., 2014; McDermott et al., 2012</td>
</tr>
<tr>
<td>CIZ1</td>
<td>Pep</td>
<td>oocyte development, nucleoplasmic compartment organization, chromatin organization</td>
<td>Reim et al., 1999; Amero et al., 1991</td>
</tr>
<tr>
<td>SFPQ</td>
<td>NonA</td>
<td>oocyte polarity, courtship behavior, visual perception</td>
<td>Reim et al., 1999; Stanek et al., 1996; Eagle &amp; Gavis, unpublished</td>
</tr>
<tr>
<td>TDP-43</td>
<td>TBPH</td>
<td>neuromuscular development</td>
<td>Buratti and Baralle, 2009; Strong 2010</td>
</tr>
</tbody>
</table>

The *Drosophila* oocyte develops within a multicellular structure called an egg chamber, which is composed of an oocyte and 15 germline-derived nurse cells surrounded by a somatic follicular epithelium. Each egg chamber develops through 14 morphologically distinct stages to give rise to a mature oocyte (Figure 1.8). The oocyte is thought to be largely quiescent and
its maturation is instead supported by the nurse cells, which produce mRNAs and proteins that are then transferred to the oocyte through ring canals. Towards the end of oogenesis, these nurse cells undergo programmed cell death and concomitantly expel their bulk cytoplasm into the oocyte, resulting in a large increase in the number of mRNAs within the oocyte. Indeed, over 3000 mRNAs are transported into the oocyte (Jambor et al., 2015), underscoring the requirement for post-transcriptional gene regulation during oogenesis.

**Figure 1.8. Overview of Drosophila oogenesis.** The *Drosophila* ovary is composed of 14-16 ovarioles, each of which contains a series of developing egg chambers (Spradling, 1993) (See also text). This figure is adapted from Becalska and Gavis, 2009.

Many of these mRNAs adopt distinct localization patterns and their expression must be appropriately regulated to specify the future embryonic body axes (Kugler and Lasko, 2009; Riechmann and Ephrussi, 2001; Weil, 2014). Four such mRNAs, *oskar* (*osk*), *nanos* (*nos*), *bicoid* (*bcd*), and *gurken* (*grk*), are particularly important for embryonic axis specification, and for this reason their regulation is especially well studied (Figure 1.9).
Figure 1.9. Localized distributions of grk, bcd, osk, and nos mRNAs. (A) Schematic showing grk (pink), bcd (green) and osk (purple) mRNA localization in mid-oogenesis (stage 9). nos mRNA is not yet localized at this stage. The anteroposterior (AP) and dorsoventral (DV) axes are indicated. (B) GFP-Staufen (green), as proxy for osk mRNA, at the posterior pole of the oocyte (oo) during mid-oogenesis. GFP-Stau is also detected in the nurse cell (nc) cytoplasm. The actin cytoskeleton is highlighted in red with phalloidin. fc, follicle cells. Orientation is the same as in A. (C-F) Visualization of endogenous mRNAs using the MS2 system: (C) grk and (D) bcd during mid-oogenesis; (E) bcd and (F) nos in late oocytes. Owing to the promoter used, the MCP-GFP and MCP-RFP fusion proteins are expressed in both the nurse cells and follicle cells, whereas the MS2- tagged mRNAs are produced only in the nurse cells. MCP-GFP/RFP that is not bound to mRNA enters both the nurse cell and follicle cell nuclei. Scale bars: 20 μm. This figure is adapted from Becalska and Gavis, 2009.

Establishment of both the anterior-posterior (A-P) and dorsal-ventral (D-V) body axes depends on signaling from the oocyte to subsets of follicle cells via Grk, a transforming growth factor α-like ligand (Gonzalez-Reyes et al., 1995; Neuman-Silberberg and Schupbach, 1993). To achieve this, grk mRNA is first localized to the posterior of the oocyte, where Grk signaling
to the overlying follicle cells triggers a reorganization of the oocyte cytoskeleton. This process enables the subsequent localization of bcd and osk mRNAs to the anterior and posterior poles of the oocyte, respectively, and the dorsal anterior localization of grk itself (Gonzalez-Reyes et al., 1995). At the anterior, Grk signaling initiates a cascade of events that will ultimately generate the embryonic D-V axis.

Osk protein synthesized at the posterior initiates the assembly of the germ plasm, a specialized cytoplasm that is maintained at the posterior pole into embryogenesis and contains the determinants that are necessary for germline and abdominal development (Lehmann and Nusslein-Volhard, 1986). Among these determinants is nos mRNA, which accumulates at the posterior of the oocyte during late-stages of oogenesis (Wang et al., 1994a). Interestingly, although nos is translated upon its localization to the posterior, Nos protein has no known function in the oocyte (Forrest et al., 2004). Similarly, Bcd, whose synthesis is repressed during oogenesis, functions only after fertilization (Driever and Nusslein-Volhard, 1988). Localization of these mRNAs during oogenesis, however, endows the newly fertilized embryo with sources for the production of opposing protein gradients that establish the A-P body axis. Bcd, acting both as a transcriptional activator and translational repressor, determines the head and thoracis regions, while Nos, a translational repressor, determines the abdomen (Thompson, 2007). In addition, localized nos mRNA is incorporated into the germ cells as they form at the posterior of the embryo, supplying the Nos protein that is necessary for germline development (Gavis et al., 2008).

Genetic manipulations that disrupt the localization and/or translation of these mRNAs provide compelling evidence that the specific subcellular distributions of these mRNAs are essential for their function in axial and germline development (Becalska and Gavis, 2009; Kugler and Lasko, 2009). Equally important is the role of RBPs in coordinating these processes. Indeed, the identification of specific regulatory factors has begun to shed light on the diversity of functions employed by these RBPs throughout development (Bandziulis et al.,
1989; Gamberi et al., 2006), and the ability to test the requirement of these proteins for developmental processes has been a major contribution of the *Drosophila* system.

### 1.5 Thesis Overview

A central feature of this thesis is the integration of structural, biochemical, genetic, and bioinformatic approaches to better understand how RBPs meet the diverse functional requirements of the cell. Since the precise roles of RBPs in regulating gene expression are likely determined by their RNA binding specificity, it is important to elucidate their RNA recognition properties as well as how they use these properties to regulate diverse RNA networks. The *Drosophila glorund* (*glo*) gene provides an ideal model for studying these interactions, as Glo is a member of the hnRNP F/H family of proteins, which function in nearly all aspects of RNA metabolism. In collaboration with the lab of Dr. Traci Hall (NIEHS), I carried out a structure/function analysis of Glo, which identified a remarkable mechanism whereby distinct RNA binding modes are used to affect specific functions during development in *Drosophila* (Chapters 2-4). These studies, accompanied by the *in silico* analysis of potential Glo RNA targets discussed in Chapter 5, offers one of the most in depth studies of hnRNP function to date, laying the foundation for addressing critical questions about the relationship of RNA binding to biological function.
CHAPTER 2:

THE DROSOPIILA hnRNP F/H HOMOLOG GLORUND USES TWO DISTINCT RNA-BINDING MODES TO DIVERSIFY TARGET RECOGNITION

THIS CHAPTER WAS ADAPTED FROM A PREVIOUSLY PUBLISHED PAPER (TAMAYO ET AL., 2017) AND REPRESENTS A COLLABORATION WITH TAKAMASA TERAMOTO AND TRACI HALL (NIEHS).
2.1 Introduction

Post-transcriptional gene regulation relies on the association of RNA with RNA-binding proteins (RBPs) that direct RNA processing, turnover, localization, and translation. Many RBPs have evolved to function in multiple aspects of RNA regulation and recognize numerous different target RNAs (Ascano et al., 2013; Gerstberger et al., 2014). These multifunctional RBPs must therefore specifically select and distinguish among their diverse target RNAs. Such multifunctionality is common among RBPs belonging to the heterogeneous nuclear ribonucleoprotein (hnRNP) family (Han et al., 2010). Although hnRNPs were first defined as a complex of proteins involved in packaging and processing nascent pre-mRNA transcripts (Dreyfuss et al., 1993), the identification of homologous proteins expanded the hnRNP family (Han et al., 2010). An increasing number of hnRNPs with general roles in pre-mRNA metabolism have now been shown to play additional, specialized roles in the regulation of specific transcripts (Piccolo et al., 2014). To understand how hnRNPs and other RBPs carry out their varied functions in vivo, it is essential to elucidate their RNA recognition properties as well as how they use these properties to regulate diverse RNA networks.

The hnRNP F/H family of proteins encompasses five mammalian homologs (hnRNP F, hnRNP H, hnRNP H′, hnRNP 2H9 and G-rich sequence factor 1 (GRSF-1)) that specifically recognize a stretch of guanine nucleotides called a G-tract using three quasi-RRMs (qRRMs) (Caputi and Zahler, 2001; Dominguez and Allain, 2006; Dominguez et al., 2010). The RBDs of hnRNP F/H proteins were denoted qRRMs because although they are structurally similar to the classical RRM, the aromatic residues that contact RNA in canonical RRMs are not as highly conserved and binding is instead mediated by residues located in three loops (Dominguez and Allain, 2006; Dominguez et al., 2010; Honore et al., 1995). Solution structures of the three qRRMs of human hnRNP F bound to a G-tract of three guanines show that the three qRRMs encage the G-tract in an identical manner, sequestering them in a single-stranded conformation (Dominguez et al., 2010). Based on these structures, it was proposed
that hnRNP F could regulate RNA metabolism by preventing G-rich RNA sequences from forming secondary structures (Dominguez et al., 2010).

The molecular mechanisms by which hnRNP F/H family members regulate RNA processing are still unclear, however. In some cases, hnRNP F/H proteins regulate alternative splicing by binding G-tracks near 5′-splice-site-donor sequences and preventing the binding of spliceosomal components to the pre-mRNA (Han et al., 2005). In other cases, hnRNP F/H members regulate polyadenylation site choice by blocking recruitment of cleavage stimulation factor (Veraldi et al., 2001). Finally, GRSF-1 has been shown to stimulate translation of influenza virus-encoded mRNAs by binding to sequences in their 5′UTRs (Kash et al., 2002). Post-transcriptional regulation by hnRNP F/H members is therefore complex, and these proteins might, depending on the RNA substrate, act at different stages of RNA metabolism through recognition of different sequence and/or structural motifs.

In Drosophila, the hnRNP F/H family is represented by a single protein, Glorund (Glo). Glo was first identified as a repressor of nanos (nos) mRNA translation through its interaction with a translational control element (TCE) in the nos 3′UTR. The TCE comprises two stem-loops, designated TCEII and TCEIII, joined to a base stem (TCEI) (Gavis et al., 1996; Kalifa et al., 2006). Previous mutational analyses of the TCE indicated that both the sequence and structure of the UA-rich stem of the TCEIII stem-loop are required for TCE function in vivo and Glo binding in vitro (Kalifa et al., 2006). Specifically, mutations in the double-stranded UA-rich motif that disrupt Glo binding or mutation of Glo itself abrogate TCE-mediated repression of nos during oogenesis (Forrest et al., 2004; Kalifa et al., 2006). However, like its mammalian counterparts, Glo has multiple functions suggesting additional mRNA targets. Glo is required for viability to adulthood (Kalifa et al., 2006), and in addition to nos regulation, Glo has recently been implicated in translational repression of nuclear-encoded mitochondrial respiratory chain complex transcripts (Gehrke et al., 2015). Glo also plays roles in ovarian nurse cell chromosome organization and dorsal-ventral patterning, most likely as a splicing regulator (Kalifa et al., 2009; Kalifa et al., 2006). How Glo distinguishes among its targets to exert the
necessary regulatory activity is not known.

To further probe the mechanistic bases for Glo’s multiple functions and, in particular, the role of RNA recognition, we took a structure-based functional approach. We determined the crystal structures of Glo’s three qRRMs and, based on these structures, identified two distinct modes of RNA recognition. We show that, like hnRNP F qRRMs, Glo qRRMs bind to single-stranded G-tracts and that mutation of loop residues equivalent to those in hnRNP F disrupt G-tract binding. We also identify a second, noncanonical RNA-binding interface present on the surface of Glo qRRM β-sheets and show that mutation of this interface disrupts recognition of the double-stranded UA-rich motif in the TCE. By engineering mutations in Glo that disable each RNA recognition mode \textit{in vitro} and evaluating the effect of these mutations on Glo function \textit{in vivo}, we demonstrate that regulation of a subset of Glo targets is mediated solely using the G-tract binding mode, whereas regulation of \textit{nos} unexpectedly requires both modes of recognition. Consistent with this finding, we identified a G-tract sequence within the TCE and show that higher affinity binding of Glo \textit{in vitro} and TCE function \textit{in vivo} require both the G-tract and the UA-rich stem. The use of two distinct RNA-binding modes to recognize different sets of target RNAs provides a molecular mechanism to explain the diverse functions of Glo and possibly other hnRNP F/H proteins. Moreover, it illustrates a general principle that RBPs may use different combinations of RNA recognition interfaces to regulate subsets of target RNAs that correlate with particular biological functions.

\textbf{2.2 Crystal Structures of Glo qRRMs Predict that Glo Recognizes G-tract RNA}

As a first step toward understanding how Glo interacts with \textit{in vivo} RNA targets, we determined the crystal structures of each of its three qRRMs: qRRM1, qRRM2 and qRRM3 (Figure 2.1; Table 2.1). Despite limited overall sequence similarity between Glo and its human homologs (24% amino acid identity/33% similarity), the residues in hnRNP F that mediate interaction with G-tract RNA are conserved in Glo (Dominguez et al., 2010; Kalifa et al., 2006). Comparison of the three crystal structures with the solution structures of hnRNP F qRRMs in
complex with single-stranded G-tract RNA indicated that the G-tract RNA interaction features are conserved in Glo (Figure 2.1). Thus, Glo qRRMs likely retain the G-tract RNA-binding ability of the hnRNP F/H family.
Figure 2.1. Crystal structures of Glo qRRM1, qRRM2 and qRRM3 domains identify RNA-binding motifs. (A) Schematic representation of the domain structure of Glo. Amino acid residues at domain boundaries are numbered, and the qRRM domains are color coded: qRRM1 (yellow), qRRM2 (cyan) and qRRM3 (orange). (B) Ribbon diagrams of the solution structure of hnRNP F qRRM1 bound to single-stranded G-tract RNA (PDB code 2KFY; Dominguez et al., 2010) and crystal structures of Glo qRRMs. The ribbon diagram of hnRNP F qRRM1 is shown in grey with a cartoon representation of bound G-tract RNA in yellow. Three critical G-tract binding residues in hnRNP F qRRM1 are displayed as stick models superimposed with transparent red space-filling spheres (R16, W20, and Y82); the structurally equivalent residues in Glo qRRM1 (yellow), qRRM2 (cyan), and qRRM3 (orange) are displayed similarly. N- and C-termini of proteins, 5´ and 3´ ends of G-tract RNA, and loops 1, 3, and 5, which contain residues that may interact with G-tract RNA, are indicated. (C) Conservation of G-tract RNA-binding residues in Glo. Comparison of the consensus G-tract binding motif sequence of hnRNP F qRRMs with corresponding residues in Glo qRRMs and hnRNP F qRRM1. Consensus residues important for stacking interaction with G-tract RNA (shown in B) are highlighted in red. Additional residues involved in G-tract RNA binding are colored magenta in the consensus sequence. (D) Ribbon diagrams of the crystal structures of classical U1A RRM bound to a fragment of U1 small nuclear RNA (PDB code 1URN; (Oubridge et al., 1994) and Glo qRRMs. The ribbon diagram of U1A RRM is shown in grey with a cartoon representation of bound U1 small nuclear RNA sequence in yellow. Critical residues within the RNP1 and RNP2 motifs of U1A are displayed as stick models superimposed with transparent green and blue space-filling spheres, respectively (Y13 and F56); the structurally equivalent residues in Glo qRRM1, qRRM2, and qRRM3 are displayed similarly. αC helices in qRRM1 and qRRM2 that occlude the RNP1 motif are colored red. N- and C-termini of proteins and 5´ and 3´ ends of the fragment of U1 small nuclear RNA are indicated; the 5´ end of the RNA is truncated in this rendering. (E) Conservation of classical RRM RNA-binding residues in Glo. Comparison of the consensus RNP motif sequences of classical RRMs with corresponding
residues in Glo qRRMs and U1A RRM. Consensus RNP1 and RNP2 residues important for RNA binding by classical RRMs are colored green and blue, respectively. See also Table 2.1.

