A high-throughput pipeline for the production of synthetic antibodies for analysis of ribonucleoprotein complexes

HONG NA,[1,8] JOHN D. LAVER,[2,8] JOUHYUN JEON,[1] FATEH SINGH,[1] KRISTIN ANCEVICIUS,[3,4] YUJIE FAN,[2] WEN XI CAO,[2] KUN NIE,[1,2] ZHENGLIN YANG,[2] HUA LUO,[2] MIRANDA WANG,[2,5] OLIVIA RISSLAND,[2,5] J. TIMOTHY WESTWOOD,[3,4] PHILIP M. KIM,[1,2,6] CRAIG A. SMIBERT,[2,7] HOWARD D. LIPSHITZ,[2] and SACHDEV S. SIDHU[1,2]

[1]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada M5S 3E1
[2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8
[3]Department of Biology, [4]Department of Cell and Systems Biology, University of Toronto at Mississauga, Mississauga, Ontario, Canada L5L 1C6
[5]Research Institute, Hospital for Sick Children, Toronto, Ontario, Canada M5G 0A4
[6]Department of Computer Science, [7]Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8

ABSTRACT

Post-transcriptional regulation of mRNAs plays an essential role in the control of gene expression. mRNAs are regulated in ribonucleoprotein (RNP) complexes by RNA-binding proteins (RBPs) along with associated protein and noncoding RNA (ncRNA) cofactors. A global understanding of post-transcriptional control in any cell type requires identification of the components of all of its RNP complexes. We have previously shown that these complexes can be purified by immunoprecipitation using anti-RBP synthetic antibodies produced by phage display. To develop the large number of synthetic antibodies required for a global analysis of RNP complex composition, we have established a pipeline that combines (i) a computationally aided strategy for design of antigens located outside of annotated domains, (ii) high-throughput antigen expression and purification in *Escherichia coli*, and (iii) high-throughput antibody selection and screening. Using this pipeline, we have produced 279 antibodies against 61 different protein components of *Drosophila melanogaster* RNPs. Together with those produced in our low-throughput efforts, we have a panel of 311 antibodies for 67 RNP complex proteins. Tests of a subset of our antibodies demonstrated that 89% immunoprecipitate their endogenous target from embryo lysate. This panel of antibodies will serve as a resource for global studies of RNP complexes in *Drosophila*. Furthermore, our high-throughput pipeline permits efficient production of synthetic antibodies against any large set of proteins.

Keywords: RNA-binding protein; phage display; synthetic antibody; Fab; ribonucleoprotein complex

INTRODUCTION

Post-transcriptional regulatory mechanisms are essential for the proper control of gene expression, and include processes occurring throughout the life cycle of an mRNA, such as polyadenylation, nuclear export, subcellular localization, translation, and degradation. These processes are controlled by *trans*-acting factors—RNA-binding proteins (RBPs) and noncoding RNAs (ncRNAs)—that, together with associated partner proteins, recognize specific sequences or structures—*cis*-elements—present in transcripts. The RBPs, along with their protein and ncRNA cofactors, reside in ribonucleoprotein (RNP) complexes, together determining the post-transcriptional fate of their mRNA targets.

Studies in recent years have sought to gain a genome-wide view of the activity of RNPs (Tenenbaum et al. 2000, 2002; Ule et al. 2003, 2005; Gerber et al. 2004; Keene 2007; Hogan et al. 2008; Morris et al. 2010; Laver et al. 2013, 2015; Chen et al. 2014; Stoiber et al. 2015). One powerful approach involves the identification of all of the RNA and protein components of particular RNPs by immunoprecipitation (IP) of the protein of interest followed by identification of the associated transcripts by microarray or next-generation sequencing analysis, and identification of the associated proteins by mass spectrometry. Analyses of the bound RNAs can yield

[8]**These authors contributed equally to this work.**

Abbreviations: CLIP, cross-linking and immunoprecipitation; ELISA, enzyme-linked immunosorbent assay; Fab, antigen-binding fragment (of an IgG); IP, immunoprecipitation; IP-MS, IP-mass spectrometry; miRNA, microRNA; RBD, RNA-binding domain; RBP, RNA-binding protein; RIP, RNA coimmunoprecipitation; RNP, ribonucleoprotein

Corresponding authors: c.smibert@utoronto.ca, howard.lipshitz@utoronto.ca, sachdev.sidhu@utoronto.ca

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.055186.115.

insights into how proteins recognize and regulate their target transcripts, and knowledge of the bound RNAs together with interacting proteins can help to elucidate the molecular, cellular, and biological processes in which the RNPs participate. These studies have also revealed fundamental aspects of post-transcriptional regulation, such as the observation that RNAs bound by a particular RBP tend to encode proteins that are functionally related (Gerber et al. 2004; Keene 2007; Hogan et al. 2008; Morris et al. 2010).

The genomes of most organisms encode hundreds of RBPs and ncRNAs, each of which likely associates with hundreds of target mRNAs. Conversely, a study in *Saccharomyces cerevisiae* estimated that an average mRNA is itself associated with ~30 different RBPs during its lifetime (Hogan et al. 2008). More recently, analyses of 20 different RBPs in *Drosophila melanogaster* tissue culture cells (Stoiber et al. 2015) identified so-called HOT (high occupancy target) RNAs that were bound by a majority of the RBPs assayed, suggesting the existence of mRNAs whose post-transcriptional regulation is likely to be very complex. This study also revealed that the mRNAs and proteins bound to the assayed RBPs are enriched for functions in RNA metabolism, further highlighting the fact that post-transcriptional regulatory factors function together, and that their mRNAs are themselves highly regulated. Taken together, these observations suggest that a complete understanding of post-transcriptional regulatory processes will require a global view of the regulation by all *trans*-factors and, hence, identification of the protein, ncRNA, and mRNA components of all RNP complexes acting in a particular biological context.

Currently, there are two major strategies for analysis of RNP complex composition. The first, referred to as RNP immunoprecipitation (RIP), carries out IP of an RNP complex protein under mild conditions with no protein–RNA cross-linking, thus capturing the entire complex or set of complexes that contain that protein (Tenenbaum et al. 2000; Keene et al. 2006). A related method involves the use of chemical cross-linking prior to IP, in order to cross-link interacting protein and RNA components of RNPs and, again, capture the entire RNP complex. The second major strategy uses UV light to cross-link RBPs to directly associated RNAs prior to IP (CLIP and PAR-CLIP) (Ule et al. 2003, 2005; Hafner et al. 2010; Spitzer et al. 2014). These experiments are conducted under conditions such that only a particular RBP and its directly bound target sequences are purified. This allows for the determination of the RBP's binding site within an mRNA but does not permit identification of the indirectly associated RNAs and proteins.

Regardless of the strategy used, one important limitation that has prevented a truly global analysis of endogenous RNP complexes in most systems has been the lack of tools required to allow the IP of all of the RNP complex proteins encoded by an organism's genome. Although there are a number of options available to immunoprecipitate individual proteins, most have significant caveats. For instance, methods relying on epitope tags often involve the use of heterologous expression, which typically results in overexpression and can, therefore, lead to spurious RBP–RNA interactions (Laver et al. 2013). Even when endogenous expression levels are achieved, the tags may affect the function of the protein. Moreover, these tagging approaches are often not amenable to high-throughput in vivo analyses in complex, multicellular organisms.