Table 2.1: Crystallographic Summary of Glo qRRMs

<table>
<thead>
<tr>
<th></th>
<th>Glo qRRM1</th>
<th>Glo qRRM2</th>
<th>Glo qRRM3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>P2(_1)</td>
<td>P4(_{212})</td>
</tr>
<tr>
<td><strong>Cell Dimensions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>115.2, 27.7, 66.4</td>
<td>45.7, 29.0, 69.5</td>
<td>55.5, 55.5, 58.7</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.0, 109.4, 90.0</td>
<td>90.0, 106.8, 90.0</td>
<td>90.0, 90.0, 90.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00–1.54 (1.60–1.54)*</td>
<td>50.00–1.55 (1.58–1.55)*</td>
<td>50.00–1.99 (2.06–1.99)*</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>6.5 (11.5)</td>
<td>5.0 (25.7)</td>
<td>6.2 (22.5)</td>
</tr>
<tr>
<td>I/σI</td>
<td>20.2 (6.0)</td>
<td>34.6 (4.0)</td>
<td>32.4 (9.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>90.9 (70.3)</td>
<td>93.4 (70.7)</td>
<td>95.9 (86.4)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.3 (1.4)</td>
<td>3.4 (2.4)</td>
<td>12.5 (11.0)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>22.37–1.54</td>
<td>32.54–1.55</td>
<td>24.81–1.99</td>
</tr>
<tr>
<td>No. reflections</td>
<td>26,927</td>
<td>23,923</td>
<td>6,311</td>
</tr>
<tr>
<td>Rwork /Rfree (%)</td>
<td>17.3/20.6</td>
<td>19.3/23.9</td>
<td>22.0/26.0</td>
</tr>
<tr>
<td><strong>No. of Atoms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>1,380</td>
<td>1,495</td>
<td>685</td>
</tr>
<tr>
<td>Ligand/Ion</td>
<td>40</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>Water</td>
<td>230</td>
<td>127</td>
<td>33</td>
</tr>
<tr>
<td><strong>B-Factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>17.6</td>
<td>25.3</td>
<td>28.1</td>
</tr>
<tr>
<td>Ligand/Ion</td>
<td>37.8</td>
<td>–</td>
<td>48.5</td>
</tr>
<tr>
<td>Water</td>
<td>29.6</td>
<td>32.0</td>
<td>31.6</td>
</tr>
<tr>
<td><strong>RMS Deviations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
<td>0.011</td>
<td>0.010</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.881</td>
<td>1.070</td>
<td>1.230</td>
</tr>
<tr>
<td><strong>Ramachandran plot (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>99.4</td>
<td>97.8</td>
<td>98.7</td>
</tr>
<tr>
<td>Allowed</td>
<td>0.6</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td>PDB ID</td>
<td>5UZG</td>
<td>5UZM</td>
<td>5UZN</td>
</tr>
</tbody>
</table>
To evaluate whether Glo can bind to G-tract sequences, we performed electrophoretic mobility shift assays (EMSAs) and determined the dissociation constants ($K_d$) of each domain with single-stranded G-tract RNA (5'-AGGGA). Each qRRM domain bound independently to G-tract RNA (Figure 2.2), and qRRM1 and qRRM3 exhibited several-fold higher affinity than qRRM2 ($K_d=0.18$ μM and $0.17$ μM, respectively, vs $0.62$ μM). This is similar to the measurement of binding affinities for the qRRMs of hnRNP F, where qRRM2 bound to G-tract RNA with lower affinity than qRRMs 1 and 3 ($K_d=4.6$ μM vs $K_d=0.4$ μM; Dominguez et al. 2010). We were unable to quantitatively evaluate binding to G-tract RNA by full-length Glo protein due to poor expression. However, a truncated protein containing the two N-terminal qRRMs (Glo qRRM1,2) bound two 5'-AGGGA RNAs with an affinity intermediate to that of the individual domains ($K_d = 0.28$ μM; Figure 2.2 and data not shown; see Chapter 7). This finding is consistent with independent binding of each qRRM to a distinct RNA. We conclude that, like hnRNP F qRRMs, Glo qRRMs are each capable of interacting with G-tract RNA.
Figure 2.2. Glo qRRMs bind to a single-stranded G-tract sequence. Binding of purified recombinant Glo qRRM1, qRRM2, qRRM3, and qRRM1,2 proteins to radiolabeled G-tract (5´-AGGGA) RNA. Representative EMSA gels are shown, with data plotted below. The highest protein concentration (right-most lanes) was 3.1 µM and 2-fold serial dilutions were assayed. No protein was added to the reactions in the lanes marked ‘-’. The positions of unbound RNA (asterisk) and Glo qRRM:RNA complexes (arrowhead) are indicated. Three technical replicates of EMSAs were performed, and apparent $K_d$ values shown are mean ± SEM.
2.3 Crystal Structures of Glo qRRMs Suggest a Second, Distinct RNA Interaction Mode

Given the conservation of the G-tract RNA interaction, it is surprising that Glo was identified by its interaction with the double-stranded UA-rich motif in the nos TCEIII stem-loop (Kalifa et al., 2006). Therefore, we inspected the Glo qRRM crystal structures for residues that might mediate interaction with TCEIII. The three qRRMs adopt a core classical RRM fold ($\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$) with a four-stranded antiparallel $\beta$-sheet on one surface and two $\alpha$-helices on the opposite surface (Figure 2.1D). The classical RRM recognizes two RNA bases using hydrophobic residues within two conserved motifs, designated RNP1 and RNP2, that are located on the face of its $\beta$-sheet surface (Figure 2.1D, E). We therefore examined whether Glo could also use these motifs to interact with nos TCE RNA.

Conserved aromatic groups are present within the RNP1 motifs of Glo qRRM1 and qRRM2 (Y97 or F193, respectively) (Figure 2.1E). However, unlike classical RRMs, the aromatic RNP1 residues in Glo qRRM1 and qRRM2 are not available for RNA recognition. Glo qRRM1 and qRRM2 each possess a third C-terminal $\alpha$-helix ($\alpha_C$; Figure 2.1D) that forms hydrophobic interactions with the respective aromatic RNP1 residues. Mutation of either Y97 or F193 to alanine resulted in aggregated proteins, suggesting that these hydrophobic interactions are important for structural stability of the protein. In contrast, Glo qRRM3 lacks both an $\alpha_C$ helix and a conserved aromatic residue in its RNP1 motif (Figures 2.1D, E). Together, the crystal structures suggest that the RNP1 motifs of Glo may not interact with RNA.

Basic residues in all three Glo qRRMs (R52 in qRRM1, K149 in qRRM2, and H484 in qRRM3) replace the aromatic residues that are conserved in RNP2 motifs of classical RRMs (Figure 2.1E). These basic residues are exposed on the protein surface (Figure 2.1D), suggesting that they are available for RNA binding. We propose that these Glo RNP2 residues, which are conserved in hnRNP F, interact with RNA, but in a manner that is distinct from that of classical RRMs. Together, our structural analysis predicts that each of the Glo qRRMs has
two RNA-binding interfaces that differ from classical RRM: the G-tract binding loops identified in hnRNP F and a basic β-strand interface within the RNP2 motif. Moreover, the exposed basic residues in the RNP2 motifs suggest a possible mechanism for recognition of the UA-rich TCEIII stem-loop by Glo.

2.4 Glo Recognizes Two Different TCE Features

In order to assess whether Glo qRRMs use either of their RNA-binding interfaces to interact with the nos TCE, we first tested the ability of individual Glo qRRMs or the tandem domain Glo qRRM1,2 protein to bind TCE RNA by EMSA. In contrast to their ability to bind G-tract RNA, single qRRM domains did not bind stably to a TCE RNA probe (Figure 2.3; $K_d$>100 µM for qRRM1 and qRRM3; $K_d$>20 µM for qRRM2; data not shown). However, Glo qRRM1,2 did bind stably to the TCE, with two plateaus apparent in the binding curve (Figure 2.3B, C). These data fit best to a curve assuming two binding sites with higher ($K_d$=0.26 µM) and lower ($K_d$=13.5 µM) affinity components, suggesting that the TCE contains two Glo recognition sites with distinct properties. We therefore re-examined the sequences that comprise the complete TCE for a second Glo binding site in addition to the UA-rich TCEIII stem-loop. This analysis identified a G-tract embedded in the TCEI base stem (Figure 2.3A) that could potentially mediate Glo interaction. Notably, the higher affinity binding of Glo qRRM1,2 to the TCE ($K_d$=0.26 µM) is similar to the affinity of Glo qRRM1,2 for single-stranded G-tract RNA ($K_d$=0.28 µM).

To determine whether the TCEI GGG sequence contributes to TCE recognition by Glo qRRM1,2, we mutated it to GAG, while maintaining base pairing with a compensatory change on the 5´ strand (TCEGAG/UUC; Figure 3A). Glo qRRM1,2 bound to TCEGAG/UUC RNA with a $K_d$ of 8.9 µM, similar to the lower affinity binding to TCE RNA ($K_d$=13.5 µM), and curve fitting now assumed a single binding site (Figure 2.3B, C). Strikingly, higher affinity binding was absent, suggesting that this component could be assigned to recognition of the TCEI G-tract sequence and the lower affinity binding that remained was due to recognition of the UA-rich TCEIII stem-
loop. The GAG mutation alone eliminated higher affinity binding ($K_d = 18.6 \mu M$) (Figure 2.3A), and the compensatory mutation alone displayed both higher and lower affinity binding (Figure 2.3B). In addition, two mutants with an intact G-tract but predicted to disrupt TCEI stem base pairing, TCE$_{UAC}$ and TCE$_{GAC}$, retained both higher and lower affinity binding components (Figure 2.4C, D), indicating that the G-tract sequence is responsible for the higher affinity binding. These data also indicate that Glo can interact with the G-tract regardless of whether it is base-paired and suggest that, like hnRNP F (Dominguez et al. 2010; (Samatanga et al., 2013), Glo binding may promote a single-stranded configuration of G-tracts.

Figure 2.3. Glo recognizes two elements in the nos TCE that contribute higher and lower affinity components. (A) Predicted secondary structures of TCE and TCEI RNAs. RNA secondary structure prediction was calculated using Mfold (Zuker, 2003). TCEI, II, and III elements are indicated. The complete sequences of the transcribed RNAs are shown. The G-tract in TCEI is boxed and TCE$_{GAG/UUC}$ mutations are indicated in red. (B) Representative EMSA gels of purified recombinant Glo qRRM1,2 binding to TCE (top), TCE$_{GAG/UUC}$ (middle), or TCEI (bottom) RNAs. The highest Glo qRRM1,2 protein concentration (right-most lanes) was 100 μM for the TCE and TCE$_{GAG/UUC}$ binding assays and 144 μM for the TCEI binding assays with 2-fold serial dilutions. No protein was added to the reactions in the lanes marked ‘-‘. The positions of unbound RNA (asterisk) and Glo qRRM1,2:RNA complex (arrowhead)
RNA are indicated. (C) Superposition of plots for binding of Glo qRRM1,2 to TCE (red) or TCE_{GAG/UUC} (black) RNA. Higher affinity binding to the TCE is selectively disrupted by mutation of the G-tract sequence in the TCEI stem. Three technical replicates of EMSAs were performed, and apparent K_d values shown are mean ± SEM. (D) Superposition of plots for binding of Glo qRRM1,2 to TCE (red) or TCEI (dotted black) RNA. The TCEI stem alone retains the higher affinity binding component of the Glo-TCE interaction, but the lower affinity component is diminished. See also Figure 2.4.
Figure 2.4. Effect of TCE mutations on Glo binding. (A-E) Predicted secondary structures of TCE_{GAG} (A), TCE_{UUC} (B), TCE_{UAC} (C) TCE_{GAC} (D) and TCEI_{III} (E) RNAs are shown at left. Middle panels show representative EMSAs of binding of Glo qRRM1,2 to the corresponding TCE mutant RNA, with data plotted at right. The highest protein concentration (right-most lanes) was 100 µM and 2-fold serial dilutions were assayed. No protein was added to the binding reactions in the lanes marked '-'. The positions of unbound RNA (asterisk) and Glo qRRM1,2:RNA complexes (arrowhead) RNA are indicated. Three technical replicates of EMSAs were performed, and apparent $K_d$ values shown are mean ± SEM.

To further assign higher and lower affinity binding to the TCE recognition sites, we tested binding of Glo qRRM1,2 to RNAs comprising TCEI and TCEIII (TCEI_{III}) or only TCEI. Glo qRRM1,2 bound to TCEI_{III} RNA with higher ($K_d$=0.24 µM) and lower ($K_d$=27.4 µM) affinity components, similar to binding to the intact TCE (Figure 2.4E). As expected, TCEII, which is recognized by the embryonic repressor Smaug (Smg; see below), is not required for recognition by Glo. When the RNA was further truncated to the TCEI stem alone (Figure 2.3A), Glo qRRM1,2 bound with a $K_d$ of 0.17 µM, similar to the $K_d$ for the higher affinity site in TCE RNA (Figure 2.3B, D) whereas the lower affinity binding component was considerably reduced ($K_d$>87 µM). We therefore conclude that binding of Glo qRRM1,2 to the TCE includes two components: higher affinity binding ($K_d$=0.3 µM) to the TCEI G-tract sequence and lower affinity binding ($K_d$=13 µM) to the UA-rich TCEIII stem-loop.

2.5 Identification of Glo Residues that Mediate G-tract or TCE RNA Binding

Our crystallographic analysis suggested that Glo has two distinct RNA-binding interfaces within each of its qRRMs, and we hypothesized that each of the two binding interfaces confers one component of Glo’s dual RNA-binding specificity. If correct, the two RNA-binding activities should be separable by mutation of the respective interfaces. We therefore generated mutations in residues within the loop and RNP2 binding surfaces and
tested whether they differentially alter the ability of Glo to bind single-stranded G-tract (5´-AGGGA) and TCE RNAs in vitro.

We first mutated to alanine Glo W58 (qRRM1) and Y155 (qRRM2), which are equivalent to hnRNP F G-tract binding residues W20 (qRRM1) and Y120 (qRRM2), respectively (Figure 2.1B, C, Figure 2.5). Binding to G-tract RNA by Glo qRRM1,2\(^{W58A,Y155A}\) protein with mutations in both qRRMs was >200-fold weaker than binding by the wild-type Glo qRRM1,2 protein, suggesting that Glo uses these equivalent residues for G-tract interaction (Figures 2.5A, 2.6, and Table 2.2). Consistent with the behavior of individual qRRM domains, a single qRRM was sufficient to bind to the single-stranded G-tract RNA, as the individual W58A and Y155A mutations individually had little to no effect on G-tract binding by Glo qRRM1,2, and the Glo qRRM1,2 mutant with an intact qRRM1 (qRRM1,2\(^{Y155A}\)) bound with slightly higher affinity than the mutant with an intact qRRM2 (qRRM1,2\(^{W58A}\)) (Figures 2.5A, 2.6A, and Table 2.2).

Figure 2.5. Identification of Glo residues critical for G-tract or TCEIII recognition. Bar graphs of mean \(K_d\) values for binding of wild-type (WT) and mutant qRRM1,2 proteins to either TCE\(_{GAG/UUC}\), 5´-AGGGA, or TCEI RNAs determined from three technical replicates. Error bars indicate SEM. P-values were calculated for mutant binding to the TCE\(_{GAG/UUC}\) or 5´-AGGGA RNA relative to the WT protein using a one-way ANOVA, excluding the double mutants whose
binding was too weak to determine precisely and are reported as a lower limit of >60 µM. The p-values for qRRM<sup>R52A,K149A</sup> mutant binding to TCE<sub>GAG/UUC</sub> or TCEI RNA relative to the WT were calculated using an unpaired, two-tailed t-test with Welch’s correction for unequal variances, and the p-value for qRRM<sup>W58A,Y155A</sup> binding to 5'-AGGGA RNA relative to the WT was calculated using an unpaired, one-tailed t-test with Welch’s correction. P-values were <0.01 with the following exceptions: binding to 5'-AGGGA RNA by qRRM<sup>W58A,Y155A</sup> (p=0.024); qRRM<sup>Y155</sup> (p=0.996); and qRRM<sup>K149A</sup> (p=0.846) mutants; and binding to TCEI RNA by the qRRM<sup>R52A,K149A</sup> mutant (p=0.28). Representative EMSA gels and plots are presented in Figures 2.6 and 2.7. See also Table 2.2.
Figure 2.6. Mutation of loop residues selectively disrupts G-tract RNA binding.
Representative EMSAs of binding of Glo qRRM1,2 mutants to AGGGA (A) or TCEGAG/UUC (B) RNAs with schematic representations of the mutants above and data plotted below. The highest protein concentrations (right-most lanes) were 100 μM for TCEGAG/UUC assays and 50 μM for AGGGA assays, and 2-fold serial dilutions were tested. No protein was added to the binding reactions in the lanes marked ‘-’. The positions of unbound RNA (asterisk) and Glo qRRM1,2:RNA complexes (arrowhead) RNA are indicated. Three technical replicates of EMSAs were performed, and apparent K_d values shown are mean ± SEM. (C) Higher affinity binding of Glo to the TCE is mediated by G-tract binding residues. Representative plots for binding of Glo qRRM1,2^{W58A,Y155A} (red) or wild-type Glo qRRM1,2 to TCE (black) are superimposed. (D) Circular dichroism spectra of Glo qRRM1,2 wild-type (black) and Glo qRRM1,2^{W58A,Y155A} (red) proteins.

Table 2.2. Identification of Glo residues critical for G-tract or TCEIII recognition.

<table>
<thead>
<tr>
<th>Glo qRRM1,2</th>
<th>RNA motif recognized</th>
<th>Mutated domain</th>
<th>TCEGAG/UUC K_d (μM)</th>
<th>K_m</th>
<th>p-value</th>
<th>AGGGA K_d (μM)</th>
<th>K_m</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>G-tract, TCEIII</td>
<td>NA</td>
<td>8.9 ± 0.6</td>
<td>1.0</td>
<td></td>
<td>0.28 ± 0.009</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>W58A</td>
<td>G-tract</td>
<td>qRRM1</td>
<td>15.9 ± 0.3</td>
<td>1.8</td>
<td>0.0006</td>
<td>0.54 ± 0.04</td>
<td>1.9</td>
<td>0.007</td>
</tr>
<tr>
<td>Y155A</td>
<td>G-tract</td>
<td>qRRM2</td>
<td>14.5 ± 0.5</td>
<td>1.6</td>
<td>0.003</td>
<td>0.25 ± 0.02</td>
<td>0.9</td>
<td>0.996</td>
</tr>
<tr>
<td>W58A,Y155A</td>
<td>G-tract</td>
<td>qRRM1&amp;2</td>
<td>22.3 ± 0.6</td>
<td>2.5</td>
<td>0.0001</td>
<td>&gt;60</td>
<td>&gt;200</td>
<td>0.024</td>
</tr>
<tr>
<td>R52A</td>
<td>TCEIII</td>
<td>qRRM1</td>
<td>19.4 ± 0.5</td>
<td>2.2</td>
<td>0.0001</td>
<td>0.59 ± 0.03</td>
<td>2.1</td>
<td>0.002</td>
</tr>
<tr>
<td>K149A</td>
<td>TCEIII</td>
<td>qRRM2</td>
<td>29.7 ± 1.9</td>
<td>3.3</td>
<td>0.0001</td>
<td>0.22 ± 0.03</td>
<td>0.8</td>
<td>0.846</td>
</tr>
<tr>
<td>R52A,K149A</td>
<td>TCEIII</td>
<td>qRRM1&amp;2</td>
<td>&gt;60</td>
<td>&gt;7</td>
<td>0.008</td>
<td>0.93 ± 0.10</td>
<td>3.3</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

1K_d values for binding of wild-type (WT) and mutant qRRM1,2 proteins to either TCEGAG/UUC, 5’-AGGGA, or TCEI RNAs are shown as mean ± SEM from three technical replicates.

2P-values were calculated for mutant binding to the TCEGAG/UUC or 5’-AGGGA RNA relative to the WT protein using a one-way ANOVA, excluding the double mutants whose binding was too weak to determine precisely and are reported as a lower limit of >60 μM. The p-values for qRRM1,2^{R52A,K149A} mutant binding to TCEGAG/UUC or TCEI RNAs relative to the WT were calculated using an unpaired, two-tailed t-test with Welch’s correction for unequal variances, and the p-value for qRRM1,2^{W58A,Y155A} binding to 5’-AGGGA RNA relative to the WT was calculated using an unpaired, one-tailed t-test with Welch’s correction.
Figure 2.7. Mutation of basic RNP2 residues selectively disrupts TCEIII UA-rich motif binding. Representative EMSAs of binding of Glo qRRM1,2 mutants to TCEGAG/UUC (A) or AGGGA (B) RNAs with schematic representations of the mutants above and data plotted below. The highest protein concentrations (right-most lanes) were 100 μM for TCEGAG/UUC assays and 50 μM for AGGGA assays, and 2-fold serial dilutions were tested. No protein was added to the binding reactions in the lanes marked ‘-’. The positions of unbound RNA (asterisk) and Glo qRRM1,2:RNA complexes (arrowhead) RNA are indicated. Three technical replicates of EMSAs were performed, and apparent $K_d$ values shown are mean ± SEM. (D) Circular dichroism spectra of Glo qRRM1,2 wild-type (black), Glo qRRM1,2$^{R52A,K149A}$ (blue) proteins.
We next tested whether mutation of these residues affects the binding of Glo to TCE RNA. For these experiments, we used primarily a TCE RNA lacking the embedded G-tract sequence, TCE_{GAG/UUC}, to probe the lower affinity interaction with the UA-rich TCEIII stem-loop. Mutation of either one or both of the G-tract binding residues had only a small effect (<2.5 fold) on binding of Glo qRRM1,2 to TCE_{GAG/UUC} RNA, indicating that these residues are not critical for recognition of the TCEIII UA-rich motif (Figures 2.5B, 2.6B, and Table 2.2). We also tested binding of Glo qRRM1,2^{W58A,Y155A} to wild-type TCE RNA. Whereas wild-type Glo qRRM1,2 protein exhibited both lower and higher-affinity binding to the TCE, Glo qRRM1,2^{W58A,Y155A} protein showed only the lower affinity component, indicating that the mutations disrupted the higher-affinity binding component attributed to recognition of the embedded TCEI G-tract sequence (Figure 2.6C). This behavior is similar to that observed for Glo qRRM1,2 binding to TCE_{GAG} RNAs (Figures 2.3C, 2.4A) and consistent with the idea that the Glo qRRM loop residues recognize both single-stranded G-tract RNA and the G-tract present in the TCEI base stem.

We next focused on the basic residues in Glo’s divergent RNP2 motif as potential candidate residues for recognition of the UA-rich TCEIII stem-loop. Mutation of both RNP2 residues (R52A, K149A) in Glo qRRM1,2 (Figures 2.1D, 2.1E, and 2.7) greatly diminished binding to TCE_{GAG/UUC} RNA by more than 7-fold (a lower limit due to the protein solubility at high concentration) as compared to the wild-type protein (Figures 2.5B, 2.7A, and Table 2.2). Moreover, these mutations preferentially affected recognition of the UA-rich TCEIII stem-loop, since strong binding to single-stranded G-tract RNA and the TCEI base stem was retained (Figures 2.5C, 2.7B, and Table 2.2). Together, these results demonstrate that Glo has two different RNA recognition modalities and uses these to recognize two distinct features of the TCE. Residues within the qRRM loops are critical for G-tract recognition in the TCEI stem, and basic residues in the divergent RNP2 motifs are important for recognition of the UA-rich TCEIII stem-loop.
2.6 Translational Repression of nos Requires Both Modes of RNA Recognition

The ability of Glo to recognize both the G-tract and the double-stranded UA-rich motif in the TCE suggests that Glo employs both modes of RNA recognition to regulate nos in vivo. We therefore sought to determine whether mutations that preferentially disrupt TCEIII stem-loop or TCEI G-tract RNA binding by Glo in vitro affect translational repression of nos in vivo (Figure 2.8A-C). Nos protein is required in the posterior of the early embryo for abdominal development but must be excluded from the anterior where it inhibits head development (Gavis and Lehmann, 1992; Wang and Lehmann, 1991). This distribution is achieved through the selective translation of nos mRNA localized at the posterior of the embryo and the repression of nos mRNA remaining within the bulk cytoplasm (Gavis and Lehmann, 1992, 1994). Localization of nos occurs during late stages of oogenesis, and Glo acts at this time to establish the repressed state of nos (Kalifa et al., 2006). Because translation of nos is restricted to the posteriorly localized transcripts, only a small amount of Nos protein accumulates in late-stage oocytes (Figure 2.8D, E; (Forrest et al., 2004; Forrest and Gavis, 2003). By contrast, Nos protein levels are dramatically increased when repression of unlocalized nos is abolished by mutation of glo (glo−; Figure 2.8D, E).
Figure 2.8. Glo uses both modes of RNA recognition for translational repression of nos.

(A) Schematic representation of the wild-type GFP-Glo protein, with the N-terminal GFP tag (green) and qRRM1 (yellow), qRRM2 (cyan), and qRRM3 (orange) indicated. (B, C) GFP-Glo proteins with alanine substitutions that were assayed for function in vivo. G-tract binding mutations are shown in red (B), TCEIII binding mutations are shown in dark blue (C). All gfp-glo transgenes were inserted into the same chromosomal location. Two independent lines of each transgene were tested for their ability to rescue the lethality of the glo null mutant and in all cases, the independent lines behaved similarly. (D, E) Immunoblot analysis of Nos protein in extracts of late ovaries from glo mutant females expressing GFP-Glo proteins with mutations in either (D) G-tract binding residues (corresponding to panel C) or (E) TCEIII binding mutations (corresponding to panel D). Kinesin heavy chain (Khc) was used as a loading control.
We expressed wild-type GFP-Glo or mutant GFP-Glo proteins in glo mutants and tested their ability to repress nos translation by assaying Nos protein levels in ovaries enriched for late-stage oocytes, referred to as late ovaries (Andrews et al., 2011). GFP-Glo proteins with a single alanine substitution in either a residue that affects G-tract binding (qRRM1: W58A; qRRM2: Y155A; qRRM3: Y490A) or a residue that affects TCEIII binding (qRRM1: R52A; qRRM2: K149A; qRRM3: H484A), behaved identically to wild-type GFP-Glo in their ability to confer translational repression of nos (Figure 2.8D, E). Similarly, GFP-Glo proteins with alanine substitutions in both qRRM1 and qRRM2 (G-tract binding: Glo\textsuperscript{W58A, Y155A}; TCEIII binding: Glo\textsuperscript{R52A, K149A}) or qRRM1 and qRRM3 (G-tract binding: Glo\textsuperscript{W58A, Y490A}; TCEIII binding: Glo\textsuperscript{R52A, H484A}) retained the ability to repress nos (Figure 2.8D, E). In contrast, Nos protein levels were dramatically increased in late ovaries from glo mutant females expressing either the G-tract binding triple mutant (GFP-Glo\textsuperscript{W58A, Y155A, Y490A}, abbreviated Glo\textsuperscript{WYY}) or the TCEIII binding triple mutant (GFP-Glo\textsuperscript{R52A, K149A, H484A}, abbreviated Glo\textsuperscript{RKH}) protein (Figure 2.8D, E). Immunoblot analysis showed that both GFP-Glo\textsuperscript{WYY} and GFP-Glo\textsuperscript{RKH} were expressed at levels comparable to proteins that do confer nos repression (Figure 2.9), indicating that their failure to repress nos translation did not result from reduced protein expression.
Figure 2.9. Expression of GFP-Glo transgenes. (A) Schematic representation of GFP-Glo protein, colored as in Figure 4, with a summary map of the alanine substitutions that were assayed for function in vivo. (B, C) Immunoblot analysis of Glo in extracts of late ovaries from transgenic females expressing wild-type GFP-Glo (GloWT) or GFP-Glo proteins with either G-tract binding mutations (B) or TCEIII binding mutations (C). The blot was probed with anti-Glo antibody to detect both endogenous Glo and GFP-Glo. Relative abundance of GFP-Glo proteins, indicated below the blots, was quantified by normalization to endogenous Glo in the same sample and to GloWT.

Together, these results indicate that, consistent with the in vitro binding studies, both modes of RNA recognition are required for translational repression of nos during oogenesis. Because Glo binds with only modest affinity to TCEIII in vitro and the qRRM1,2R52A,K149A mutation that abolishes this binding has an additional, albeit minor, effect on TCEI binding, we considered an alternative model in which another protein, instead of Glo, binds to the TCEIII UA-rich motif to regulate nos. This model predicts that: 1) TCEI G-tract mutations would eliminate binding of Glo to the TCE; 2) binding of Glo to the G-tract would be the same as binding to the full TCE; and 3) the GloRKH mutation, which selectively abrogates binding to the TCEIII UA-rich stem, would have only a minor effect on Glo function in vivo compared to the GloWYY mutation, which selectively abrogates binding to the TCEI G-tract. None of these predictions is borne out. Furthermore, the binding specificity might well be higher for the full-length proteins in vivo. Thus, although we cannot entirely rule out the possibility that only G-tract binding is relevant to Glo regulation of nos translation in vivo, our data best support a dual recognition model for Glo interaction with the TCE.

2.7 TCE Function in vivo Requires Both Glo Binding Motifs

To determine whether the requirements for TCE function are the same as or different from the requirements for recognition of the TCE by Glo, we generated a series of mutations
that alter either the TCEI G-tract, the TCEIII UA-rich motif, or both in an otherwise wild-type nos transgene (Figure 2.10A). These TCE mutants were then assayed for their ability to repress nos translation by analyzing Nos protein levels in late ovaries from transgenic females. Mutation of either the G-tract (GAG) or TCE UA-rich stem (IIA) resulted in a modest increase in Nos protein as compared to the wild-type nos transgene, indicating that both motifs contribute to translational repression of nos (Figure 2.10B). Consistent with this, the TCEGAG.IIIA double mutation had an additive effect, resulting in a dramatic increase in Nos protein levels (Figure 2.10B). We therefore conclude that nos translational regulation depends upon both the UA-rich and G-tract motifs in the nos TCE and their interaction with Glo qRRMs.

**Figure 2.10.** TCE function *in vivo* requires both Glo recognition sites. (A) Predicted TCE secondary structure, with the TCEI G-tract and TCEIII UA-rich motifs indicated by boxes. Nucleotide changes creating the GAG, IIA, and SRE mutations are indicated. (B) Immunoblot analysis of Nos protein in extracts of late ovaries from females expressing either a wild-type nos transgene (WT) or nos transgenes with the GAG, IIA, or GAG.IIIA mutation as indicated. Nos protein levels were normalized to the Khc loading control. The mean value ± SEM for the
relative abundance of Nos shown below each lane was determined from three independent experiments. A one-way ANOVA showed no significant difference. (C) Representative cuticular phenotypes of embryos from females expressing either the WT or the indicated TCE mutant transgenes. Arrowheads indicate the extent of the head skeleton formed, which reaches the third thoracic segment in WT. Head structures in the TCE\textsubscript{GAG,IIIA} cuticle are foreshortened and malformed (reduced classification in D); no head structures are visible in the TCE\textsubscript{GAG,IIIA,SRE} cuticle (absent classification in D). (D) Quantification of the fraction of embryos exhibiting defects in the development of head structures for the TCE mutants shown in (C). Values represent the average percentages obtained from analysis of two independent lines (n>200 embryos/line) (E) Quantitative RT-PCR of transgenic nos RNA. Values are the percent mean number of transcripts ± SEM of three technical replicates from each of three biological replicates.

The ultimate physiological consequence of the failure to repress nos is the suppression of head development (Gavis and Lehmann, 1992, 1994). This defect can be easily visualized in the larval cuticle, providing a sensitive, cumulative readout of nos regulation in the oocyte and early embryo (Dahanukar and Wharton, 1996; Gavis et al., 1996; Smibert et al., 1996). The repression of unlocalized nos that is initiated by Glo during oogenesis is maintained in the early embryo by the SAM-domain protein Smg, which binds to a motif in TCE stem-loop II called the Smg Recognition Element (SRE; Dahanukar et al., 1999; Smibert et al., 1999; Smibert et al., 1996). To evaluate the relative contributions of the TCEI G-tract, TCEII UA-rich motif, and SRE to nos regulation, we assayed the effects of mutations in these sequences, individually and in combination, on embryonic head development.

Mutation of either the TCE G-tract (TCE\textsuperscript{GAG}), TCE UA-rich stem (TCE\textsubscript{IIIA}), or SRE (TCE\textsubscript{SRE}) individually did not significantly affect head development (Figure 2.10C, D), indicating that the elevated Nos protein levels observed by immunoblot analysis are not sufficient to suppress the formation of anterior structures. Strikingly, mutation of both Glo
recognition motifs (TCE_{GAG,IIIA}) was sufficient to disrupt head development, even though Smg binding was not disrupted (Figure 2.10C, D). Consistent with the requirement for Glo and Smg, however, the most severe phenotype was produced by combining the GAG, IIIA, and SRE\textsuperscript{−} mutations (TCE_{GAG,IIIA,SRE\textsuperscript{−}}), with over 95% of embryos exhibiting loss of head structures (Figure 2.10D). Similar effects on head development were observed when these transgenes were introduced into nos mutants, indicating that the increase in nos activity is not due simply to increased gene dosage (data not shown). Furthermore, all transgenes examined produce similar levels of RNA, as indicated by quantitative RT-PCR analysis (Figure 2.10E). Therefore, differences in the behavior of the transgenes are unlikely to result from differences in the levels of expression or stabilities of the RNAs. Finally, all transgenes behaved similarly in oskar (osk) mutants, where nos localization is abolished, confirming that the assay monitors de-repression of unlocalized nos RNA (data not shown). Taken together, these results confirm the physiological relevance of the in vitro binding data and establish a critical role for both the TCEI G-tract and TCEIII UA-rich motif in TCE function during oogenesis. Moreover these results further demonstrate that translational repression of unlocalized nos during oogenesis is critical to achieve the spatially restricted synthesis of Nos needed for embryonic patterning.

2.8 Glo Functions During Mid-oogenesis Are Mediated Solely by the G-tract Binding Mode

Like many hnRNPs, Glo has diverse biological functions. Indeed, the pleiotropy of the glo mutant phenotype, which includes dorsal-ventral and nurse cell nuclear morphology defects, suggests that Glo acts at different stages of oogenesis to regulate RNAs in addition to nos. Although the specific target RNAs whose regulation is important for these processes are not known, the identification of mutations that discriminate the two Glo binding modes in vivo allowed us to investigate how Glo recognizes the RNAs relevant for these functions.

In wild-type Drosophila ovaries, the nurse cell chromosomes initially exhibit a polytene morphology, but subsequently become dispersed throughout the nucleus (Dej and Spradling,
This chromosomal dispersion fails to occur in glo mutant ovaries (Kalifa et al., 2009; see also Figure 2.11). In addition, Glo indirectly regulates the localization and translation of gurken (grk) mRNA, which is required for dorsal fate specification. As a result, the dorsally located respiratory appendages of the egg shell are broad and often laterally expanded in glo mutants as compared to wild-type (Kalifa et al., 2009). To better understand how Glo interacts with the RNAs required for these processes, we assayed the functionality of mutant GFP-Glo proteins during mid-oogenesis (Figure 2.11). GFP-Glo proteins carrying a single mutation in residues that affect either G-tract or TCEIII binding behaved like wild-type GFP-Glo and completely rescued the glo mutant chromosome organization, grk localization, and dorsal appendage defects (Figures 2.11B, 2.11C, 2.11F, and 2.11). Similarly, expression of GFP-Glo proteins with mutations in both qRRM1 and qRRM2 (G-tract binding: GloW58A,Y155A; TCEIII binding: GloR52A,K149A) or qRRM1 and qRRM3 (G-tract binding: GloW58A,Y490A; TCEIII binding: GloR52A,H484A) also rescued these defects (Figure 2.12). Strikingly, while expression of the GFP-GloRKH protein fully rescued the chromosome organization and grk regulation defects, expression of the GFP-GloWYY protein failed to rescue either of these glo mutant phenotypes (Figure 2.11D-F). Similarly to glo mutants, the nurse cell chromosomes of ovaries expressing GFP-GloWYY remained polytene throughout oogenesis (Figure 2.11). In addition, grk was mislocalized along the anterior cortex (Figure 2.11D) and the resulting eggs exhibited abnormal dorsal appendages (Figure 2.11F). We therefore conclude that regulation of Glo target RNAs required for chromosome organization and grk regulation specifically require the G-tract RNA binding mode whereas the TCEIII binding mode is dispensable for these Glo functions. Furthermore, the behavior of the qRRM1,2 (Glow58A,Y155A) and qRRM1,3 (Glow58A,Y490A) mutant proteins indicate that RNA binding by a single qRRM is sufficient for regulating these targets.
Figure 2.11. Regulation of chromosome dispersion and dorsal-ventral patterning are mediated solely by the G-tract binding mode of Glo. (A) Schematic representation of the GFP-Glo protein with a summary map of the alanine substitutions shown in Figure 2.8B, C and analyzed in (B-F). (B-E) Confocal sections of a glo mutant egg chamber (B) and glo mutant egg chambers expressing GFP-Glo (B), GFP-Glo\textsuperscript{WYY} (C), or GFP-Glo\textsuperscript{RKH} (D). Nurse cells (nc) and oocyte (oo) are indicated. grk mRNA (red) is detected by smFISH, nurse cell and follicle cell nuclei (blue) are stained with DAPI. Egg chambers are oriented with anterior to the left, dorsal up. Arrows indicate mislocalized grk mRNA. Scale bar=10 µM. (F) Quantification of the fraction of eggs laid by glo mutant females expressing GFP-Glo, GFP-Glo\textsuperscript{WYY}, or GFP-Glo\textsuperscript{RKH} that show dorsal appendage defects, including shortened, laterally expanded, fused, and/or missing appendages. See also Figures 2.12 and 2.13.
Figure 2.12. The Glo G-tract binding mode is required for Glo function in nurse cell chromosome dispersion. (A) Schematic representation of GFP-Glo protein, colored as in Figure 2.8, with a summary map of the alanine substitutions that were assayed for function in vivo. (B,C) Confocal images of glo mutant egg chambers expressing GFP-Glo with either G-tract binding (B) or TCEIII binding (C) mutations. Nuclei were stained with DAPI (blue) and the actin cytoskeleton was stained with phalloidin (green). The nurse cells (nc) and oocyte (oo) are indicated. Scale bar=20 µm. (D, E) Quantification of the percentage of eggs from G-tract binding mutants (B) or TCEIII binding mutants (D) with dorsal appendage defects, including shortened, laterally expanded, fused, and/or missing appendages.