To allow the IP of endogenous RNP complexes, it is, therefore, desirable to develop antibodies against their protein components. The generation of synthetic antibodies by phage display technology offers an approach for high-throughput production of functional monoclonal antibodies (Adams and Sidhu 2014). To generate synthetic antibodies, libraries are engineered to express and display antigen-binding fragments (Fabs) on the surfaces of bacteriophage particles, and Fabs that recognize an antigen of interest are selected from the library through multiple rounds of in vitro selection. Following the selections, the DNA sequence of the Fabs can be determined, and they can be expressed and purified from *Escherichia coli*. Such synthetic antibody libraries have been used to generate high-affinity antibodies against a wide variety of antigens, and offer a number of advantages (Sidhu 2012). For instance, given the recombinant nature of synthetic antibodies, they can be engineered or tagged in a variety of ways, and they represent an inexhaustible resource. In addition, through subsequent mutagenesis of a particular Fab and additional selections, higher affinity antibodies against a particular protein can be produced in a process referred to as affinity maturation (Li et al. 2009; Huang et al. 2015). Finally, synthetic antibodies are not associated with the ethical issues related to the use of antibodies raised in animals.

Although we have previously used synthetic Fabs produced in a low-throughput manner to identify endogenous RBP–RNA interactions genome-wide for three *Drosophila* RBPs, Staufen, Brain tumor, and Pumilio (Laver et al. 2012, 2013, 2015), their general utility for elucidation of RNP complex composition remained unclear. One potential issue relates to the fact that stable, independently folding, compact protein regions are required as antigens when producing synthetic antibodies (Hornsby et al. 2015). In principle, RNA-binding domains could be used as such antigens for RBPs; however, a subset of the Fabs would then be likely to disrupt RNA–protein interactions (Laver et al. 2012) and, thus, would not be useful for identification of the complex's mRNA components. Although non-RNA-binding domains could be used as antigens, the fact that a substantial fraction of RBPs have no annotated domains other than their RNA-binding domain(s) suggested that generating Fabs useful for the elucidation of RNP complex composition could be problematic.

Here, we report a high-throughput pipeline for the production of synthetic Fabs for use in global studies of RNP complexes. Our pipeline combines methods for antigen design, high-throughput antigen expression and purification from *E. coli*, and high-throughput antibody selection and

screening (Fig. 1). Importantly, our method incorporates several novel aspects compared with other high-throughput synthetic antibody production methodologies (Schofield et al. 2007; Hornsby et al. 2015; Huang et al. 2015). These include (i) a computationally aided approach for designing effective antigens that lie outside of annotated domains, (ii) improved protocols for high-throughput antigen expression and purification, and (iii) a streamlined strategy to screen for individual antibodies and subclone them for expression in *E. coli*. Using this pipeline, we designed, purified, and performed synthetic antibody selections for 93 antigens representing 90 different RNP complex proteins from *D. melanogaster*, obtaining a total of 279 antibodies against 61 of these proteins. Together with the antibodies produced in low-throughput experiments, we now have a panel of 311 Fabs against 67 RNP complex proteins. We demonstrate that our antibodies are effective in IP of their endogenous target proteins from *Drosophila* embryos, underscoring their utility in global studies of RNP composition, as well as the usefulness of our pipeline for high-throughput production of functional synthetic antibodies against any large set of proteins.

RESULTS

To develop the pipeline outlined in Figure 1, we selected 90 proteins encoded by the *D. melanogaster* genome with a variety of known and predicted post-transcriptional functions. Sixty of these proteins have canonical RNA-binding domains, whereas 30 either bind RNA directly but do not possess a canonical binding domain or are likely to associate with RNA indirectly as part of the RNP complex (Table 1).

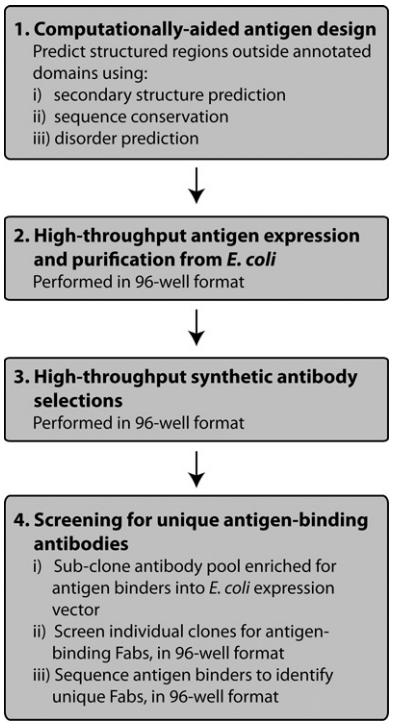

FIGURE 1. Overview of the high-throughput pipeline for the production of synthetic antibodies.

Computationally guided identification of antigenic protein regions

Our first challenge was to identify protein regions outside of the RNA-binding domain that would serve as optimal antigens. In particular, we wanted to select regions that are likely to fold independently into stable structures, since such regions are required to optimize the chances of yielding antibodies by phage display approaches (Hornsby et al. 2015). First, we searched for annotated domains since these have served as effective antigens for synthetic antibody production in other studies (Colwill and Graslund 2011; Huang et al. 2015). However, as described above, we wanted to avoid choosing canonical RNA-binding domains as antigens in order to minimize the chances of producing antibodies that might interfere with RBP–RNA interactions. Although other regions of a protein may be involved in protein–protein interactions important for RNP complex formation or stability, in most cases, these have not been mapped. Furthermore, for 45 of the 60 RBPs on our list that have canonical RNA-binding domains (75%), these were the only annotated domains present.

To identify potential structured regions lying outside of these domains, and still allow the antigen design process to be amenable to high-throughput approaches, we developed a computational algorithm that integrates physicochemical properties of amino acid sequences and structure-related features of proteins, some of which have been described previously as useful for optimization of successful expression of a protein fragment (Dyson et al. 2004). The features analyzed by the algorithm were (i) annotated protein domains (identified using InterProScan, see Zdobnov and Apweiler 2001); (ii) regions of predicted secondary structure (from PSIPRED, see Buchan et al. 2010); (iii) predicted regions of disorder (from DISOPRED2, see Ward et al. 2004); (iv) sequence conservation among 12 *Drosophila* species, calculated based on orthologs provided on FlyBase (http://flybase.org) (using Rate4Site, see Mayrose et al. 2004); (v) predicted antigenicity (using Antigenic from EMBOSS, see Rice et al. 2000); and (vi) predicted hydrophobicity (using Hmoment from EMBOSS, see Rice et al. 2000).