2.9 Discussion

RNA-binding proteins are often modular, with multiple RNA-binding domains that in combination can increase both specificity and affinity (Lunde et al., 2007). Our structure/function analysis of Glo uncovers another layer of complexity by showing that a single RNA-binding domain, the Glo qRRM, encompasses two discrete RNA-binding interfaces. These binding surfaces allow Glo to specifically recognize two very different RNA substrates, single-stranded G-tract RNA and a structured UA-rich motif. By assaying the effect of mutations that preferentially disrupt each recognition mode in vitro on Glo function in vivo, we show that regulation of a subset of Glo’s targets in vivo is mediated solely using the G-tract binding mode whereas regulation of nos requires both modes of recognition. The discovery of distinct RNA-binding modes within a single RNA-binding domain suggests a mechanism for the diversification of target RNA recognition that expands the functional repertoire of individual RNA-binding proteins.

Recognition of G-tracts by qRRMs differs from core RNA recognition by classical RRM s in utilization of loop residues rather than β-sheet residues. Moreover, β-sheet basic residues in Glo qRRMs have diverged from the canonical RNP1/RNP2 interface of the
classical RRM s. Analysis of both protein and RNA mutations confirmed that these residues endow Glo with a second RNA-binding mode that interacts with the double-stranded UA-rich motif in nos, thereby diversifying the repertoire of Glo target RNAs and its biological functions. Interestingly, these basic RNP2 residues are conserved in mammalian hnRNP F/H proteins, and a recent study using crosslinking and immunoprecipitation coupled with high-throughput sequencing demonstrated that hnRNP F often bound G-tract motifs near UA-rich regions (Huelga et al., 2012). It is therefore possible that dual modes of recognition also impart functional versatility on the human proteins.

Recognition of the nos TCE by Glo requires at least two qRRMs, and both Glo binding modes are required for regulation of nos in vivo. In contrast, recognition of targets involved in nurse cell chromosome dispersion and dorsal-ventral patterning requires binding of a single qRRM to G-tract motifs. Thus, it is the combination of the two binding modes within the three qRRMs that confers multifunctionality at the level of target RNA selection and, potentially at the level of regulatory function. Glo qRRM1,2 protein binds with >50-fold higher affinity to the TCEI G-tract than to the TCEIII UA-rich motif in vitro, suggesting that G-tract binding dominates stable TCE interaction whereas TCEIII recognition dominates target specificity. Moreover, the finding that G-tract or TCEIII binding residues must be mutated in all three qRRMs to abolish nos repression suggests that, unlike other proteins containing RRM s, the G-tract or TCEIII motif interaction may not map to individual domains.

In addition to discriminating among RNA targets, the use of different binding modes may influence how Glo regulates those targets. For example, the engagement of Glo with both TCE motifs may be required for the recruitment of a larger repressor complex and/or for interactions with components of the translation machinery. HnRNP F/H proteins are best known for their roles in the regulation of alternative splicing by binding to G-tracts, and G-tract recognition by single hnRNP F qRRMs is sufficient for alternative splicing of Bcl-x (Dominguez et al., 2010). We showed previously that Glo forms a complex with two Drosophila splicing factors, Half-pint (Hfp)/PUF68 and Hrp48 (Kalifa et al., 2009). Mutations in glo, hfp, or hrp48
cause similar defects in dorsal-ventral patterning and nurse cell chromosome dispersion, suggesting that Glo might collaborate with Hfp and Hrp48 to regulate the splicing of transcripts needed for these processes. Our finding that the G-tract binding mode alone is required for Glo function in dorsal-ventral patterning and chromosome dispersion supports this idea. Interaction of Glo with Hfp and Hrp48 is RNA-independent and although GFP-Glo$^{WYY}$ cannot bind to G-tracts, it retains the ability to bind Hfp and Hrp48 as assayed by coimmunoprecipitation (Figure 2.13). Thus, it is possible that Glo can only bind to G-tracts and not to TCEIII-like motifs when complexed with Hfp and Hrp48, thereby targeting the relevant Glo regulatory function to the proper targets.

**Figure 2.13. Glo mutant proteins interact with Hfp and Hrp48.** (A) Schematic representation of GFP-Glo protein, colored as in Figure 2.8, with a summary map of the alanine substitutions that were assayed for function *in vivo*. Anti-GFP immunoprecipitation of extracts from ovaries expressing either wild-type GFP-Glo (Glo$^{WT}$) or the GFP-Glo G-tract binding triple mutant (Glo$^{WYY}$) or the GFP-Glo TCEIII binding triple mutant (Glo$^{RKH}$). Total extract and immunoprecipitates were analyzed by immunoblotting with antibodies to Hfp and Hrp. Ovaries expressing an unrelated RNA-binding protein, MCP-GFP, were used as a negative control.
Glo is required for viability to adulthood and although the relevant target RNAs are not known, we found that, as for nos repression, both the Glo G-tract and TCEIII binding modes are essential (See Chapter 3). This suggests that Glo similarly recognizes other RNAs with bipartite motifs in addition to nos (See Chapter 4 and Chapter 5). As other RNA targets of Glo are identified, computational analyses should search for both G-tract and TCEIII-like motifs. Analysis of how the new targets are regulated and what accessory factors are engaged will deepen our understanding of how recognition of G-tract or TCEIII-like elements influences RNA metabolism.
CHAPTER 3:

FURTHER PROBING THE RNA-BINDING REQUIREMENTS OF GLO IN VIVO
3.1 Introduction

The diversity of functions of RNA-binding proteins (RBPs) would suggest a correspondingly large diversity in the structures that are responsible for RNA recognition. However, most RBPs are built from relatively few RNA binding domains (RBDs). Instead, it is generally thought that by combining these motifs in various structural arrangements, RBPs can recognize RNA with the affinity and selectivity that is required to bind specific sequences, while retaining the versatility required to regulate numerous RNAs. Structural biology has provided the molecular details of how individual domains recognize RNA, but many of these proteins require multiple copies of these domains in order to function. It is therefore important to understand how multiple RBDs bind RNA, as well as how they are used to dictate biological function.

The *Drosophila* hnRNP F/H homolog Glorund (Glo) was previously identified as a repressor of *nanos* (*nos*) mRNA translation through its interaction with both a G-tract and structured UA-rich motif in the *nanos* 3’ untranslated region (*nos* 3’UTR) (Kalifa et al., 2006; Tamayo et al., 2017). Like its mammalian counterparts, however, Glo has multiple functions, suggesting additional mRNA targets. Glo is required for viability to adulthood (Kalifa et al., 2006) and in addition to *nos* regulation, Glo also plays roles in ovarian nurse cell chromosome organization and dorsal-ventral patterning, most likely as a splicing regulator (Kalifa et al., 2009; Kalifa et al., 2006).

Our discovery that Glo contains three quasi-RNA recognition motifs (qRRMs) that each encompass two discrete RNA-binding interfaces for binding both G-tract and UA-rich motifs (Tamayo et al., 2017) raised the question of why multiple qRRMs are required to stably bind the translational control element of (TCE) of *nos* *in vitro* (See Section 2.4). It seemed possible, for example, that both the TCEI G-tract and TCEIII UA-rich stem could be bound by a single qRRM. Alternatively, multiple qRRMs could be required to recognize the noncontiguous G-tract and UA-rich motifs within the TCE. Indeed, the finding that G-tract or TCEIII binding
residues must be mutated in all three qRRMs to abolish nos repression suggests that these interactions may not be relegated to individual domains.

To further probe the RNA-binding requirements of Glo in vivo, we expanded our library of GFP-Glo mutants described in Chapter 2 to include variants with mutations in both TCEI G-tract and TCEIII UA-rich binding residues. We show that, consistent with our in vitro binding studies, regulation of nos requires at least two qRRMs, and that each qRRM confers one component of Glo’s dual RNA-binding modes. Similarly, the RNA binding requirements of Glo for viability involves multiple qRRMs, although in this case the relevant target RNAs may require all three qRRMs for proper regulation. This is in contrast to the regulation of Glo targets necessary for chromosome dispersion and dorsal-ventral patterning, which require only one qRRM to be available for G-tract recognition. Thus, our findings show how multiple RBDs are used to specify biological function and highlight their importance for achieving the diverse functional requirements of the cell.

3.2 Translational Repression of nos Requires Binding by Two qRRMs in vivo

The requirement for multiple qRRMs to stably interact with the TCE in vitro, suggested that multiple qRRMs are required to regulate nos in vivo. To examine this, we first tested whether mutations that disrupt the ability of individual qRRMs to interact with both UA-rich and G-tract sequences would affect TCE recognition in vivo (Figure 3.1). GFP-Glo proteins with mutations affecting a single qRRM (qRRM1: R52A,W58A; qRRM2: K149A,Y155A; qRRM3: H484A,Y490A) behaved similarly to wild-type GFP-Glo in their ability to interact with nos by coimmunoprecipitation, regardless of which qRRM was mutated (Figures 3.1B, 3.1C, and 3.1D). Consistent with this result, Nos protein levels were not affected in these mutants (Figure 3.1E). By contrast, nos could not be detected in immunoprecipitates from Glo proteins containing mutations in multiple qRRMs (qRRM1/2: R52A,W58A, K149A,Y155A – abbreviated RWKY; qRRM1/3: R52A,W58A,H484A,Y490A – abbreviated RWHY; qRRM2/3: K149A,Y155A, H484A,Y490A – abbreviated KYHY). As expected, Nos protein levels were
dramatically increased in these mutants, indicating that they are unable to repress *nos* translation (Figure 3.1E). Immunoblot analysis showed that these GFP-Glo transgenes are expressed at levels comparable to proteins that do confer *nos* repression (Figures 3.2B and 3.2D) indicating that their failure to repress *nos* translation did not result from reduced protein expression. Taken together, these results indicate that, consistent with the previous *in vitro* binding studies (Tamayo et al., 2017), at least two qRRMs are required for TCE recognition and translational repression of *nos* in vivo. Furthermore, they demonstrate that the three qRRMs function redundantly to regulate *nos*, since any two qRRMs are capable of regulating *nos* translation *in vivo*.

![Figure 3.1. Translational Repression of nos Requires Binding by Two qRRMs in vivo.](image)

(A) Schematic representation of the wild-type GFP-Glo protein, with the N-terminal GFP tag (green) and qRRM1 (yellow), qRRM2 (cyan), and qRRM3 (orange) indicated. (B) GFP-Glo
proteins with alanine substitutions that were assayed for function in vivo. G-tract binding mutations are shown in red. UA-rich TCEIII binding mutations are shown in dark blue. All gfp-glo transgenes were inserted into the same chromosomal location. (C) RT-PCR to detect nos or tub mRNA co-immunoprecipitated from ovaries expressing GFP-Glo mutant proteins (corresponding to panel B; See also text). Total RNA from the extracts used for immunoprecipitation was used as a positive control for the RT-PCR reaction (T). Reactions were performed in the presence (+RT) or absence (~RT) of reverse transcriptase. (D) Immunoblot of anti-GFP immunoprecipitates shown in (C), demonstrating efficient isolation of GFP-Glo mutant proteins. (E) Immunoblot analysis of Nos protein in extracts of late ovaries from glo mutant females expressing GFP-Glo mutant proteins (Corresponding to panel B; See also text). Kinesin heavy chain (Khc) was used as a loading control. (RWKY155: R52A,W58A,K149A,Y155A; RWHY: R52A,W58A,H484A,Y490A; KYHY: K149A,Y155A,H484A,Y490A).
**Figure 3.2. Expression of GFP-Glo transgenes.** (A) Schematic representation of the wild-type GFP-Glo protein, colored as in Figure 3.1. (B,C) GFP-Glo proteins with alanine substitutions that were assayed for function in vivo. G-tract binding mutations are shown in red. UA-rich TCEIII binding mutations are shown in dark blue. All gfp-glo transgenes were inserted into the same chromosomal location. (D,E) Immunoblot analysis of Glo in extracts of ovaries from transgenic females expressing wild-type GFP-Glo or GFP-Glo mutant proteins (Corresponding to panels B and C). The blot was probed with anti-Glo antibody to detect both endogenous Glo and GFP-Glo. Relative abundance of GFP-Glo proteins, indicated below the blots, was quantified by normalization to endogenous Glo in the same sample and to wild-type GFP-Glo. (RKY155: R52A,W58A,K149A,Y155A; RWHY: R52A,W58A,H484A,Y490A; KYHY: K149A,Y155A, H484A,Y490A); RWKY490: R52A,W58A,K149A,Y490A; RWYH: R52A,W58A,Y155A, H484A; RYHY: R52A,K149A,Y155A,Y490A; WKHY: W58A,K149A, Y155A,H484A; RYHY: R52A,Y155A,H484A,Y490A; WKHY: W58A,K149A, H484A,Y490A).

### 3.3 Each qRRM Confers One Component of TCE Recognition

The requirement for at least two qRRMs and both RNA-binding modes to regulate nos translation in vivo suggested a testable hypothesis for how individual qRRMs contribute to RNA binding. In particular, we hypothesized that while one qRRM interacts with the TCEI G-tract, one of the other two qRRMs interacts with the UA-rich TCEIII (Figure 3.3). In this model, one qRRM would only require an intact G-tract binding surface, whereas the other qRRM would only require an intact UA-rich TCEIII binding surface. Furthermore, the third qRRM would not be required at all, since only two qRRMs are required regulate to nos translation.
Figure 3.3. Model for Glo-TCE Interaction. Binding to the TCE requires two Glo qRRMs (Section 3.2). This model postulates that one qRRM binds to the UA-rich TCEIII through the β-sheet residues, while the other binds to the TCEI G-tract through the loop residues. Although binding by qRRMs 1 and 2 is depicted, data presented above suggest that any combination of two qRRMs is allowed.

To test this model, we engineered GFP-Glo proteins with various combinations of UA-rich- and G-tract binding mutations within all three qRRMs and examined their ability to interact with nos by coimmunoprecipitation (Figure 3.4). Strikingly, nos was associated with all mutants tested and, in all cases, Nos protein levels were comparable to that of wild-type GFP-Glo (Figures 3.4B, 3.4C, 3.4D, and 3.4E). Immunoblot analysis showed that all GFP-Glo transgenes were expressed at comparable levels (Figures 3.2C and 3.2D), indicating that their ability to repress nos translation did not result from changes in protein expression. Together, these results indicate that translational repression of nos requires two qRRMs, each of which contributes one of Glo’s dual RNA binding modes.
Figure 3.4. Each qRRM Confers One Component of TCE Recognition. (A) Schematic representation of the wild-type GFP-Glo protein, colored as in Figure 3.1. (B) GFP-Glo proteins with alanine substitutions that were assayed for function in vivo. G-tract binding mutations are shown in red. TCEIII binding mutations are shown in dark blue. All gfp-glo transgenes were inserted into the same chromosomal location. (C) RT-PCR to detect nos or tub mRNA co-immunoprecipitated from ovaries expressing GFP-Glo mutant proteins (corresponding to panel B; See also text). Total RNA from the extracts used for immunoprecipitation was used as a positive control for the RT-PCR reaction (T). Reactions were performed in the presence (+RT) or absence (–RT) of reverse transcriptase. (D) Immunoblot of anti-GFP immunoprecipitates shown in C, demonstrating efficient isolation of GFP-Glo mutant proteins. (E) Immunoblot analysis of Nos protein in extracts of late ovaries from glo mutant females expressing GFP-Glo mutant proteins (Corresponding to panel B; See also text). Kinesin heavy chain (Khc) was used as a loading control. RWKY490: R52A,W58A,K149A,Y490A; RWYH:
Glo is required for viability to adulthood and although the relevant target RNAs are not known, we found that, as for nos repression, both the Glo G-tract and TCEIII binding modes are essential (Tamayo et al., 2017). We therefore wondered whether multiple qRRMs are also required to regulate the RNAs needed for viability. To further probe the RNA-binding requirements of Glo in vivo we assayed the ability of our new Glo mutant proteins to rescue the lethality associated with glo (Figure 3.5). GFP-Glo proteins with mutations affecting a single qRRM (qRRM1: R52A,W58A; qRRM2: K149A,Y155A; qRRM3: H484A,Y490A) were mostly able to rescue the lethality associated with glo, although not to the extent of the wild-type GFP-Glo protein (Figure 3.5). By contrast, GFP-Glo proteins with mutations in multiple qRRMs (qRRM1/2: R52A,W58A,K149A,Y155A – abbreviated RWKY155; qRRM1/3: R52A,W58A, H484A,Y490A – abbreviated RWHY; qRRM2/3: K149A,Y155A, H484A,Y490A – abbreviated KYHY) were unable to confer survival to adulthood, suggesting that, similar to the regulation of nos, multiple qRRMs are required to rescue lethality (Figure 3.5C). Surprisingly, GFP-Glo proteins containing only one available G-tract binding surface and one available UA-rich TCEIII binding surface were unable to rescue lethality (Figure 3.5C, right panel), despite their ability to rescue nos repression (Figure 3.4). This suggests that the RNA binding requirements of Glo for viability are distinct from those of nos repression, and may instead require all three qRRMs. Alternatively, the relevant target RNAs for viability may only require two qRRMs, with one of them contributing to both RNA-recognition modes. Since GFP-Glo proteins with deletions of individual qRRMs are unable to rescue the lethality associated with Glo (Y. Peng and E.R Gavis, unpublished data), we favor the former hypothesis. This is supported by the decrease in viability exhibited by mutants containing both G-tract and UA-rich binding mutations in
individual qRRMs (Figure 3.5C, left panel) Taken together, our findings highlight the importance of multiple RBDs for achieving the diverse functional requirements of Glo in vivo.

**Figure 3.5. Survival to Adulthood Requires Multiple qRRMs and Both RNA-Recognition Modes of Glo.** (A) Crossing scheme used to determine rescue of lethality associated with glo. DF glo/TM6B virgin females were mated to FRT82B glo−/TM6TbSb males carrying a single copy of wild-type or mutant GFP-Glo. Complete rescue by the gfp-glo transgene is expected to result in 33% of F1 progeny with the gfp-glo/+; FRT82B glo−/Df glo genotype. To calculate percent viability, this expected frequency was compared to the experimentally determined frequency (See 7.12) (B) Schematic representation of the wild-type GFP-Glo protein, which rescues viability, colored as in Figure 3.1. (C) GFP-Glo proteins with alanine substitutions that were assayed for function in vivo. G-tract binding mutations are shown in red. TCEIII binding mutations are shown in dark blue. All gfp-glo transgenes were inserted into the same chromosomal location. Percent viability, compared to GFP-GloWT, is shown for each mutant protein.
3.5 Discussion

Although single RBDs can use very diverse and sophisticated modes of RNA recognition, the use of multiple RBDs in a single protein has long been thought to help satisfy the diverse functional requirements of RBPs in the cell. By combining these RBDs in various arrangements, RBPs can recognize different and often longer RNA substrates in vitro (Lunde et al., 2007). Therefore, in proteins containing multiple RBDs, the domains can collaborate to achieve specific functions that cannot be achieved by the individual domains. Recognition of the nos TCE by Glo requires at least two qRRMs, and both Glo binding modes are required for regulation of nos in vivo. In contrast, we have previously shown that recognition of targets involved in nurse cell chromosome dispersion and dorsal-ventral patterning requires only the G-tract binding mode (Tamayo et al., 2017), and analysis of these new set of mutants further indicates that these RNAs only require binding of a single qRRM to G-tract motifs (Figure 3.6). Thus, it is the combination of the two binding modes within the three qRRMs that confers multifunctionality at the level of target RNA selection and at the level of regulatory function.
Figure 3.6. Glo Functions During Mid-oogenesis Requires Binding of a Single qRRM to G-tract motifs. (A) Schematic representation of the GFP-Glo protein, colored as in Figure 3.5, with a summary of the alanine substitutions that were assayed for function in vivo. (B,C) Confocal images of glo mutant egg chambers expressing GFP-Glo proteins with various RNA binding point mutations shown above. Nuclei were stained with DAPI (blue) and the actin cytoskeleton was stained with phalloidin (green). Scale bar = 20 µm. (B) GFP-Glo proteins containing mutations in both RNA-binding modes of individual qRRMs. (C) GFP-Glo proteins containing mutations in various combinations (see text). (D,E) Quantification of the percentage of eggs from flies expressing the proteins shown B (D) or C (E) with dorsal appendage defects, including shortened, laterally expanded, fused, and/or missing appendages. (RWKY155: R52A,W58A,K149A,Y155A; RWHY: R52A,W58A,H484A,Y490A; KYHY: K149A,Y155A, H484A,Y490A); RWKY490: R52A,W58A,K149A,Y490A; RWHY: R52A,W58A,Y155A, H484A; RKYY: R52A,K149A,Y155A,Y490A; WKHY: W58A,K149A, Y155A,H484A; RYHY: R52A,Y155A,H484A,Y490A; WKHY: W58A,K149A,H484A,Y490A; RWKY490: R52A,W58A,K149A,Y490A; RWHY: R52A,W58A, Y155A, H484A; RKYY: R52A,K149A,Y155A,Y490A; WKHY: W58A,K149A, Y155A,H484A; RYHY: R52A,Y155A,H484A,Y490A; WKHY: W58A,K149A,H484A,Y490A)

Glo is required for viability to adulthood and although the relevant target RNAs are not known, we found that, as for nos repression, both the Glo G-tract and TCEIII binding modes are essential, suggesting that Glo similarly recognizes other RNAs with bipartite motifs in addition to nos. Significantly, the RNA binding requirements of Glo for viability are distinct from those of nos repression, and may instead require all three qRRMs. Thus, although the relevant target RNAs may be similar to nos with respect to sequence and structure, their regulation may be more sensitive to perturbations in the RNA binding domains of Glo. Alternatively, the
targets required for viability may contain multiple G-tract and/or UA-rich motifs, requiring the recognition by all three qRRMs. Finally, it is possible that mutating both RNA binding surfaces within a qRRM also disrupts a protein-protein interaction domain that is required for Glo mediated survival to adulthood. Although we cannot currently distinguish between these possibilities, the work presented here highlights the importance of multiple RBDs for regulating Glo targets. As other RNA targets are identified, computational analyses should search for all possible combinations throughout the genome (See Chapter 5).

Recent studies suggest that the combinatorial use of RBDs in proteins may be more common than was previously thought (Castello et al., 2012; Mitchell et al., 2013), and structures of tandem RBDs bound to RNA have revealed the diversity in the mechanisms employed by RBPs to interact with RNA. For example, the structure of Prp24 in complex with U6 RNA show that three of the RRMs encircle the large RNA loop to form an interlocked protein-RNA ring, with one RRM making noncanonical contacts with the double-stranded region of the RNA (Montemayor et al., 2014). Thus, in addition to identifying the Glo-RNA interactome, it will be of interest to obtain co-crystals of Glo with the TCE. To date, there has been no systematic analysis of how RNA binding confers regulation in the context of an intact cell. Our discovery of Glo’s dual RNA-binding modes and distinct requirements for each mode in vivo (Summarized in Table 3.1) highlight the importance of Glo as a model for investigating the relationship between RNA recognition, regulatory activity, and biological role.
Table 3.1. Summary of Glo-RNA Binding Point Mutants

<table>
<thead>
<tr>
<th>Glo Point Mutant Proteins</th>
<th>Viability</th>
<th>nurse cell nuclear morph.</th>
<th>gurken regulation</th>
<th>nanos translation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>R52A</td>
<td>98%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>K149A</td>
<td>97%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>H484A</td>
<td>99%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>R52A K149A</td>
<td>77%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>R52A</td>
<td>78%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>K149A</td>
<td>79%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>R52A K149A</td>
<td>0%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>W58A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>Y155A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>Y490A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>W58A Y155A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>Y490A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>W58A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>Y155A</td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>W58A Y155A</td>
<td>0%</td>
<td>pollytene</td>
<td>mislocalized</td>
<td>unregulated</td>
</tr>
<tr>
<td>R52A</td>
<td>78%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>Y155A K149A</td>
<td>75%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>Y480A H484A</td>
<td>81%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>W58A R52A K149A</td>
<td>0%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>W58A R52A Y490A H484A</td>
<td>0%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>Y155A K149A Y490A H484A</td>
<td>0%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
</tbody>
</table>
Table 3.1 (continued). Summary of Glo-RNA Binding Point Mutants

<table>
<thead>
<tr>
<th>Glo Point Mutant Proteins</th>
<th>Viability</th>
<th>nurse cell nuclear morph.</th>
<th>gurken regulation</th>
<th>nanos translation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>W58A R52A K149A Y490A</td>
<td>15%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>W58A R52A Y155A H484A</td>
<td>14%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>R52A K149A Y490A</td>
<td>19%</td>
<td>dispersed</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>W58A K149A H484A</td>
<td>17%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>R52A Y155A H484A</td>
<td>18%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
<tr>
<td>W58A K149A H484A</td>
<td>15%</td>
<td>dispersed</td>
<td>WT</td>
<td>unregulated</td>
</tr>
</tbody>
</table>
CHAPTER 4:

GLO USES BOTH RNA-BINDING MODES TO REGULATE OSKAR \textit{IN VIVO}

THIS CHAPTER WAS ADAPTED FROM A MANUSCRIPT CURRENTLY IN PREPARATION AND REPRESENTS A COLLABORATION WITH TAKAMASA TERAMOTO AND TRACI HALL (NIEHS)
4.1 Introduction

In principle, every step in the pathway from RNA to protein can be controlled. Indeed, most genes require multiple layers of regulation, and the interplay of translational control with other post-transcriptional regulatory mechanisms such as mRNA transport allow cells to precisely control the expression of a particular protein. Post-transcriptional gene regulation by RNA-binding proteins (RBPs) has proved crucial for developmental events including the establishment and patterning of body axes, asymmetric cell fate decisions, and morphogenesis. In *Drosophila*, for example, specification of both the germ line and abdomen depends on localization of *oskar* (*osk*) mRNA to the posterior of the oocyte (Ephrussi et al., 1991; Kim-Ha et al., 1991). This process requires the microtubule plus-ended motor Kinesin (Zimyanin et al., 2008) as well as several RBPs, including the heterogeneous nuclear ribonucleoproteins (hnRNP s) Squid, PTB/hnRNP I, and Hrp48/Hrb27C (Besse et al., 2009; Huynh et al., 2004; Norvell et al., 2005; Yano et al., 2004). Nevertheless, only a few RBPs have been shown to directly interact with *osk in vivo* and little is known about how RBPs interact with the sequences that mediate *osk* localization to the posterior.

mRNA localization is often by mediated by the interaction of RBPs with sequences in the 3’ untranslated region (UTR) of transcripts (Abaza and Gebauer, 2008). However, previous studies indicated that this region is not sufficient for posterior localization of *osk* (Hachet and Ephrussi, 2004), suggesting that other features of the mRNA are crucial for localization. Consistent with this, splicing of the first intron of *osk* was shown to be required for posterior localization in addition to the 3’ UTR (Hachet and Ephrussi, 2004). Splicing of this intron creates a short RNA stem-loop structure, termed the spliced *oskar* localization element (SOLE), which acts in association with the exon junction complex (EJC) and is thought to constitute a recognition element for an as yet unknown protein (Ghosh et al., 2012). Localization of *osk* is therefore complex and may be dictated by factors that are recruited in the nucleus as well as by nuclear processing events.
The ability to undergo nucleocytoplasmic shuttling is a hallmark of the hnRNPs and makes them especially suited to regulate all aspects of post-transcriptional gene regulation, from RNA processing to mRNA localization and translation, while also providing a link between these events. Indeed, an increasing number of hnRNPs with general roles in pre-mRNA metabolism have now been shown to play additional, specialized roles in the regulation of specific transcripts (Piccolo et al., 2014). hnRNPs K and E1, which repress translation of 15-lipoxygenase (LOX) mRNA during erythrocyte differentiation, and hnRNP I, which participates in localization of Vg1 and VegT mRNAs in Xenopus oocytes, recognize specific sequence motifs in the 3′ UTRs of their target mRNAs that are required for their regulation (Cote et al., 1999; Kress et al., 2004; Ostareck et al., 1997). Similarly, in *Drosophila* oocytes, the hnRNP F/H homolog Glorund represses the translation of *nanos* (*nos*) mRNA and this regulation has been shown to be mediated by two distinct RNA binding modes in Glo that interact with both a G-tract and structured UA-rich motif in the *nos* 3′ UTR (Kalifa et al., 2006; Tamayo et al., 2017).

Like its mammalian counterparts, which are involved in splicing, polyadenylation, mRNA stability, and translation, Glo has multiple functions (Dreyfuss et al., 2002a; Kalifa et al., 2009; Kalifa et al., 2006; Tamayo et al., 2017). Glo is required for viability to adulthood (Kalifa et al. 2006), and in addition to *nos* regulation, Glo has recently been implicated in translational repression of nuclear-encoded mitochondrial respiratory chain complex transcripts (Gehrke et al., 2015). Glo also plays roles in ovarian nurse cell chromosome organization and dorsal-ventral patterning, most likely as a splicing regulator (Kalifa et al., 2009; Tamayo et al., 2017). Unlike regulation of *nos* however, these functions appear to be mediated solely by the G-tract binding mode of Glo, suggesting that Glo uses distinct RNA-binding modes to differentially regulate its targets (Tamayo et al., 2017).

We have previously identified a role for Glo in the localization of *osk* during oogenesis (Kalifa et al., 2009). Consistent with this, Glo is part of a complex that contains Hrp48 (also called Hrb27C), which has been shown to interact with *osk* and colocalize with *osk* RNPs.
Notably, immunolocalization studies indicate that Glo is both nuclear and cytoplasmic, suggesting that Glo could provide the functional link between osk splicing in the nucleus and posterior localization in the oocyte cytoplasm (Kalifa et al., 2009; Kalifa et al., 2006). Indeed, although Glo does not affect the splicing of osk mRNA (Kalifa et al., 2009), mammalian hnRNP F has been shown to interact with components of the EJC (Singh et al., 2012). Here, we demonstrate that Glo regulates osk localization specifically during late stages of oogenesis. We show that Glo interacts specifically with the double-stranded UA-rich motif in the SOLE as well as a proximate G-tract and that, similar to nos, both RNA-binding modes of Glo are required to regulate osk. Our findings suggest a requirement for the SOLE during late stages of oogenesis and illustrate the importance of Glo’s dual RNA binding modes for diversifying target recognition in vivo.

4.2 Glo is Required for osk Localization During Late-Stages of Oogenesis

As a maternal mRNA, osk is transcribed in the ovarian nurse cells and is transported from the nurse cells into the oocyte early in oogenesis (Ephrussi et al., 1991; Kim-Ha et al., 1991). Subsequently, osk is localized to the posterior pole of the oocyte in two mechanistically distinct phases. During mid-oogenesis, osk is actively transported along microtubules by kinesin and is subsequently translated at the oocyte posterior (Brendza et al., 2000; Zimyanin et al., 2008). During later oogenesis however, osk is localized by a diffusion and entrapment mechanism, which ensures robust abdomen and germ cell formation during embryogenesis (Glotzer et al., 1997; Sinsimer, 2011). To better assess the requirement for Glo in osk localization, we analyzed the distribution of osk mRNA and accumulation of Osk protein during both mid- and late-stages of oogenesis. Although posterior localization of osk appears wild-type in glo mutants during mid-stages of oogenesis (Figure 4.1A), a small and variable fraction of late-stage oocytes exhibit significantly reduced osk mRNA at the posterior (Figure 4.1A). Consistent with this, Osk protein levels are substantially reduced in glo mutant ovary extracts that are enriched for late-stage oocytes, referred to as late ovaries (Andrews et al., 2011)
Taken together, these data indicate that Glo is specifically required during late-stages of oogenesis to regulate osk.

4.3 Glo Associates with osk mRNA in vivo

To determine whether Glo might interact directly with osk mRNA to mediate osk localization, we performed co-immunoprecipitation experiments from late ovary extracts to test the interaction of Glo with osk mRNA in vivo. To facilitate Glo immunoprecipitation, we took advantage of a transgenic line that expresses a functional GFP-Glo fusion protein (Tamayo et al., 2017). Following immunoprecipitation with an anti-GFP antibody, RNA was isolated and

**Figure 4.1 Glo Regulates osk During late-stages of oogenesis.** (A) Fluorescence in situ hybridization to osk (green) in wild-type (WT) or glo mutant egg chambers during mid- (stage 9) and late- (stage 12) stages of oogenesis. Nuclei are stained with DAPI (blue). Oocytes are oriented with anterior to the left. Scale bar, 10 µM. (B) Immunoblot analysis of Osk protein in extracts of late-stage enriched ovaries from WT and glo mutant females. Kinesin heavy chain (khc) was used as a loading control.
analyzed by RT-PCR using primers for either osk mRNA or a control tubulin RNA (tub). osk mRNA is detected in immunoprecipitates from GFP-Glo ovary extracts but not from ovaries expressing an unrelated RNA-binding protein (MCP-GFP) (Figure 4.2A). Moreover, the control tub RNA is not detected in either of the immunoprecipitates, indicating specificity of the interaction between Glo and osk. We therefore conclude that Glo associates with osk mRNA in vivo.

Figure 4.2. Glo Associates with osk mRNA in vivo. (A) RT-PCR to detect osk mRNA coimmunoprecipitated from late ovaries expressing either MCP-GFP or GFP-Glo using anti-GFP antibody. tub serves as a negative control. Total RNA from the extracts used for immunoprecipitation was used as a positive control for the RT-PCR reaction. Reactions were performed in the presence (+RT) or absence (-RT) of reverse transcriptase. Immunoblot analysis of anti-GFP immunoprecipitates demonstrates efficient isolation of MCP-GFP and GFP-Glo (bottom panel).
4.4 Glo Binds the osk SOLE in vitro.