The output from the algorithm was used as follows. Fifteen of the 60 RBPs with canonical RNA-binding domains and 10 of the 30 RNP complex proteins that lacked canonical RNA-binding domains or that associated with RNA indirectly were predicted by InterProScan to have one or more annotated domains that were not canonical RNA-binding domains. For these 25 proteins, we chose annotated non-RNA-binding domains as antigens. For four of the canonical RBPs for which there were no annotated domains other than

TABLE 1. List of RNP complex proteins for which antigens were produced and Fab screens conducted using either low- or high-throughput strategies

Protein name	RBD name	Antigen coordinates (isoform_amino-acid-range)	Annotated domain chosen as antigen?	# Fabs obtained
Antigens generated and screened through the high-throughput pipeline				
UNR	CSP	CG7015-PC_586-685	–	4
BEL	DEAD	CG9748-PA_1-61	–	2
CG7878	DEAD	CG7878-PA_210-263	–	4
eIF-4A	DEAD	CG9075-PA_240-403	Helicase carboxyl-terminal	1
eIF4AIII	DEAD	CG7483-PA_1-240	Q motif and DEAD box helicase domain	2
GEM3	DEAD	CG6539-PA_684-874	–	1
ME31B	DEAD	CG4916-PA_396-459	–	6
CG1434	dsRBD	CG1434-PA_272-369	–	4
CG8273	dsRBD	CG8273-PA_370-426	–	3
DCR-2	dsRBD, DEAD, and PAZ	CG6493-PA_220-359	–	1
DIP1	dsRBD	CG17686-PH_156-210	–	10
DROSHA	dsRBD	CG8730-PA_655-734	–	3
LOQS	dsRBD	CG6866-PA_1-130	–	2
PASHA	dsRBD	CG1800-PA_474-642	–	1
R2D2	dsRBD	CG7138-PA_186-311	–	11
DP1	KH	CG5170-PA_96-164	–	1
IMP	KH	CG1691-PA_231-293	–	2
MASK	KH	CG33106-PA_1779-2052	–	4
OSK	LOTUS	CG10901-PC_260-468	SGNH hydrolase-type esterase domain	2
AGO1	PAZ and PIWI	CG6671-PB_297-464 [AGO1 (2)]	PAZ	4
AGO2	PAZ and PIWI	CG7439-PE_1-115	Protein argonaute, amino-terminal	1
CG11123	Puf	CG11123-PA_612-665	–	2
ARET	RRM	CG31762-PD_130-270	–	2
CBP20	RRM	CG12357-PA_1-154	Full-length protein, includes RRM	1
CNOT4	RRM and ZnF	CG31716-PG_547-733	–	19
CSTF-64	RRM	CG7697-PA_353-419	Transcription termination and cleavage factor, carboxyl-terminal domain	4
FUS	RRM	CG8205-PH_1-45	–	1
GW	RRM	CG31992-PA_715-810	GW182 M domain	14
NONA	RRM	CG4211-PA_493-665	–	4
NONA-L	RRM	CG10328-PA_456-574	–	3
PABP	RRM	CG5119-PA_552-634	PABC domain	15
PABP2	RRM	CG2163-PA_1-80	–	3
SHEP	RRM	CG32423-PA_498-578	–	5
SWA	RRM	CG3429-PA_204-273	–	13
TBPH	RRM	CG10327-PA_1-106	–	6
TSU	RRM	CG8781-PA_1-63	–	4
XMAS-2	RRM	CG32562-PA_642-832	–	2
eIF-2α	S1	CG9946-PA_91-295	Translation initiation factor 2, α subunit, middle and carboxyl-terminal domains	2
SMG	SAM	CG5263-PA_197-280 [SMG (2)]	–	4
MEI-P26	TRIM-NHL	CG12218-PA_360-530	B-box carboxyl-terminal domain	8
MDLC	ZnF	CG4973-PA_96-194	–	3
MKRN1	ZnF	CG7184-PA_157-209	–	1
NOS	ZnF	CG5637-PA_91-184	–	4
ROQ	ZnF	CG16807-PA_95-396	–	2
UNK	ZnF	CG4620-PA_461-553	–	5
ZN72D	ZnF	CG5215-PE_409-540	Partial overlap with DZF domain	2
APT	–	CG5393-PE_412-469	–	20
BICD	–	CG6605-PB_399-511	–	12
CBP80	–	CG7035-PA_28-294	MIF4G-like, type 3	7
DIS3	–	CG6413-PA_810-945	Nucleic acid-binding, OB-fold	1
EGL	–	CG4051-PB_1-280	–	9

Continued

TABLE 1. *Continued*

Protein name	RBD name	Antigen coordinates (isoform_amino-acid-range)	Annotated domain chosen as antigen?	# Fabs obtained
eIF-4B	–	CG10837-PE_309-388	–	5
eIF4G	–	CG10811-PA_27-254	–	2
eIF4G2	–	CG10192-PA_1905-2072	MIF4G-like domain	6
eIF-5A	–	CG3186-PA_13-88	Ribosomal protein L2 domain 2	1
HRG	–	CG9854-PA_41-295	Poly(A) polymerase, central domain	2
LOST	–	CG14648-PB_20-280	5-Formyltetrahydrofolate cyclo-ligase-like domain	3
MAEL	–	CG11254-PA_51-120	–	5
MSL-3	–	CG8631-PA_200-320	MRG domain	3
QIN	–	CG43726-PB_244-373	–	4
SMN	–	CG16725-PA_1-53	–	1
SPN-E	DEAD	CG3158-PA_646-730	–	0
ADAR	dsRBD	CG12598-PL_250-300	Partial overlap with adenosine deaminase/editase domain	0
CG5641	dsRBD	CG5641-PA_154-330	DZF	0
DCR-1	dsRBD, DEAD, and PAZ	CG4792-PA_971-1086	–	0
BICC	KH	CG4824-PD_279-435	–	0
AGO1	PAZ and PIWI	CG6671-PB_113-247 [AGO1 (1)]	Protein argonaute, amino-terminal	0
AGO1	PAZ and PIWI	CG6671-PB_580-915 [AGO1 (3)]	PIWI	0
AGO3	PAZ and PIWI	CG40300-PD_125-283	–	0
AUB	PAZ and PIWI	CG6137-PC_37-201	Protein argonaute, amino-terminal	0
PIWI	PAZ and PIWI	CG6122-PA_96-253	Protein argonaute, amino-terminal	0
CG3594	RRM	CG3594-PA_193-250	–	0
CYP33	RRM	CG4886-PA_141-300	Cyclophilin-type peptidyl-prolyl *cis–trans* isomerase domain	0
eIF3-S9	RRM	CG4878-PA_594-690	–	0
ORB	RRM	CG10868-PC_407-559	Cytoplasmic polyadenylation element-binding protein, ZZ domain	0
SPEN	RRM	CG18497-PH_2316-2605	–	0
RRP4	S1	CG3931-PA_173-298	KH	0
SMG	SAM	CG5263-PA_69-155 [SMG (1)]	–	0
ARMI	–	CG11513-PC_444-624	–	0
BCD	–	CG1034-PG_401-494	–	0
CPSF160	–	CG10110-PB_491-683	–	0
CUP	–	CG11181-PB_105-219	–	0
DCP2	–	CG6169-PB_380-564	–	0
eIF-4E	–	CG4035-PA_71-259	–	0
EXU	–	CG8994-PA_33-197	–	0
HEPH	–	CG31000-PA_164-217	–	0
IRP-1B	–	CG6342-PA_1-251	Aconitase/3-isopropylmalate dehydratase large subunit, α/β/α, subdomain 1/3	0
MSI	–	CG5099-PH_1-90	–	0
OTU	–	CG12743-PA_17-209	OTU domain	0
RRP45	–	CG9606-PA_197-297	Exoribonuclease, phosphorolytic domain 2	0
TRAL	–	CG10686-PE_546-642	–	0
TUD	–	CG9450-PA_2467-2515	–	0
VIR	–	CG3496-PA_1233-1379	–	0
Antigens generated and screened through low-throughput approaches				
HOW	KH	CG10293-PA_261-369	–	1
FMR1	KH	CG6203-PA_360-473	Fragile X-related 1 protein, carboxyl-terminal core	2
ORB2	RRM	CG43782-PH_163-446	Contains nucleotide-binding α-β plait domain	18
STAU[a]	dsRBD	CG5753-PA_113-310	–	2
PUM[b]	Puf	CG9755-PA_726–882	–	1
BRAT[b]	TRIM-NHL	CG10719-PA_375–565	–	1