Glo is the sole Drosophila representative of the hnRNP F/H family of proteins, which have been shown to bind RNA through three quasi-RNA Recognition motifs (qRRMs) (Dominguez and Allain, 2006; Dominguez et al., 2010; Kalifa et al., 2006; Tamayo et al., 2017). We have recently shown that the Glo qRRM contains two discrete RNA-binding interfaces that allow Glo to specifically recognize both G-tracts and structured UA-rich motifs in vitro (Tamayo et al., 2017). Notably, the osk SOLE contains a double-stranded UA-rich motif as well as a proximate G-tract (Figure 4.3A) and previous mutational analysis indicated that the double-stranded UA-rich motif is important for osk localization (Ghosh et al., 2012; Simon et al., 2015). We therefore hypothesized that Glo mediates osk localization through its interaction with the SOLE. To test this, we performed an electrophoretic mobility shift assay (EMSA). Although we were unable to evaluate binding to the SOLE by the full-length Glo protein due to poor expression, we found that a truncated protein containing the two N-terminal qRRMs (qRRM1,2) bound stably to the SOLE, with two plateaus apparent in the binding curve (Figures 4.3B and 4.3C). These data are best fit by a curve assuming two binding sites with higher ($K_D = 2.5 \mu M$) and lower ($K_D = 39.1 \mu M$) affinity components, suggesting that Glo qRRM1,2 binds both the G-tract and double-stranded UA-rich motifs in the SOLE. This behavior is similar to what we have observed previously for Glo binding to the nos translational control element (TCE) (Tamayo et al., 2017).
Figure 4.3. Glo Recognizes Two Elements in the osk SOLE that Contribute Higher- and Lower-Affinity Components. (A) Predicted secondary structure of the SOLE and SOLE_{UA} RNAs. RNA secondary structure was calculated using Mfold (Zuker, 2003). The G-tract and UA-rich motifs are boxed, and the SOLE_{GAG} mutation is indicated in red. (B) Representative EMSA gels of purified recombinant Glo qRRM1,2 binding to the SOLE (top), SOLE_{GAG} (middle) and SOLE_{UA} (bottom RNAs). The highest Glo qRRM1,2 protein concentration (right-most lanes) was 100 μM. No protein was added to the reactions in the lanes marked '-' . The positions of unbound RNA (asterisk) and Glo qRRM1,2:RNA complex (arrowhead) RNA are indicated. (C) Superposition of plots for binding of Glo qRRM1,2 to SOLE (black) or SOLE_{GAG} (red) RNA. Higher affinity binding to the SOLE is selectively disrupted by mutation of the G-tract sequence. Three technical replicates of EMSAs were performed, and apparent K_D values shown are mean ± SEM. (D) Superposition of plots for binding of Glo qRRM1,2 to SOLE (black) or SOLE_{UA} (dotted black) RNA. The SOLE_{UA} RNA retains the higher affinity binding component of the Glo-SOLE interaction, but the lower affinity component is diminished.
To determine whether recognition of the SOLE requires both G-tract and UA-rich motifs similar to recognition of the TCE, we first mutated the G-tract sequence in the SOLE to GAG (SOLE\textsubscript{GAG}; Figure 4.3A). Glo qRRM1,2 bound to the SOLE\textsubscript{GAG} with a \(K_D\) of 37.7 \(\mu\)M, similar to the lower affinity binding to SOLE RNA (\(K_D = 39.1 \mu\)M), and curve fitting now assumed a single binding site (Figures 4.3B and 4.3C). The loss of higher affinity binding indicates that the G-tract sequence is responsible for this component. Conversely, when the UA-rich stem was deleted in the presence of an intact G-tract (SOLE\textsubscript{3UA}) Glo qRRM1,2 bound with a \(K_D\) of 1.1 \(\mu\)M, similar to the \(K_D\) for the higher affinity binding site in the SOLE RNA (Figures 4.3B and 4.3D), whereas the lower affinity binding component was considerably reduced (\(K_D > 100 \mu\)M). We therefore conclude that Glo qRRM1,2 binds the SOLE and that this binding includes two components: higher affinity binding to the SOLE G-tract and lower affinity binding to the UA-rich stem.

4.4 Localization of osk Requires Both RNA-Recognition Modes of Glo

The ability of Glo to recognize both the G-tract and the double-stranded UA-rich motif in the SOLE suggests that Glo employs both modes of RNA recognition to regulate osk \textit{in vivo}. We therefore sought to determine whether mutations that have been shown to preferentially disrupt the ability of Glo to bind either G-tract or UA-rich stem sequences \textit{in vitro} affect osk localization \textit{in vivo}. To assess the RNA binding requirements of Glo for osk localization, we expressed wild-type GFP-Glo or mutant GFP-Glo proteins in glo mutants and analyzed the distribution of osk mRNA and accumulation of Osk protein during late-stages of oogenesis (Figure 4.4). In contrast to wild-type GFP-Glo, GFP-Glo proteins with mutations in either G-tract binding residues (GFP-Glo\textsuperscript{W58A,Y155A,Y490A}, abbreviated Glo\textsuperscript{WYY}) or UA-rich binding residues (GFP-Glo\textsuperscript{R52A,K149A,H484A}, abbreviated Glo\textsuperscript{RKH}) were unable to rescue the posterior localization defect of glo mutants (Figure 4.4C, 4.4D, and 4.4E), indicating that both RNA binding modes are required for posterior localization of osk. Consistent with this, Osk protein levels were substantially reduced in late ovary extracts from flies expressing either GFP-
Importantly, immunoblot analysis showed that both GFP-Glo\textsuperscript{WYY} and GFP-Glo\textsuperscript{RKH} were expressed at levels comparable to wild-type GFP-Glo (Tamayo et al., 2017), indicating that the reduced localization of osk at the posterior did not result from a reduction in the expression of the proteins. Together, these results indicate that, consistent with the \textit{in vitro} binding studies, both modes of RNA-recognition are required for localization of osk during late-stages of oogenesis.

\textbf{Figure 4.4 Glo Uses Both Modes of RNA Recognition for Localization of osk to the Posterior.} (A) Schematic representation of the wild-type GFP-Glo protein, with a summary map of the alanine substitutions that were assayed for function \textit{in vivo}. The N-terminal GFP tag (green) and qRRM1 (yellow), qRRM2 (cyan), and qRRM3 (orange) are indicated. (B-E) Fluorescence \textit{in situ} hybridization to osk (green) in a stage 12 glo mutant oocyte (B) and glo mutants expressing GFP-Glo (C), GFP-Glo\textsuperscript{WYY} (D), or GFP-Glo\textsuperscript{RKH} (E). Nuclei are stained with DAPI (blue). Oocytes are oriented with anterior to the left. Scale bar, 10 µM. (F) Immunoblot analysis of Osk protein in extracts of late ovaries from glo mutant females expressing GFP-
Glo proteins with mutations in either G-tract binding residues or UA-rich stem binding residues. Kinesin heavy chain (khc) was used as a loading control.

4.5 Discussion

As soon as transcripts emerge from RNA polymerase II, they are covered in RBPs, which mediate various RNA processing reactions including splicing, 5′-end capping, 3′-end cleavage and polyadenylation (Dreyfuss et al., 2002a; Maniatis and Reed, 2002; Moore, 2005). Of all the RBPs that assemble on mRNA, the most abundant are the heterogeneous nuclear ribonucleoproteins. Indeed, although all mRNAs are associated with hnRNPs and many hnRNPs recognize numerous mRNAs, recent studies have identified roles for several hnRNPs in the localization and/or translation of specific mRNAs. We previously identified the Drosophila hnRNP F/H homolog Glo as an ovarian repressor of nos translation (Kalifa et al., 2006). Here, we have investigated an additional role for Glo in the localization of osk through an its interaction with the SOLE. Several lines of evidence establish Glo as an osk localization factor. First, binding of Glo to the SOLE in vitro correlates with the ability of this element to localize osk in vivo. Second, RNA co-immunoprecipitation of GFP-Glo demonstrates that Glo associates with osk mRNA in vivo. Finally, loss of glo or mutation of the residues in Glo required for RNA recognition results in a lack of posteriorly localized osk mRNA during late-stages of oogenesis.

The fact that both the SOLE and EJC are essential for osk mRNA localization indicates that splicing and cytoplasmic localization are mechanistically coupled. Interestingly, Glo is both nuclear and cytoplasmic, and in addition to its cytoplasmic function as a translational repressor, our previous work and precedence from human homologs suggest a nuclear function for Glo in splicing (Chou et al., 1999; Gamberi et al., 1997; Kalifa et al., 2009; Tamayo et al., 2017). Although we have not detected a defect in osk splicing in glo mutant ovaries (Kalifa et al., 2009), our discovery that Glo binds the SOLE to mediate
posterior localization of osk, suggests that Glo provides the functional link between osk splicing in the nucleus and posterior localization in the oocyte cytoplasm.

Previous studies have demonstrated the importance of splicing at the first exon-exon junction for the localization of osk mRNA to the posterior of the oocyte. However, these experiments only assessed the requirement for the SOLE during mid-oogenesis (Hachet and Ephrussi, 2004; Ghosh et al., 2012). Our analysis of Glo mutants suggests that the SOLE may also be required during late-stages of oogenesis and that interaction with Glo is required specifically during these stages. It will therefore be of interest to determine whether the SOLE is indeed required during late-stages of oogenesis and, similarly, whether both G-tract and UA-rich stem motifs are required for osk localization during mid-oogenesis.

Osk protein synthesized at the posterior initiates the assembly of the germ plasm, a specialized cytoplasm that contains nos as well as other determinants necessary for germline and abdominal development (Lehmann and Nusslein-Volhard, 1986; Wang et al., 1994b). Interestingly, recognition of the SOLE by Glo appears is similar to the what we have previously described for recognition of the nos TCE (Tamayo et al., 2017). In contrast, recognition of targets involved in nurse cell chromosome dispersion and dorsal-ventral patterning requires only the ability to bind G-tract motifs. Thus, it is possible that Glo uses similar RNA binding modes to regulate coherent biological programs, such as axial patterning. In support of this possibility, recent advances in the use of genomic techniques have revealed that many RNA binding proteins function as ‘posttranscriptional operons’, in some cases using sequence elements present in the UTRs of transcripts to coordinate expression of gene products needed collectively for a biological process (Keene and Tenenbaum, 2002). As other RNA targets of Glo are identified, computational analyses should search for both G-tract and UA-rich motifs. Analysis of the sequences directly associated with Glo will provide a basis for understanding how RNA-binding proteins are able to function in a variety of processes and the rules by which they do so.
CHAPTER 5:

IN SILICO ANALYSIS OF POTENTIAL GLO RNA TARGETS

THIS CHAPTER REPRESENTS A COLLABORATION WITH ROBERT LEACH OF THE BIOINFORMATICS CORE AT THE LEWIS-SIGLER INSTITUTE FOR INTEGRATIVE GENOMICS AT PRINCETON UNIVERSITY
5.1 Introduction

Identifying the regulatory sequence motifs that mediate post-transcriptional gene regulation is a major challenge in RNA biology. Unfortunately, the experimental determination of RNA binding sites can be tedious and difficult. Alternatively, regulatory sequence motifs can be predicted by computational methods. Although less accurate than experimental observations, in silico analyses can be sufficiently accurate to prompt functional hypotheses and guide experiments, e.g. to identify potential RNA targets. Our discovery of Glo's dual RNA-binding modes and distinct requirements for each mode in vivo highlight the importance of Glo as a model for investigating the relationship between RNA recognition, regulatory activity, and biological role. To further unravel this regulatory layer, we have used a bioinformatic approach based on known targets of Glo to identify other potential RNA targets of Glo. Our transcriptomic analysis identified a list of 806 potential target RNAs that contain both a double-stranded UA-rich stem as well as a proximate G-tract. Notably, both nos and osk are identified by this method and many of the other candidates have been shown to undergo post-transcriptional gene regulation, demonstrating the strength of this approach. Analysis of how these targets are regulated as well as which RBDs of Glo are engaged will deepen our understanding of how RNA recognition by Glo influences RNA metabolism.

5.2 Identification of Potential RNA Targets of Glo

We have shown that Glo binds the nos TCE and osk SOLE by interacting with both a double-stranded UA-rich stem as well as a proximal G-tract (Tamayo et al., 2017; See also Chapter 4). Like its mammalian counterparts, however, Glo has multiple functions, suggesting additional mRNA targets. Glo is required for viability to adulthood (Kalifa et al. 2006) and although the relevant target RNAs are not known, we found that, as for the regulation of nos and osk, both the Glo G-tract and UA-rich binding modes are essential. This suggests that Glo similarly recognizes other RNAs with similar motifs in addition to nos and osk. We therefore sought to use computational methods to identify other potential Glo target RNAs throughout
the genome that contain similar motifs. Using the nos TCE and osk SOLE as templates, we identified parameters that would allow for the identification of both nos and osk mRNAs by computational methods (See Appendix). This analysis identified 806 potential target RNAs that contain both a double-stranded UA-rich stem as well as a proximate G-tract (Table 5.1). Interestingly, many of the motifs identified are predicted to form higher order structures, with the G-tract and double stranded UA-rich sequences present in distinct stem-loops like the TCE and SOLE (Figure 5.1). Furthermore, both nos and osk are identified by this method and many of the other candidates have been shown to undergo post-transcriptional gene regulation. Thus, we argue that this in silico analysis may be useful for identifying other potential Glo target RNAs.