[a]Laver et al. (2012, 2013).
[b]Laver et al. (2015).

their RNA-binding domains (as well as for AGO1 for which we had already selected a non-RBD antigen), we chose the RNA-binding domains as antigens. For the remaining 61 proteins (41 with canonical RNA-binding domains and 20 other RNP complex proteins), we manually inspected the outputs from (ii) through (vi) and designed antigens that mapped outside of annotated domains but, instead, were from regions that met one or more of the following criteria: had predicted secondary structure, were conserved among *Drosophilids*, and/or had low predicted disorder. Where possible, we further refined our choice of antigens by selecting regions with high predicted antigenicity and low predicted hydrophobicity. Whenever possible, the boundaries of the antigens were chosen to correspond precisely to those of the predicted regions of secondary structure or, for antigens where secondary structure was not predicted, boundaries were selected based on regions of conservation. As a final criterion, we selected regions present in all predicted protein isoforms (i.e., present in all mRNA splice variants as defined on FlyBase, http://flybase.org).

Examples of the algorithm output and choice of antigens are shown in Figure 2, and output of the algorithm is available upon request for all 90 RNP complex proteins. Using this approach, we designed 93 antigens for the 90 proteins (Table 1). Of these, 31 represented annotated domains and the remaining 62 represented regions that were not annotated but met the criteria described above.

High-throughput expression and purification of antigens

The 93 antigens were expressed and purified from *E. coli* as fusion proteins, with amino-terminal hexaHis and GST tags. We performed the purifications according to our previously published high-throughput protocol for protein expression and affinity purification with the hexaHis tag (Huang and Sidhu 2011). Seventy-two of the 93 antigens gave average yields of >10 μg of total protein from 2.4 mL of bacterial culture (Supplemental File 1). To further optimize the yields of the purified antigens, we tested whether antigen yields could be improved by the addition of 1% sodium lauroyl sarcosinate (sarkosyl, an ionic detergent used to increase solubility of proteins upon purification from *E. coli*) (see Frankel et al. 1991) during the lysis and purification. Upon inclusion of sarkosyl, the yields of 14 antigens increased >1.5-fold on average, whereas yields of only four antigens decreased by a similar amount (Supplemental File 1). Notably, the addition of sarkosyl was particularly effective at improving the recovery of antigens that gave low yields in the absence of this reagent: Of the 14 antigens whose yields increased >1.5-fold when sarkosyl was included, eight gave yields of <10 μg and 11 gave yields of <15 μg in the absence of sarkosyl. Overall, 80 antigens gave average yields of >10 μg from 2.4 mL of bacterial culture when sarkosyl was included in the purifications (Supplemental File 1).

Examination of the quality of the antigens purified in the presence of sarkosyl, by SDS-PAGE followed by "InstantBlue" staining, revealed that 89 of the 93 preparations included protein at or near the molecular weight predicted for the hexaHis–GST–antigen fusion protein (Fig. 3). This high success rate provided strong evidence that our computationally guided antigen design approach successfully identified structured regions lying outside of annotated domains.

High-throughput antibody selections

We performed high-throughput synthetic antibody selections using Library F (Persson et al. 2013), a highly validated phage-displayed Fab library that we previously used to generate antibodies against RBPs in a low-throughput manner (Laver et al. 2012, 2015). We applied Library F to 96-well plates in which each well was coated with a different antigen. Given that the vast majority were successfully purified as at least partially intact protein, we performed selections against all 93 antigens. Moreover, to further assess the optimal conditions for antigen purification and their impact on the outcome of the selections, we performed parallel selections with antigens purified with or without sarkosyl.

For each set of antigens, four rounds of selection were carried out, with the library preincubated with GST before application to the selection plates at each round to reduce the number of GST-specific and nonspecific Fab-phage (see Materials and Methods). After the fourth round of selection, success for each antigen was assessed by determining the enrichment of antigen-binding Fab-phage in the final phage pool. This enrichment was measured using ELISAs in which the binding to antigen-coated wells or GST-coated wells was compared for the final Fab-phage pool. Pools with an antigen-binding to GST-binding ratio greater than two were considered to be enriched for antigen binders. As shown in Figure 4A, these ELISAs indicated that, for antigens purified either in the presence or absence of sarkosyl, the selections were successful in enriching for antibodies against a large proportion of antigens. As expected, based on the increased antigen yields obtained upon purification with sarkosyl, 59 antigens purified in the presence of sarkosyl, but only 38 of those purified in the absence of sarkosyl, generated enriched Fab-phage pools (Fig. 4A; Supplemental File 1). Moreover, only five of the 38 antigens that generated enriched pools when purified without sarkosyl did not generate enriched pools when purified with sarkosyl, and two of the five were antigens whose yields decreased >1.5-fold upon addition of sarkosyl. Together, these data verify that inclusion of sarkosyl improves the high-throughput antigen purification protocol. Importantly, the presence or absence of sarkosyl during antigen extraction and purification did not appear to affect the conformation of the purified antigens, as Fab-phage pools selected against antigens purified in the presence of sarkosyl still recognized antigen that was purified without sarkosyl, and vice versa (Supplemental File 2).