Table 5.1: Potential RNA Targets of Glo Identified by In Silico Analysis

<table>
<thead>
<tr>
<th>5-HT1A</th>
<th>asl</th>
<th>bw</th>
<th>CG11030</th>
</tr>
</thead>
<tbody>
<tr>
<td>ace</td>
<td>asta-r1</td>
<td>c1galta</td>
<td>CG11138</td>
</tr>
<tr>
<td>acj6</td>
<td>asta</td>
<td>c3g</td>
<td>CG11155</td>
</tr>
<tr>
<td>acp1</td>
<td>asun</td>
<td>ca-alpha1t</td>
<td>CG11210</td>
</tr>
<tr>
<td>ag5r</td>
<td>atg13</td>
<td>cag</td>
<td>CG11263</td>
</tr>
<tr>
<td>ago1</td>
<td>atg17</td>
<td>cah1</td>
<td>CG11317</td>
</tr>
<tr>
<td>ahcy</td>
<td>b4</td>
<td>cah2</td>
<td>CG11360</td>
</tr>
<tr>
<td>akt1</td>
<td>babo</td>
<td>calx</td>
<td>CG11505</td>
</tr>
<tr>
<td>alk</td>
<td>bbx</td>
<td>cana-14f</td>
<td>CG11537</td>
</tr>
<tr>
<td>alpha-est8</td>
<td>ben</td>
<td>cap</td>
<td>CG11594</td>
</tr>
<tr>
<td>alpha-man-ii</td>
<td>betanactes2</td>
<td>cdep</td>
<td>CG11825</td>
</tr>
<tr>
<td>almn</td>
<td>bili</td>
<td>cdi</td>
<td>CG11843</td>
</tr>
<tr>
<td>amn</td>
<td>blimp-1</td>
<td>cdsa</td>
<td>CG12179</td>
</tr>
<tr>
<td>ampdeam</td>
<td>boil</td>
<td>CG10431</td>
<td>CG12278</td>
</tr>
<tr>
<td>ance-3</td>
<td>bol</td>
<td>CG10492</td>
<td>CG12402</td>
</tr>
<tr>
<td>anne</td>
<td>brat</td>
<td>CG10638</td>
<td>CG12502</td>
</tr>
<tr>
<td>ap-1-2beta</td>
<td>bru1</td>
<td>CG10700</td>
<td>CG12516</td>
</tr>
<tr>
<td>ap-2alpha</td>
<td>bruce</td>
<td>CG10702</td>
<td>CG12531</td>
</tr>
<tr>
<td>aqp</td>
<td>brwd3</td>
<td>CG10804</td>
<td>CG12605</td>
</tr>
<tr>
<td>ara</td>
<td>bsg</td>
<td>CG10859</td>
<td>CG12659</td>
</tr>
<tr>
<td>arf102f</td>
<td>btd</td>
<td>CG10898</td>
<td>CG12769</td>
</tr>
<tr>
<td>arl5</td>
<td>btv</td>
<td>CG10920</td>
<td>CG12912</td>
</tr>
<tr>
<td>ars2</td>
<td>bugz</td>
<td>CG1103</td>
<td>CG12917</td>
</tr>
<tr>
<td>CG12963</td>
<td>CG13003</td>
<td>CG13004</td>
<td>CG13126</td>
</tr>
<tr>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
<td>Gene</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>CG43693</td>
<td>CG5966</td>
<td>CG9727</td>
<td>dl</td>
</tr>
<tr>
<td>CG43707</td>
<td>CG6051</td>
<td>CG9782</td>
<td>dnc</td>
</tr>
<tr>
<td>CG43729</td>
<td>CG6123</td>
<td>CG9813</td>
<td>dop1r2</td>
</tr>
<tr>
<td>CG43813</td>
<td>CG6168</td>
<td>CG9896</td>
<td>dop2r</td>
</tr>
<tr>
<td>CG43901</td>
<td>CG6175</td>
<td>chrb</td>
<td>dpr4</td>
</tr>
<tr>
<td>CG44085</td>
<td>CG6191</td>
<td>cht10</td>
<td>dmr</td>
</tr>
<tr>
<td>CG44245</td>
<td>CG6231</td>
<td>cht2</td>
<td>dsp1</td>
</tr>
<tr>
<td>CG44341</td>
<td>CG6321</td>
<td>ci</td>
<td>dtn</td>
</tr>
<tr>
<td>CG44403</td>
<td>CG6398</td>
<td>cib</td>
<td>dyrk2</td>
</tr>
<tr>
<td>CG44422</td>
<td>CG6435</td>
<td>cka</td>
<td>dyrk3</td>
</tr>
<tr>
<td>CG44428</td>
<td>CG6443</td>
<td>ckalpha</td>
<td>eag</td>
</tr>
<tr>
<td>CG4467</td>
<td>CG6664</td>
<td>cln3</td>
<td>ear</td>
</tr>
<tr>
<td>CG44838</td>
<td>CG6700</td>
<td>cnn</td>
<td>eas</td>
</tr>
<tr>
<td>CG45002</td>
<td>CG6959</td>
<td>comm3</td>
<td>ecr</td>
</tr>
<tr>
<td>CG45049</td>
<td>CG7149</td>
<td>con</td>
<td>ect4</td>
</tr>
<tr>
<td>CG45086</td>
<td>CG7335</td>
<td>cp7fb</td>
<td>efn1c.2</td>
</tr>
<tr>
<td>CG45088</td>
<td>CG7367</td>
<td>cpo</td>
<td>egfr</td>
</tr>
<tr>
<td>CG45413</td>
<td>CG7372</td>
<td>cpr5c</td>
<td>eif4a</td>
</tr>
<tr>
<td>CG4572</td>
<td>CG7470</td>
<td>cpx</td>
<td>eif4g2</td>
</tr>
<tr>
<td>CG4619</td>
<td>CG7564</td>
<td>creba</td>
<td>eig71ei</td>
</tr>
<tr>
<td>CG46306</td>
<td>CG7646</td>
<td>crk</td>
<td>eip93f</td>
</tr>
<tr>
<td>CG46320</td>
<td>CG7708</td>
<td>crm</td>
<td>elav</td>
</tr>
<tr>
<td>CG46321</td>
<td>CG8034</td>
<td>croc</td>
<td>ena</td>
</tr>
<tr>
<td>CG4658</td>
<td>CG8097</td>
<td>crol</td>
<td>endoa</td>
</tr>
<tr>
<td>CG4686</td>
<td>CG8177</td>
<td>crtc</td>
<td>ens</td>
</tr>
<tr>
<td>CG4747</td>
<td>CG8230</td>
<td>csp</td>
<td>ent3</td>
</tr>
<tr>
<td>CG4766</td>
<td>CG8272</td>
<td>ctrip</td>
<td>epac</td>
</tr>
<tr>
<td>CG4793</td>
<td>CG8500</td>
<td>ctsb1</td>
<td>epg5</td>
</tr>
<tr>
<td>CG4853</td>
<td>CG8516</td>
<td>cul3</td>
<td>ephrin</td>
</tr>
<tr>
<td>CG4911</td>
<td>CG8547</td>
<td>cyck</td>
<td>eps-15</td>
</tr>
<tr>
<td>CG4968</td>
<td>CG8563</td>
<td>cyp4aa1</td>
<td>esn</td>
</tr>
<tr>
<td>CG5065</td>
<td>CG8641</td>
<td>cyp4s3</td>
<td>ethr</td>
</tr>
<tr>
<td>CG5157</td>
<td>CG8786</td>
<td>cyp6g1</td>
<td>ets96b</td>
</tr>
<tr>
<td>CG5440</td>
<td>CG8925</td>
<td>dally</td>
<td>evi5</td>
</tr>
<tr>
<td>CG5850</td>
<td>CG9005</td>
<td>dar1</td>
<td>evg</td>
</tr>
<tr>
<td>CG5863</td>
<td>CG9132</td>
<td>dco</td>
<td>fas2</td>
</tr>
<tr>
<td>CG5897</td>
<td>CG9220</td>
<td>deaf1</td>
<td>fd96ca</td>
</tr>
<tr>
<td>CG5921</td>
<td>CG9391</td>
<td>dh31</td>
<td>fend</td>
</tr>
<tr>
<td>CG5937</td>
<td>CG9425</td>
<td>dhit</td>
<td>fife</td>
</tr>
<tr>
<td>CG5953</td>
<td>CG9684</td>
<td>dia</td>
<td>fit1</td>
</tr>
</tbody>
</table>
fkbp14  heix  krz  mkk4
fkh    hers  ksr  mkp-4
fne    hipk  kug  mlp84b
foxp   his3.3b  lambda try  mmp2
fray   hk  lap  mnt
fred   hmt4-20  lawc  mo25
frl    hmu  lbl  msp
frq1   hmx  ldsdh1  mthl3
fru    hnf4  lic  mub
fs(1)h hppy  lili  mur89f
ftz-f1 hr3  lip4  myc
fus    hr4  lola  myo31df
futsch hs3st-a  lpr1  mys
fwd    hsc70-4  lrch  mys45a
fz2    hth  lrp4  na
gaa1   htra2  luna  nachralpha1
gaba-b-r1 hts  mad  nachralpha3
galphao hwt  maf1  nachralpha5
galphaq ih  mahe  nachralpha7
gbs-76a imp  mam  nacp60e
gcs1   inaf-b  mamo  nak
gdi    inaf-d  mbc  nckx30c
gfmeso inr  mbl  ndae1
ggamma30a ints6  mbs  nej
glurid  inv  mcu  nemy
gogo   ipod  mdr50  nep3
gp150  ir7d  mef2  netb
gpp    jbug  mei-p26  neto
gr28b  jeb  melt  neur
gramd1b jef  meltin  nfi
grh    jigr1  metr s-m  nkain
grmd   jil-1  mew  nknd
gs1    jmj4  mex1  nona
gs1    kcc  mf  nmo
gss2   kdn  mfs10  nop56
gstz1  kirre  mfs17  nop60b
gw     kl-2  mfs10  nord
glyc beta100 kl-5  mif3  nos
hcf    klu  mini  npc1a
hdc    kr  miro  nrg
nrv3
nrx-1
nse4
nsf2
nwk
oamb
obp18a
oc
octbeta1r
olf413
or22a
or45a
or65c
or85d
orb
orc2
org-1
osk
Pab p2
pak
pak3
pan
papst2
para
pde8
pdfr
pdk1
pdp1
pex2
pgant1
pgant35a
pgrp-la
pgrp-sc1b
phcl-2
piezo
pigs
pins
pka-c1
pka-r2
pkc98e
pkcdelta
ple
plexa
plexb
pmca
pmp70
png
polybromo
porin
pp2b-14d
prge
prate
pro
pten
ptp10d
ptp36e
pur-alpha
pvf3
pzi
qkr54b
rab10
rab3-gef
rab35
rangap
rap1
rasp
rbfox1
rbp6
rbp9
rdga
rdgb
rdl
reph
rg
rgn
rh7
rhogap18b
rhogap19d
rhogap93b
rhogapp190
rim
ril
rnasez
robo3
root
roq
rpa3
rph
rpi38
rpt4
rrp4
rtgef
rudhira
rut
s6kii
sano
santa-maria
sap-r
sap47
sbb
scap
scrt
sdc
sec71
sema1a
sep4
serca
sesb
sf2
sfl
sh
shakb
shaw
shd
shg
side
sif
simj
sk
skd
skpa
sky
slc45-1
slik
slo
slo2
slpr
smog
smyda-4
snoo
snp
snx3
snz
sowah
sp
sp1
sp2353
spase25
spin
spn77ba
spps
spri
sqa
sqd
sti
str-mlck
stw
sty
su(dx)
su(tpl)
su
sun
sur
| sut1 | thd1 | u2af50 | vnd  |
| svr  | tio  | ubc7   | vrp1 |
| sxl  | tk   | ubl3   | vsg  |
| syn  | tm9sf4 | ubx   | vvl  |
| syn2 | tn   | uch-l5r | waw  |
| synd | tnks | ulp1   | wcy  |
| syngr | tob | unc-13 | wdb  |
| synj | tom20 | unc-76 | wech |
| syt1 | tou  | unr    | wrapper |
| syt12 | tpnc4 | upd1   | wts  |
| sytbeta | trc | usf    | x11lbeta |
| tank | trc8 | usp16-45 | yrt  |
| task6 | trpa1 | usp30  | ytr  |
| tbc1d15-17 | trs23 | vap   | zfh2 |
| teh2 | trxt | vglut  | zip102b |
| teh4 | tsp97e | vha100-1 | ziz  |
| tep1 | tth  | vha14-2 | zld  |
| tgo  | tun  | vmat   | zyd  |
Figure 5.1. Predicted Secondary Structure of RNAs identified by *In silico* Analysis. (A-D) Predicted secondary structure of UA-rich/G-tract motifs from *sqd* (A) *brat* (B) *eIF4G2* (C) and *imp* (D) RNAs. RNA secondary structure prediction was calculated using Mfold (Zuker, 2003).

5.3 Gene Ontology Analysis of Potential RNA Targets of Glo

We next used the Protein Analysis Through Evolutionary Relationships (PANTHER) database (Mi et al., 2017) to perform a Gene Ontology term (GO term) enrichment analysis on the potential RNA targets of Glo (Figure 5.1). By classifying genes based on their molecular function and biological process, we found that a large number of potential targets of Glo are involved in binding, suggesting that they may also be involved in nucleic acid metabolism (Figure 5.1A). Indeed, both *nos* and *osk* are RNA-binding proteins that have been shown to
function as post-transcriptional gene regulators (Ephrussi et al., 1991; Sonoda and Wharton, 1999). This analysis also identified a significant number of genes involved in metabolic processes (Figure 5.1B), consistent with previous studies implicating Glo in the translational repression of mitochondrial respiratory chain complex transcripts (Gehrke et al., 2015). Overall, this GO term enrichment analysis identified genes involved in numerous biological processes and therefore suggests that the function of Glo may be much more widespread than we previously appreciated. Of course, it will be important to analyze these targets further to determine if Glo indeed regulates these transcripts directly.

Figure 5.2. Gene Ontology Analysis of Potential Glo Target RNAs. Genes are classified according to their molecular function (A) and biological process (B).

5.4 Discussion

In-depth analysis of RNA–protein interactions is crucial to understand how the gene expression output of cells is regulated. Yet hundreds of RBPs are encoded by genomes and only a few have been analyzed. Progress towards identifying in vivo RNA targets and binding sites has revealed that in silico approaches to determining binding specificity is largely valid, but consensus binding motifs alone cannot reveal endogenous targets. The development of models for protein–RNA interactions will therefore depend on experimental data from structure analysis and from immunoprecipitation-based methods, e.g. CLIP. Here, we have used in silico methods guided by structural, biochemical, and genetic analyses to identify...
potential Glo target RNAs with both g-tract and structured UA-rich motifs. Our analysis yielded a list of 806 potential target RNAs that should be further analyzed to determine if they are indeed regulated by Glo. Ultimately, analysis of how these targets are regulated as well as which RBDs of Glo are engaged will deepen our understanding of how RNA recognition by Glo influences RNA metabolism.
CHAPTER 6:

PERSPECTIVES
6.1 Concluding Remarks

Since the definitive identification of the hnRNP proteins and the discovery of the first consensus motifs in RBPs more than thirty years ago, the catalog of RBPs and the multitude of functions in which they participate has expanded enormously. It is clear that RBPs are critical components of the gene expression pathway in eukaryotes and their ability to regulate every aspect of the biogenesis and function of RNA is truly remarkable. It is also clear, however, that a great deal of information is lacking about the structure of RBPs, their mode of interaction with RNAs, and the relationship of RNA binding to biological function. Given the impressive progress that has already been made, the vast number of RBPs that remain to be characterized, and the rich arsenal of tools available to study them, the promise of what the study of RBPs still has in store for understanding biology and many diseases is extremely exciting.

An increasing number of hnRNP and other RBPs with general functions in RNA metabolism, including Drosophila Squid and Hrp48 (hnRNP A/B family), Rumpelstiltskin (Rump; hnRNP M), and Syncrip (hnRNP Q), have been shown to have specialized roles in the regulation of specific transcripts (Gavis et al., 2007; Kugler and Lasko, 2009). Similarly, Smaug, which functions as a major regulator of maternal transcript degradation at the maternal-zygotic transition, also acts as the embryonic repressor of nos translation (Tadros et al., 2007). Thus, while such multifunctionality is widespread among RBPs and is likely essential to their function, its basis is poorly understood.

Our discovery of Glo’s dual RNA-binding modes and distinct requirements for each mode in vivo highlight the importance of Glo as a model for investigating the relationship between RNA recognition, regulatory activity, and biological role. Indeed, the structure/function analysis described here offers one of the most in depth studies of hnRNP function to date and addresses critical questions about the relationship of RNA binding to biological function. Nevertheless, many important questions still need to be addressed (see
Section 6.2), and we envisage that future studies of Glo will facilitate a global understanding of the functional diversity of all RBPs.

### 6.2 Future Directions

To better understand the functional roles of Glo, further studies of its RNA targets are required, and more effort needs to be invested in refining the models for the Glo-TCE and Glo-SOLE interaction through genetic and biochemical approaches. In particular, it will be important to co-crystallize Glo with target RNAs containing a UA-rich motif, like the nos TCE. Because full-length Glo behaves poorly in solution however, likely due to the 242-aa GN-rich linker between qRRMs 2 and 3, the qRRM1,2 protein should provide a good starting point for co-crystallization. It may also be useful to generate constructs with truncated linker domains to facilitate crystallization of a protein containing all three qRRMs, although these should be tested for function in vitro and in vivo first. Alternatively, mammalian hnRNP F, which has a shorter linker, could provide a possible alternative for crystallization. Indeed, although hnRNP F has only been reported to bind G-tracts (Honore et al., 1995), the UA-rich binding residues of Glo are conserved in mammalian hnRNP F. It will therefore be important to determine whether hnRNP F also interacts with the TCE.

The molecular mechanisms by which mammalian hnRNP F/H family members regulate RNA processing are still unclear. In some cases, hnRNP F/H proteins regulate alternative splicing by binding G-tracts near 5'-splice-site-donor sequences and preventing the binding of spliceosomal components to the pre-mRNA (Han et al., 2005). In other cases, hnRNP F/H members regulate polyadenylation site choice by blocking recruitment of cleavage stimulation factor (Veraldi et al., 2001). Finally, GRSF-1 has been shown to stimulate translation of influenza virus-encoded mRNAs by binding to sequences in their 5'UTRs (Kash et al., 2002). Post-transcriptional regulation by hnRNP F/H members is therefore complex, and
these proteins might, depending on the RNA substrate, act at different stages of RNA metabolism through recognition of different sequence and/or structural motifs.

Our previous identification of Glo in a complex with the hnRNP A/B protein Hrp48 and the splicing factor Half-pint (Hfp) (Kalifa et al., 2009) implicates Glo in splicing in addition to localization and translational regulation. Furthermore, the pleiotropic phenotype of \textit{glo} mutants indicates that Glo has a diverse set of targets and, taken together, our data suggest that Glo exerts different regulatory activities – e.g., splicing vs. translational repression – on different targets. We hypothesize that the particular Glo recognition mode (i.e., G-tract, UA-rich motif, or both) influences how Glo regulates a target RNA. However, testing these ideas requires the identification of additional Glo target RNAs and Glo binding sites. We attempted to identify the Glo-RNA interactome experimentally by CLIP-seq (data not shown), which uses crosslinking of RNA-protein complexes \textit{in situ} followed by stringent immunopurification to yield target RNAs that can be amplified and sequenced (Barckmann et al., 2015; Darnell, 2010). However, these experiments were unsuccessful. Nevertheless, future methods should be considered to obtain the RNA targets of Glo \textit{in vivo}. As these transcripts are identified, it will be interesting to see how they compare to the \textit{in silico} analysis described here. In complementary studies, it will be interesting to identify Glo interacting proteins, since these should help to understand how Glo-RNA interactions ultimately confer biological regulation. Finally, the role of subcellular localization of Glo on RNA regulation should also be examined, especially considering that Glo is both nuclear and cytoplasmic. In light of the functional complexity of the hnRNPs, answers to these questions await a comprehensive integration of genetic, biochemical, and structural analyses focused on Glo protein-RNA and protein-protein complexes.
CHAPTER 7:

METHODS

7.1 Protein Expression and Purification

The cDNA sequences encoding the qRRM domains of Glo (qRRM1; 45-141, qRRM2; 142-234, qRRM3; 475-562, and qRRM1,2; 1-244) were subcloned into pET15b (Novagen), which encodes an N-terminal His<sub>6</sub>-tag. Individual qRRM domains were expressed in *E. coli* strain BL21-CodonPlus (DE3)-RIL (Agilent Technologies) at 20 °C overnight after induction with 0.5 mM IPTG. The cells were collected by centrifugation, and pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl) and stored at −80 °C until use. The cells were disrupted by sonication and the soluble fraction was applied to a Ni-NTA agarose column (Thermo Scientific). After thorough washing with lysis buffer containing 20 mM imidazole, protein was eluted with lysis buffer containing 400 mM imidazole. For crystallization, the His<sub>6</sub> tag was cleaved overnight with 5 U of thrombin (Novagen). Glo qRRM proteins were further purified using a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated with lysis buffer, and peak fractions were pooled. For Glo qRRM1 purification, the pooled fractions were dialyzed against a buffer containing 50 mM Tris-Cl, pH 8.0, 20 mM NaCl and purified further using a HiTrap Heparin column (GE Healthcare). Bound protein was eluted using a 0-1 M NaCl linear gradient in 50 mM Tris-Cl, pH 8.0. Peak fractions were pooled and dialyzed against a buffer containing 20 mM Tris-Cl, pH 8.0, 100 mM NaCl and then concentrated to 10 mg/ml. For Glo qRRM2 purification, the pooled fractions were dialyzed against a buffer containing 50 mM Tris-Cl, pH 8.0, 100 mM NaCl and purified further using a HiTrap Heparin column (GE Healthcare). Bound protein was eluted using a 0.1-1 M NaCl linear gradient in 50 mM Tris-Cl, pH 8.0. Peak fractions were pooled and dialyzed against a buffer containing 20 mM Tris-Cl, pH 8.0, 500 mM NaCl and then concentrated to 9 mg/ml. For Glo qRRM3 purification, pooled fractions were dialyzed against a buffer containing 50 mM Tris-Cl, pH 8.0, 100 mM NaCl. The dialyzed sample was applied to HiTrap Heparin column, HiTrap Q column (GE Healthcare), and HiTrap SP column (GE Healthcare) sequentially to remove contaminating proteins that bound to these columns. Unbound fractions containing Glo qRRM3 protein were
pooled and dialyzed against a buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, then concentrated to 10 mg/ml.