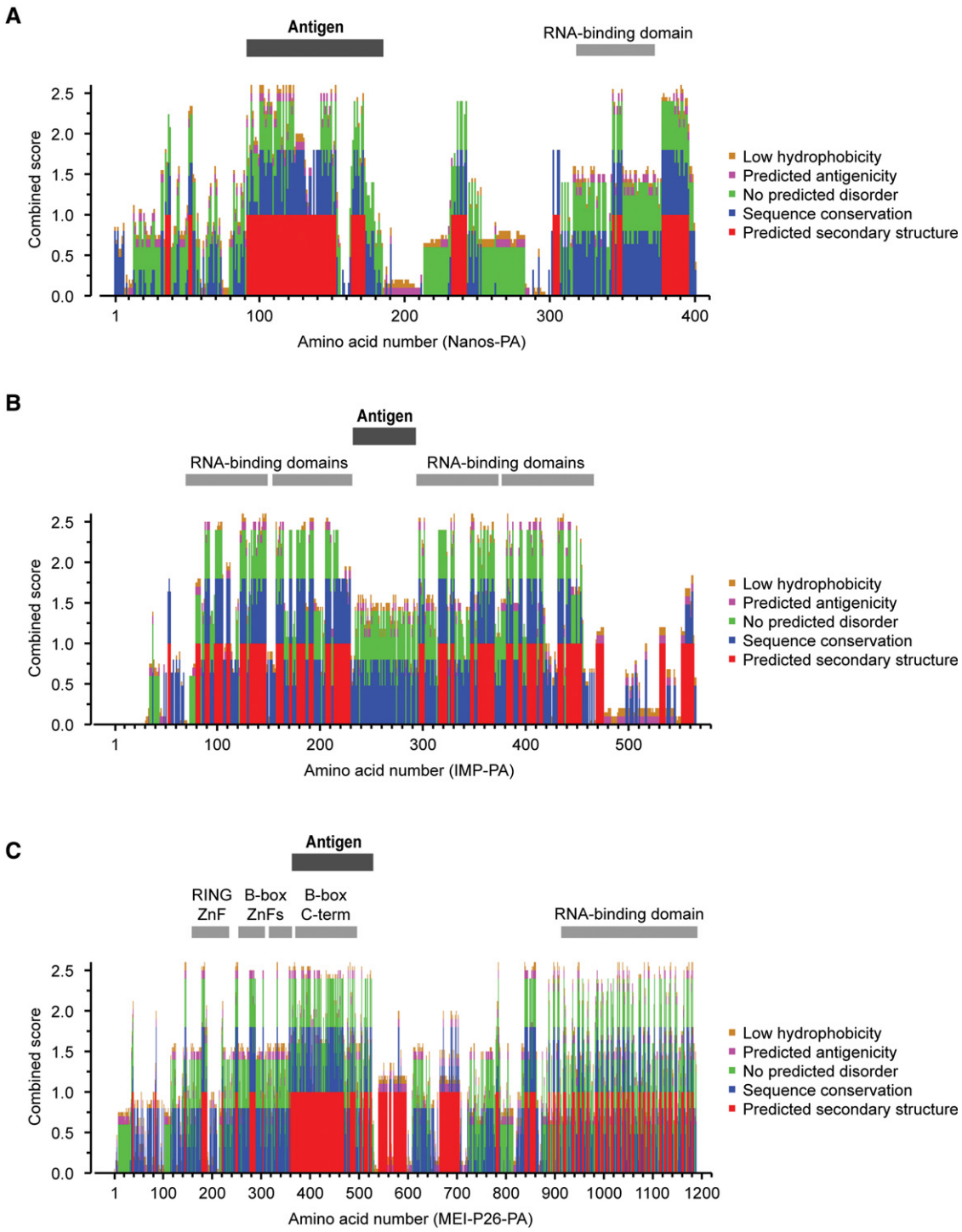

FIGURE 2. Examples of output from the computational antigen design algorithm. Output from the computational antigen design algorithm is shown for three RNP complex proteins, (*A*) Nanos, (*B*) IMP, and (*C*) MEI-P26, represented as stacked bar plots depicting weighted scores for the different sequence features analyzed, for each residue. The scores are presented such that for each sequence feature, a higher score indicates a better candidate region for an antigen. The features depicted are as follows: predicted secondary structure (assessed by PSIPRED) where a score of 1 indicates predicted secondary structure and a score of 0 indicates no predicted secondary structure; sequence conservation among 12 *Drosophila* species (calculated by Rate4Site), with scores ranging from 0 to 0.8 for each residue where 0.8 indicates maximal conservation; predicted disorder (assessed by DISOPRED) where a score of 0 indicates predicted disorder and a score of 0.6 indicates no predicted disorder; predicted antigenicity (assessed by antigenic from EMBOSS) where 0.1 indicates predicted antigenicity and 0 indicates no predicted antigenicity; hydrophobicity (assessed by Hmoment from EMBOSS) with scores from 0 to 0.1, where 0.1 indicates no hydrophobicity and 0 indicates hydrophobicity. The relative weights of the different features approximately reflect the relative importance given to each feature in selecting antigens. Bars *above* the plots indicate annotated domains and the regions chosen as antigens for each protein.

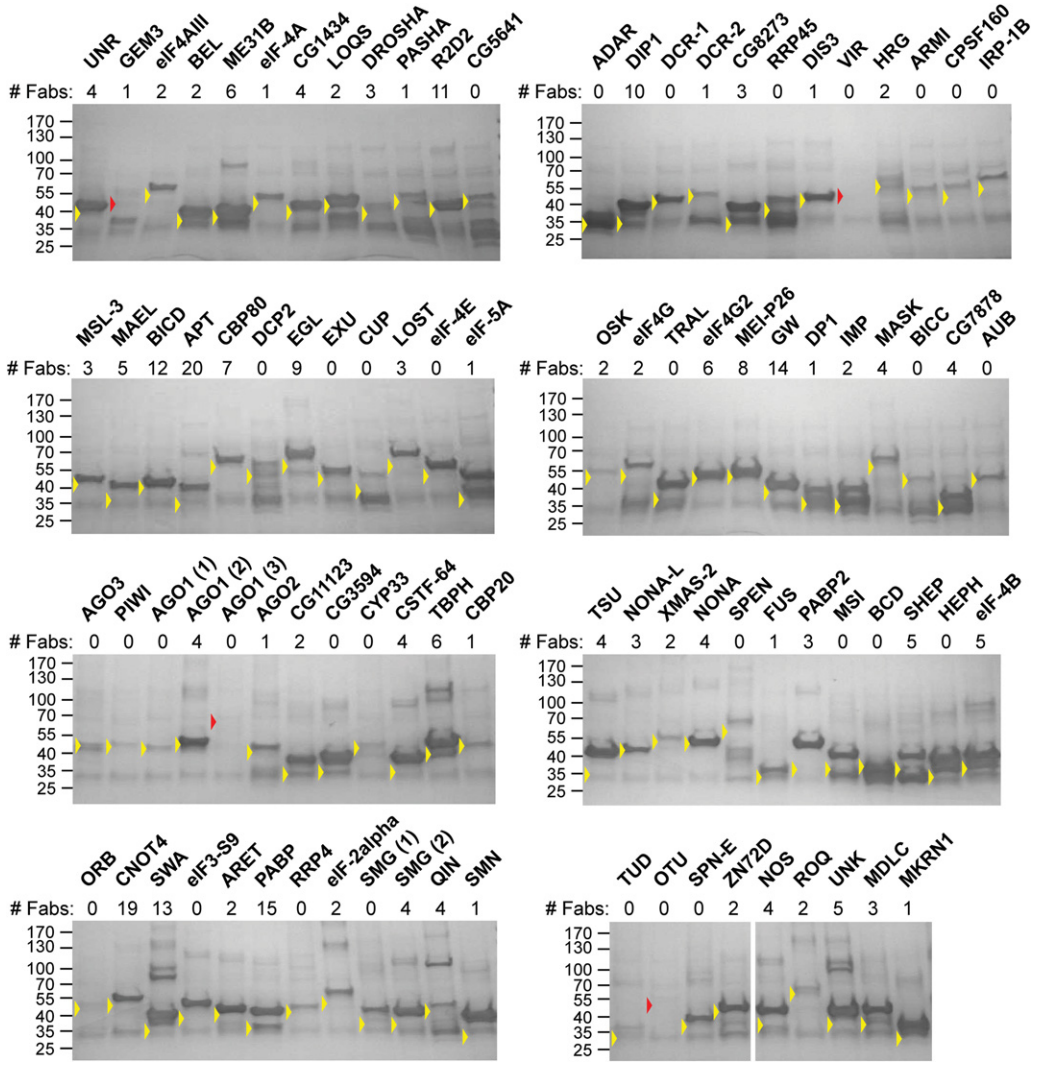

FIGURE 3. Purification of 93 RNP complex protein antigens and the number of synthetic Fabs obtained against each antigen. The 93 antigens were expressed and purified in *E. coli*, and purified antigens were visualized by SDS-PAGE followed by staining with InstantBlue. Molecular mass markers (kDa) are shown to the *left* of each gel, and arrowheads indicate the expected size of each antigen: Yellow arrowheads denote antigens for which some protein is present at or near the expected size, and red arrowheads denote antigens for which very little or no protein is present at the expected size. The number of Fabs obtained for each antigen after synthetic antibody selections is also indicated *above* each lane. For AGO1 and SMG, which had more than one antigen, numbers *after* the protein name indicate the identity of each antigen (see Table 1).

As a further point of optimization, in addition to the aforementioned selections, which were performed using naive Library F, selections were performed for each set of antigens using the phage that had been previously incubated once with the set of 93 antigens, and then repooled (see Materials and Methods). This was done to assess whether reusing the library might be a cost-effective approach to increase the coverage of the library's diversity for each antigen (i.e., each antigen would be exposed to a greater absolute number of phage since each antigen would be exposed both to the naive library and the repooled library). As measured by ELISA, selections with once-used library yielded Fab-phage pools enriched for binders for a similar number of antigens to that observed using the naive library although, in this case, the difference between the number of successful selections for antigens purified with sarkosyl versus without was less dramatic (Supplemental File 1). This suggests that repooling and reusing an antibody library after exposure to a panel of antigens is a high-throughput-amenable and cost-effective method to increase library coverage for individual antigens.

High-throughput screening and isolation of unique Fabs

We next sought to isolate individual antigen-binding Fabs from our enriched phage pools. To accomplish this in a high-throughput manner, we developed a new procedure to screen for, and purify, unique antigen-binding Fabs.

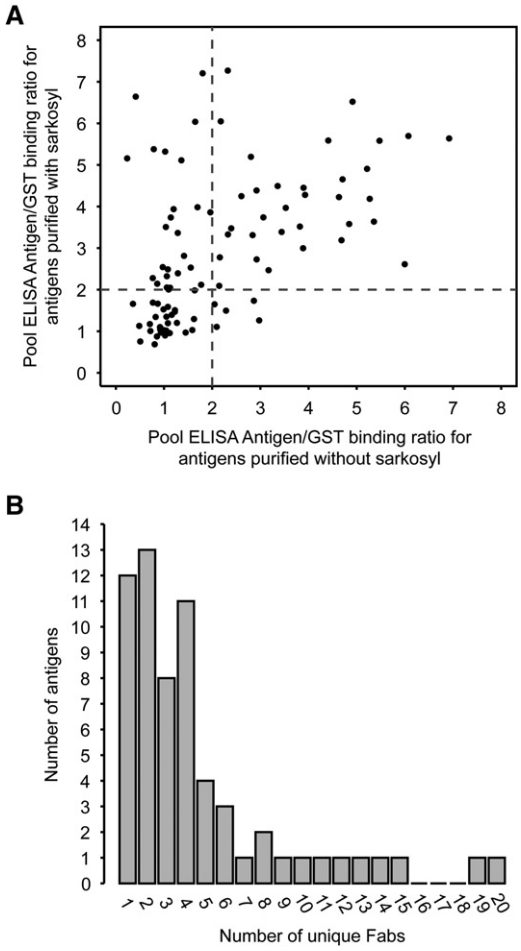

FIGURE 4. Results of high-throughput synthetic antibody selections. (*A*) Scatterplot comparing, for antigens purified in the presence of sarkosyl (*y*-axis) versus the absence of sarkosyl (*x*-axis), the antigen/GST-binding enrichment ratios determined by ELISA for the Fab-phage pools after four rounds of selection. Dashed lines indicate an antigen-/GST-binding ratio of 2, which we considered to be the minimum fold enrichment to consider the selections successful in enriching for antigen binders. (*B*) Histogram indicating the number of antigens for which different numbers of unique Fabs were obtained.

Previously, the isolation and preparation of individual Fabs typically involved (i) isolation of individual Fab-phage clones and screening for antigen-binding activity by clonal phage ELISAs, (ii) sequencing of antigen-binding Fab-phage to identify unique Fabs, (iii) PCR amplification of the Fab sequences from each of these, followed by (iv) individual subcloning of each antigen-binding Fab into an expression vector for expression and purification from *E. coli* (Hornsby et al. 2015).

To streamline this procedure and avoid subcloning each unique Fab individually, we established a new procedure in which we (i) PCR-amplified and subcloned the entire Fab-phage pool into a vector for expression from *E. coli*, (ii) isolated and screened individual Fab clones from this pool by clonal ELISAs, and (iii) sequenced antigen-binding clones to identify unique Fabs (see Materials and Methods for details). Using this new procedure, upon identification of unique Fabs, these could immediately be purified from *E. coli* as they were already inserted into the correct expression vector.

To identify unique antigen-binding Fabs using this new procedure, we first combined, for each antigen, the Fab-phage pools from the final round of any of the four selections (±sarkosyl, naive versus reused library) for which the pool ELISA data showed antigen-binding enrichment of at least twofold. Alternatively, for antigens that showed less than twofold enrichment by pool ELISA for all of the selection conditions, the pool from the selection condition with the highest antigen-/GST-binding ratio was chosen. These pools were subcloned into an expression vector, 24–48 individual clones were screened for antigen-binding activity by ELISA, and binding clones were sequenced. This led to the identification of a total of 279 unique Fabs against 61 different RNP complex proteins (Table 1; Figs. 3–7), an overall success rate of 66% (61 antigens with one or more Fabs out of 93 total antigens designed and screened).

Importantly, we achieved similar success rates for the 31 antigens that represented annotated domains (antibodies against 19 of 31 antigens, 61%) (Figs. 6, 7) and the 62 that mapped outside of annotated domains (antibodies against 42 of 62 antigens, 68%) (Fig. 5). Moreover, this success rate for antibody production is comparable with that achieved when high-throughput selections are carried out against well-characterized protein domains as antigens (e.g., SH3 domains) (Huang et al. 2015). In addition to validating our streamlined method for screening for individual antigen-binding Fabs, this success rate validates our computational approach for choosing predicted structured regions that lie outside of annotated domains as a highly effective method for designing antigens for synthetic antibody selection.