Individual Glo qRRM domains for \textit{in vitro} binding assays were purified by Ni-NTA agarose chromatography as described above, followed by purification on a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated with lysis buffer. The peak fractions were pooled and dialyzed against a buffer containing 20 mM HEPES-NaOH pH 7.9, 200 mM NaCl, and then concentrated.

WT Glo qRRM1,2 protein and qRRM1,2 mutants generated by site-directed mutagenesis (Agilent Technologies) were purified by the same procedure as individual Glo qRRM proteins for \textit{in vitro} binding assays. Following purification on a HiLoad 16/60 Superdex 75 column (GE Healthcare), peak fractions were pooled and dialyzed against a buffer containing 50 mM Tris-HCl pH 8.0, 50 mM NaCl, then, purified further using a HiTrap Heparin column (GE Healthcare). Bound proteins were eluted using a 0.05-1 M NaCl linear gradient in 50 mM Tris-HCl, pH 8.0. Peak fractions were pooled and dialyzed against a buffer containing 20 mM HEPES-NaOH, pH 7.9, 200 mM NaCl, and then concentrated. To confirm correct folding of the mutant proteins, we assessed protein folding by circular dichroism (CD) and found no differences in the spectra of wild-type Glo qRRM1,2, Glo qRRM1,2$^{W58A,Y155A}$, and Glo qRRM1,2$^{R52A,K149A}$ proteins (Figure S2, Figure S3). The CD spectra were measured on a JASCO J-810 CD spectrometer at room temperature. For each sample (200 μl in a 0.1 cm light-path cell), four scans were accumulated in the wavelength range of 190–260 nm with a 0.2 nm step size. Protein samples were 100 μg/ml in 20 mM Na phosphate, pH 8.0, 100 mM NaCl. The raw CD data were adjusted by subtracting a buffer blank. CD spectra of wild-type and mutant proteins displayed negative ellipticities at 208/222 nm and 215 nm, which indicate the presence of α helices and β strands, respectively.
7.2 Crystallization, Data Collection, Structure Determination and Refinement

Crystals of Glo qRRM1 were prepared by the sitting-drop vapor diffusion method at 4 °C. Sitting drops contained 250 nl of protein mixed with 250 nl of reservoir solution (0.2 M ammonium sulfate, 0.1 M Na acetate pH 4.6, 30% w/v PEGME 2000). Crystals of Glo qRRM2 or qRRM3 were prepared by the hanging-drop vapor diffusion method at 20 °C. Hanging drops contained 1 µl of protein mixed with 1 µl of reservoir solution (qRRM2: 0.2 M ammonium acetate, 20% v/v PEG 3350; qRRM3: 0.5 M ammonium sulfate, 0.1 M Na cacodylate pH 6.4, 1.0 M lithium sulfate). Prior to data collection, crystals were transferred to a cryoprotectant solution containing 15% glycerol and flash cooled to -180 °C. X-ray diffraction data for crystals of qRRM1 and qRRM3 were collected using a conventional X-ray source (Rigaku 007HF rotating anode equipped with VariMax HF mirrors and a Saturn 944 CCD detector) and for crystals of qRRM2 were collected at the SER-CAT Beamline 22-ID at the Advanced Photon Source, Argonne National Laboratories. Diffraction data were processed using the program package HKL2000 (HKL Research Inc. (Otwinowski and Minor, 1997). The resolution limits of data collection for qRRM1, qRRM2 and qRRM3 were restricted to 1.54, 1.55 and 1.99 Å, respectively, due to the detector settings during data collection, although the crystals might diffract to higher resolution. The structures were determined by molecular replacement using the program Molrep (Vagin and Teplyakov, 2000). The NMR structures of qRRM1, qRRM2 and qRRM3 of human hnRNP F were used as the search models for Glo qRRM1, qRRM2 and qRRM3, respectively. Model building was carried out with the program Coot (Emsley and Cowtan, 2004). The programs Refmac5 (Murshudov et al., 1997) and Phenix.refine (Adams et al., 2010) were used for refinement. The structures displayed good geometry when analyzed by MolProbity (Chen et al., 2010).

7.3 In vitro Transcription

TCE and TCE mutant RNAs were produced by in vitro transcription. The DNA templates were amplified from oligonucleotides by PCR using a primer containing the T7 promoter sequence.
TCE mutations were encoded in the corresponding template oligonucleotides. The PCR products were used directly as templates for *in vitro* transcription. *In vitro* transcription with T7 RNA polymerase was performed in a buffer containing 40 mM Tris-HCl, pH 8.0, 2.5 mM spermidine, 26 mM MgCl\(_2\), 0.01% v/v Triton X-100, 1 mM DTT, 4 mM NTPs, 16 mM GMP, 40 U RNase inhibitor, and 0.5 U pyrophosphatase. Reactions were incubated for 4 hr at 37 °C. Reactions were then supplemented with 50 U of DNase I and 10x DNase reaction buffer (400 mM Tris-HCl, pH 7.9, 100 mM NaCl, 60 mM MgCl\(_2\), 10 mM CaCl\(_2\)) and incubated for 30 min at 37 °C. RNA products were purified by phenol/chloroform extraction followed by ethanol precipitation. RNA was further purified by electrophoresis on a 10% polyacrylamide-urea gel (Invitrogen) in 1X TBE buffer. Target bands were detected by UV shadowing, excised from the gel, and incubated in RNA elution buffer (20 mM Tris-HCl, pH 7.5, 250 mM Na acetate, 1 mM EDTA, 0.25% w/v SDS) overnight at room temperature. Eluted RNA was diluted to 10 ml and purified further using a HiTrap Q column (GE Healthcare). Bound RNAs were eluted using a 0.05-1 M NaCl linear gradient in 50 mM Tris-HCl, pH 8.0. Peak fractions containing target RNAs were pooled, ethanol precipitated, and resuspended in DEPC-treated water.

### 7.4 Electrophoretic Mobility Shift Assay

TCE and TCE mutant RNAs were produced by *in vitro* transcription as described above, and a 5´-AGGGA RNA oligonucleotide was generated by RNA synthesis (GE Dharmacon). RNAs were radiolabeled at the 5´ end using [γ-\(^{32}\)P] ATP and T4 polynucleotide kinase, then purified using an Illustra MicroSpin G-25 column (GE Healthcare). TCE and TCE mutant RNAs were prepared by heating to 90 °C for 5 min, and then slowly cooling to room temperature. The 5´-AGGGA RNA oligonucleotide was incubated at 90 °C for 5 min in 50 mM Tris, pH 8.0 and then cooled on ice. This procedure yielded single-stranded RNA that bound to qRRMs as analyzed by size exclusion chromatography. RNA-binding reactions included 0.9 nM radiolabeled RNA and increasing concentrations (2-fold) of protein in a binding buffer containing 20 mM HEPES, pH 7.9, 50 mM NaCl, 50 mM KCl, 2 mM MgCl\(_2\), 0.02% v/v Tween-20, 1 μg/ml yeast tRNA, 0.1
mg/ml poly(rU), and 5% v/v glycerol. Binding reactions were incubated for 1 hr at room temperature and immediately separated by electrophoresis on 10% polyacrylamide gels (Invitrogen) in 1X TBE buffer at 100 V for 30 min at 4 °C. Gels were dried and exposed to storage phosphor screens for 6–20 hr, scanned with a Typhoon 8600 Imager, and the band intensities were quantified with ImageQuant 5.2. The data were analyzed and $K_d$ values were calculated via non-linear regression analysis for one- or two-site binding with GraphPad Prism 6. Three technical replicates of all binding assays were performed, and $K_d$ values are reported as mean ± standard error of the mean (SEM).

We assessed the stoichiometry of Glo qRRM1,2 protein binding to RNA by EMSA using 100 µM 5´-AGGGA or 500 µM TCE RNA and protein-RNA ratios of 0.1, 0.25, 0.5, 0.75, 1.0, 1.5 and 2.0. Glo qRRM1,2 binding to 5´-AGGGA RNA saturated at a protein-RNA ratio of ~0.5, corresponding to two 5´-AGGGA molecules per Glo qRRM1,2. Glo qRRM1,2 binding to TCE RNA was more complex and saturated at a protein-RNA ratio of ~1.5, which is not conclusive. Interpretation of the stoichiometry EMSA is complicated due to the two-site binding model, but the 1.5 protein-RNA ratio may correspond to a mixture of 1:1 and 2:1 protein-RNA complexes. Since full-length Glo contains three qRRM domains, the 2:1 complexes might include the binding of a third qRRM domain from a second protein.

7.5 Construction of gfp-glo and nos Transgenes and Transgenic Lines

A functional gfp-glo transgene was generated by insertion of the egfp coding sequences at the start codon of a 5.6 kb genomic glo rescue fragment (g-glo; Kalifa et al., 2009; Kalifa et al., 2006) and cloned into the pattB vector (http://www.flyc31.org/sequences_and_vectors.php). Point mutations were introduced into the glo sequences by PCR-based site-directed mutagenesis. All transgenes were inserted into the attp40 landing site by the phiC31 integrase method to eliminate position effects on gene expression. Phenotypic analysis was performed by crossing transgenes into the null glo$^{162x}$ background (Kalifa et al., 2006). Mutations in the nos 3´UTR were engineered by PCR into a 4.3 kb genomic nos rescue fragment (Gavis and
Lehmann, 1992) and cloned into the pattB vector (Bischof et al., 2013). All nos transgenes were inserted into the attP40 landing site.

### 7.6 Immunoblotting

For analysis of GFP-Glo levels during oogenesis, ovaries were dissected from well-fed females in PBS and frozen in liquid nitrogen. For analysis of Nos and Osk levels during late-stages of oogenesis, ovaries enriched for late-stage oocytes were generated as described previously (Andrews et al., 2011), dissected in PBS, and frozen in liquid nitrogen. Ovary extract preparation and immunoblotting followed the procedures of Forrest et al. (2004), using nitrocellulose membrane. The following primary antibodies were used: 1:1000 rabbit anti-Nos (gift of A. Nakamura); 1:2000 anti-Osk (gift of A. Ephrussi); 1:1000 anti-Glo (5B7; (Kalifa et al., 2006); 1:10,000 rabbit anti-kinesin heavy chain (Cytoskeleton). Proteins were visualized by chemiluminescence (Roche) and quantified using ImageJ.

### 7.7 qPCR

Total RNA was extracted and reverse transcribed from dissected ovaries using a standard random-primed cDNA protocol (Qiagen). qPCR was conducted using gene-specific primers and three technical replicates from each of three biological replicates were performed per genotype (TaqMan, ThermoFisher). Data are expressed as the amount of target cDNA relative to the amount of rpl7 cDNA and normalized to the wild-type transgene.

### 7.8 smFISH

smFISH was performed on ovaries dissected from well-fed females as described previously (Abbaszadeh and Gavis, 2016) using probes for either the coding region of grk or osk conjugated to ATTO-565 of ATTO-647(ATTO-Tec), respectively. DNA was visualized with DAPI (1:1000; Molecular Probes). Oocytes were imaged using a Nikon A1R confocal microscope with a 60x/1.4 NA oil objective.
7.9 Analysis of Eggshell and Embryonic Phenotypes

Embryos were collected on yeasted apple juice plates (Wieschaus and Nüsslein-Volhard, 1986) at RT and immediately analyzed for eggshell phenotypes. For the analysis of embryonic phenotypes, the embryos were aged for >24 hours at RT, after which larval cuticle preparations were made (Wieschaus and Nüsslein-Volhard, 1986).

7.10 Immunoprecipitation

Ovaries from well-fed females were dissected in PBS and homogenized in IP buffer [10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 0.1mM PMSF, 1x complete protease inhibitor cocktail (Roche)]. Extracts were subsequently cleared by centrifugation at 13,000 rpm for 10 min at 4° C and supplemented with 100 µg/ml RNase A and 100 units/ml RNase One (Promega). Cleared extracts were then incubated with GFP_TRAP®_M beads (ChromoTek) for 2 hours at 4° C. Eluted protein complexes were resolved by SDS-PAGE, transferred to nitrocellulose membrane and detected by immunoblotting and chemiluminescence. Primary antibody concentrations: 1:20 anti-Hfp (6G10); 1:5000 anti-Hrp48.

7.11 RNA co-immunoprecipitation

Ovaries from well-fed females were enriched for late-stage oocytes as described previously (Andrews et al., 2011), dissected in PBS and homogenized on ice in RNase-Free IP buffer [10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, 0.1mM PMSF, 1x complete protease inhibitor cocktail (Roche), 300 U RNasin (Promega)]. Extracts were subsequently cleared by centrifugation at 13,000 rpm for 10 min at 4° C and incubated with GFP_TRAP®_M beads (ChromoTek) for 2 hours at 4° C. RNA was isolated with TRIzol (Invitrogen), precipitated with ethanol, and resuspended with DEPC-treated distilled H₂O. For RT-PCR, RNA was eluted from the beads by incubating samples for 5 minutes at 80° C. Half of the eluted RNA was reverse transcribed using Superscript IV (Invitrogen) and primers for
either nos (5′-GCGATCAAGGCGGAATCG-3’ and 5′-ATAGGATCCGAAAGTGTTCTTGGCTA-3’), osk (5′-TCTTGATGCTCGATATCGTG-3’ and 5′-ACCAGATCTTCGTAAGGTCC-3’) or tub (5′-CCAAGCGGCACATTCAATCGTC G-3’ and 5′-CATAATTATTGGATATTGTAATGA-3’) and the other half was treated similarly, except that Superscript IV was omitted (-RT). Target genes were then amplified by PCR using the same primers used for RT-PCR.

7.12 Viability Assay
Viability assays were conducted to determine the extent to which GFP-Glo proteins were able to rescue the lethality observed in glo homozygous mutant flies. Df glo is a deficiency chromosome that contains a chromosomal deletion spanning the glo gene locus. The Df glo/FRT82B glo− genotype is normally lethal but can be rescued by expression of a functional GFP-Glo protein, where it is expected to occur in 33% F1 progeny. Viability is calculated by comparing the expected frequency of this genotype to the observed frequency.

7.13 Statistical methods
The Kd values are reported as mean +/- standard error of the mean (SEM), which indicates the discrepancy among the technical replicates. We also examined the error of the Kd values for each technical replicate and calculated mean errors for the sets of three technical replicates. This analysis indicated that the SEM is a reasonable representation of the accuracy of the Kd values.

P-values for Glo qRRM1,2 mutant binding to the TCEGAG/UUC or AGGGA RNA relative to wild-type Glo qRRM1,2 protein were calculated using a one-way ANOVA, excluding the double mutants (qRRM1,2W58A,Y155A and qRRM1,2R52A,K149A) whose binding was too weak to determine precisely and are reported as a lower limit of >60 μM. The p-value for qRRM1,2W58A,Y155A binding to 5´-AGGGA RNA relative to wild-type qRRM1,2 was calculated using an unpaired, one-tailed t-test with Welch’s correction and the p-values for qRRM1,2R52A,K149A binding to TCEGAG/UUC and TCEI RNAs relative to wild-type qRRM1,2 were
calculated using an unpaired, two-tailed t-test with Welch’s correction for unequal variances. Differences in Nos protein levels in late ovaries expressing nos transgenes were analyzed by one-way ANOVA.

### 7.14 Accession Numbers

Coordinates and structure factors for qRRMs 1-3 have been deposited in the RCSB Protein Data Bank (www.rcsb.org) with the accession codes: 5UZG, 5UZM, 5UZN.
A.1 Script to Identify Potential RNA Targets of Glo on a Genome-Wide Basis

# Author Robert Leach (Bioinformatics Core, Lewis-Sigler Institute)

# This script was imported into Galaxy (Princeton University installation) and is designed to
# search for UA-rich motifs that may form a double-stranded structure with a proximate G-tract
# consisting of GGG. In order to search for UA-rich motifs that would likely form double-
# stranded structures, we searched specifically for transcripts containing at least two UA-rich
# motifs that each consist of at least four consecutive UA nucleotide base pairs (e.g., UAUA
# or AUAU) and that are separated by at least 3-20 nucleotides. Those transcripts were then
# queried for proximate G-tracts by looking for a GGG sequence within 5-25 nucleotides of
# either UA-rich sequence forming the stem. This resulted in a list of ~800 potential RNA
# targets that include known targets of Glo, like nos and osk.

// script = MotifSearch

//Command to run MotifSearch:

motifSearch.pl -s 'TATA.{3,20}A
TAT.{5,25}GGG|GGG.{5,25}TATA.{3,20}ATAT|ATAT.{3,20}TATA.{5,25}GGG|
GGG.{5,25}ATATA.{3,20}TATAT' -e -i dmel-all-transcript-r6.20.fasta --verbose -d forward > uauau3-20auaua5-25ggg.out

#!/bin/tcsh

#USAGE: transcript_presence_filter.tcsh count_file motifSearch_output_files
#EXAMPLE: transcript_filter.tcsh transcript_counts_from_galaxy.txt uauau*.out

setenv COUNTSFILE `echo "$argv"   | cut -f 1 -d " "`
setenv MOTIFFILES `echo "$argv"   | cut -f 2-999 -d " "`

echo
echo "Running $0"
echo "COUNTS FILE: $COUNTSFILE"
echo "MOTIF FILES: $MOTIFFILES"

sort -k 2 -n "$COUNTSFILE" | grep -v "\t0" | cut -f 1 > "$COUNTSFILE.1ormore.ids"

echo -n "Number of transcripts with at least 1 read mapped: ":
cat "$COUNTSFILE.1ormore.ids" | wc -l

echo

foreach f ( $MOTIFFILES )

echo "DOING $f"
grep -f "$COUNTSFILE.1ormore.ids" "$f" > "$f filt"

end
References