Validation of synthetic antibodies for immunoprecipitation of endogenous target proteins

Given that a major future goal is to characterize all endogenous RNP complexes via RIP, we next assessed the ability of the Fabs to immunoprecipitate their endogenous target proteins from early *Drosophila* embryo extract. To test this, we selected eight Fabs produced by our high-throughput pipeline against eight different RNP proteins (IMP, ME31B, PABP, NANOS, MEI-P26, EGL, DP1, and eIF4G). Since an average of more than four Fabs was obtained for each of the 61 RNP proteins (range: 1–20) (Figs. 4–7), we used clonal ELISA results as a guide to choosing which Fab to test, and selected the Fab with the highest antigen-/GST-binding ratio for each of the eight RBP antigens.

We expressed and purified the eight Fabs from *E. coli*, and performed IPs on extract from 0- to 3-h-old *Drosophila* embryos, isolating the Fabs and their bound target proteins via the Flag tag present on the Fab light chain. We then assessed whether the Fab had immunoprecipitated the endogenous

RBP name	Fab ID	CDR-L3	CDR-H1	CDR-H2	CDR-H3
		107 108 109 110 112 113 114 115 116	30 35 36 37 38 39	55 56 57 58 59 62 63 64 65 66	107 108 109 110 111 111.1 111.2 111.3 111.4 111.5 112.5 112.4 112.3 112.2 112.1 112 113 114 115
UNR	P001	Y Y F Y - G G L I	I Y S S S M	S I S P Y S S Y T S	Y G G V Y - - - - - - - - - - P V A L
	P002	S Y W - - - Y L F	I Y Y S Y I	S I S P Y Y G Y T S	A W S - - - - - - - - - - - - - W A F
	P003	S W Y - - - Y L I	I Y S S Y M	S I Y S S Y G Y T S	V A G H P - - - - - - - - - - W G A M
	P004	V W - - - - G L F	I S S S S M	S I Y P S S G Y T Y	G A Y G W - - - - - - - - - P Y S G L
BEL	P005	Y W Y - - - H L F	L Y Y S S M	S I Y P Y Y S Y T Y	A F H S Y Y Y W G - A A P P V S W A F
	P006	S Y F - - - F P I	F S G A D I	Y I Y T S G G D T A	Y G W S Y G - - - - - - - - W G G G M
CG7878	P007	Y W - - - - S P I	L S S S S M	S I Y S Y Y G Y T Y	G H G Y Y - - - - - - - - - - H S A L
	P008	W W - - - - S P I	I Y S S S I	S I Y S Y Y G Y T Y	G A S G Y - - - - - - - - - Y H H A I
	P009	W H Y - - - W P I	I S Y Y S M	S I S P Y Y S Y T Y	Y S Y W A F - - - - - - - S G P G A L
	P010	H Y W - - - S P I	L S Y Y Y I	S I S S S Y G Y T Y	Y S W W W P - - - - - - - - Y W Y G M
GEM3	P011	Y Y F - - - F P I	L S S Y S I	S I S P Y S S Y T S	A W - - - - - - - - - - - - - - - A F
ME31B	P012	G P Y P Y Y G L I	L S Y S S I	S I Y S Y Y G Y T Y	W Y W Y - - - - - - - - - - - G W A L
	P013	S S Y G - S Y L I	I Y S Y S I	S I Y P S Y G Y T Y	W S V Y Y S P A - - - Y Y S Y Y G G M
	P014	S W Y - - F G L I	L Y Y Y S I	S I S P S Y S Y T Y	F H P S - - - - - - - - - - - P W A L
	P015	A P W F - G S L I	L Y S S Y I	S I S P Y Y S Y T Y	G A S F Y Y P W Y - H G F Y G A G G I
	P016	S Y F V S G S P I	I Y S S S I	S I Y P Y Y S S T Y	S Y Y Y Y Y Y A - - - Y H S W S G A M
	P017	H Y Y - - W A P I	I Y Y S S M	S I S S Y S G S T Y	W Y Y Y - - - - - - - - - - - V W A M
CG1434	P018	G Y V - - W Y L F	I Y S S S M	S I S S Y S S Y T S	P A P Y W Y Y - - - - - - G P F Y A L
	P019	W F A - - - S P I	I Y Y S Y I	S I S S Y Y G Y T Y	G F V - - - - - - - - - - - - - - A L
	P020	W W W - - - Y P I	I Y Y Y Y M	S I Y P S S G S T Y	S Y S W W S - - - - - - - Y S G H G L
	P021	Y G S F - Y S L I	L S Y Y S M	S I S S S Y G Y T Y	S G - - - - - - - - - - - - - - - A L
CG8273	P022	V S H - - Y W P F	L Y S Y S M	Y I Y S Y Y G Y T S	S A F G - - - - - - - - - - - G G A M
	P023	Y W Y - - - Y L F	I S S S S I	S I Y P Y Y S S T S	G Y Y W Y - - - - - - - - - - S A G L
	P024	S S Y - - - S L I	L Y S S S I	S I S S S Y G Y T S	G Y A A G Y S Y W - - S S A Y Y S G M
DCR-2	P025	W A A - - - V P I	L S Y Y S I	Y I S P Y Y G Y T Y	Y V W - - - - - - - - - - - - - G G M
DIP1	P026	P S G - - - S L F	L S Y S S M	S I S P Y S S Y T Y	Y S F P W F - - - - - - - Y Y G G A L
	P027	S Y - - - - G P F	L S Y S S M	Y I S P Y S G Y T Y	Y S Y P W F F G G - G W F G A P G A L
	P028	P F Y - - S S P I	I S Y S S M	S I Y S Y Y G Y T Y	P Y P Y A - - - - - - - - - - W G A F
	P029	A A S - - - G P I	L Y Y S S M	S I S P Y S G Y T Y	W G A P W F S V Y V W G G Y W G S G F
	P030	W H W G G G G L I	L S Y Y S M	Y I S P Y Y G Y T S	V G A A A - - - - - - - - - - W G G F
	P031	Y A - - - - Y L F	I S Y S Y M	S I Y P Y S G Y T S	G W G - - - - - - - - - - - - - - G L
	P032	Y A - - - - F L F	I S Y S Y M	S I Y P Y S S Y T S	G W G - - - - - - - - - - - - - - G L
	P033	Y W P - - Y Y P I	F S S S S I	Y I S P S S G Y T Y	G - - - - - - - - - - - - - - - - G M
	P034	Y W P - - W Y P I	F S S S S I	Y I S S Y S S S T S	G - - - - - - - - - - - - - - - - G M
	P035	Y W P - - F Y P I	I S Y S S I	Y I S P S Y G Y T S	A - - - - - - - - - - - - - - - - A F
DROSHA	P036	S Y S - - - Y L F	I S Y S Y I	S I S S Y Y G Y T S	A Y P Y Y - - - - - - - - - A Y V A M
	P037	G Y Y - - A S L I	I S Y S Y M	S I S P S Y G Y T Y	S V Y - - - - - - - - - - - - - F G L
	P038	W H S - - - V L I	I Y Y S Y M	S I S P S Y S Y T Y	S V F - - - - - - - - - - - - - S A I
LOQS	P039	Y W S - - Y A L I	I S Y Y S M	S I Y P S Y G Y T S	G F P F H F H - - - - - - W P A A G F
	P040	Y A G P S H S L I	L Y S S S I	S I S P Y S S Y T Y	G V Y G W Y - - - - - - - W F G S A M
PASHA	P041	Y W Y H W S W P I	L Y Y S S M	S I S S S Y G Y T Y	S G Y - - - - - - - - - - - - - - G L
R2D2	P042	Y Y Y Y V S Y P I	L Y S Y S I	S I S P Y Y G Y T Y	Y V G S - - - - - - - - - - - - Y G L
	P043	A Y Y Y Y Y Y P I	L S Y Y S I	S I S P Y Y G Y T S	Y S S S - - - - - - - - - - - A W G F
	P044	Y Y W Y - P P L F	L Y Y S S M	S I S S Y S G Y T Y	Y S S G - - - - - - - - - - - - S A L
	P045	Y W W G - G F P I	I S Y Y Y I	S I S S Y S G Y T S	P Y G F - - - - - - - - - - - H S A M
	P046	Y F W - - Y A P F	I Y Y Y S I	Y I S P S S G Y T Y	Y F A V - - - - - - - - - - - - Y G M
	P047	Y W F - - G S L I	I Y Y Y S M	S I Y P Y S S S T Y	Y Y Y V - - - - - - - - - - - - S A M
	P048	S S V W S Y G L F	I S Y Y S M	S I S P Y S G Y T Y	A P W - - - - - - - - - - - - - Y G M
	P049	Y H P S - F A L I	L S Y Y S I	Y I S P Y Y G Y T Y	S Y A H - - - - - - - - - - - - Y G L
	P050	F W A - - G S L I	L Y Y Y S M	Y I Y P S S G S T S	S P W - - - - - - - - - - - - - Y G L
	P051	Y S V - - F S L I	L Y Y S S M	S I S P Y S G Y T Y	G Y S Y - - - - - - - - - - - - F G L
	P052	Y Y F Y F Y W P F	I Y S Y S I	S I S P Y Y S Y T Y	Y Y S Y - - - - - - - - - - - - Y A L
DP1	P053	S S Y - - - S L I	F S S S S I	Y I Y S Y S G Y T Y	G V S P W Y H Y Y - W Y Y G S F V A L
IMP	P054	H W - - - - W L I	L S Y S Y I	S I Y S Y Y S Y T Y	W W V V - - - - - - - - - - - - Y G M
	P055	Y A - - - - F L I	L Y S Y S I	Y I Y S S S G Y T Y	G G G G W W S Y W - F S Y G G G S A I

FIGURE 5. Complementarity determining region (CDR) sequences of Fabs produced against antigens that represent computationally predicted structured regions lying outside of annotated domains. Positions randomized within each of the CDRs are shown at the *top* of each column, numbered according to IMGT standards (Lefranc et al. 2003). Each Fab is named with a unique identifier, and antigens are listed by the FlyBase gene symbol of the parent protein.

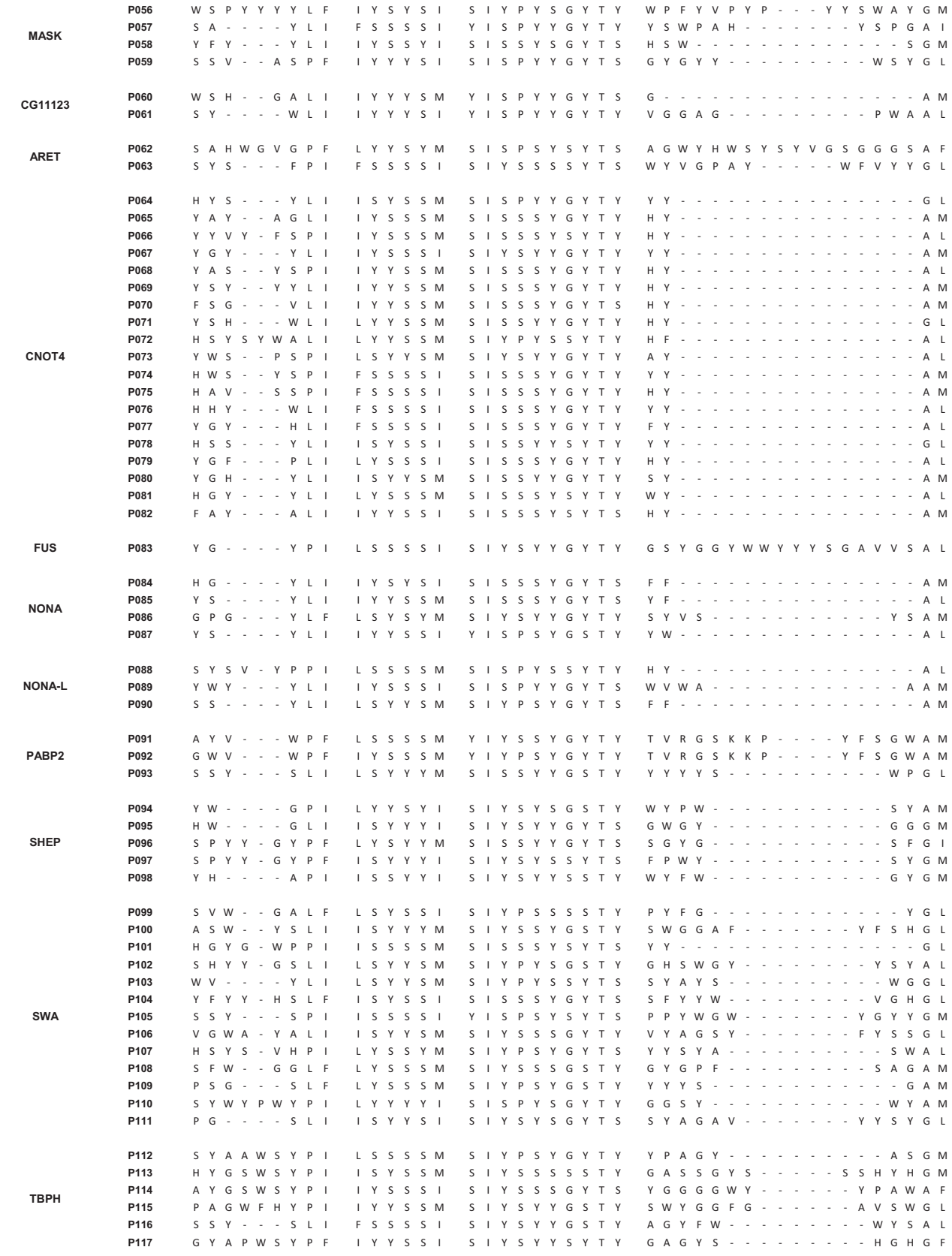

FIGURE 5. *Continued.*

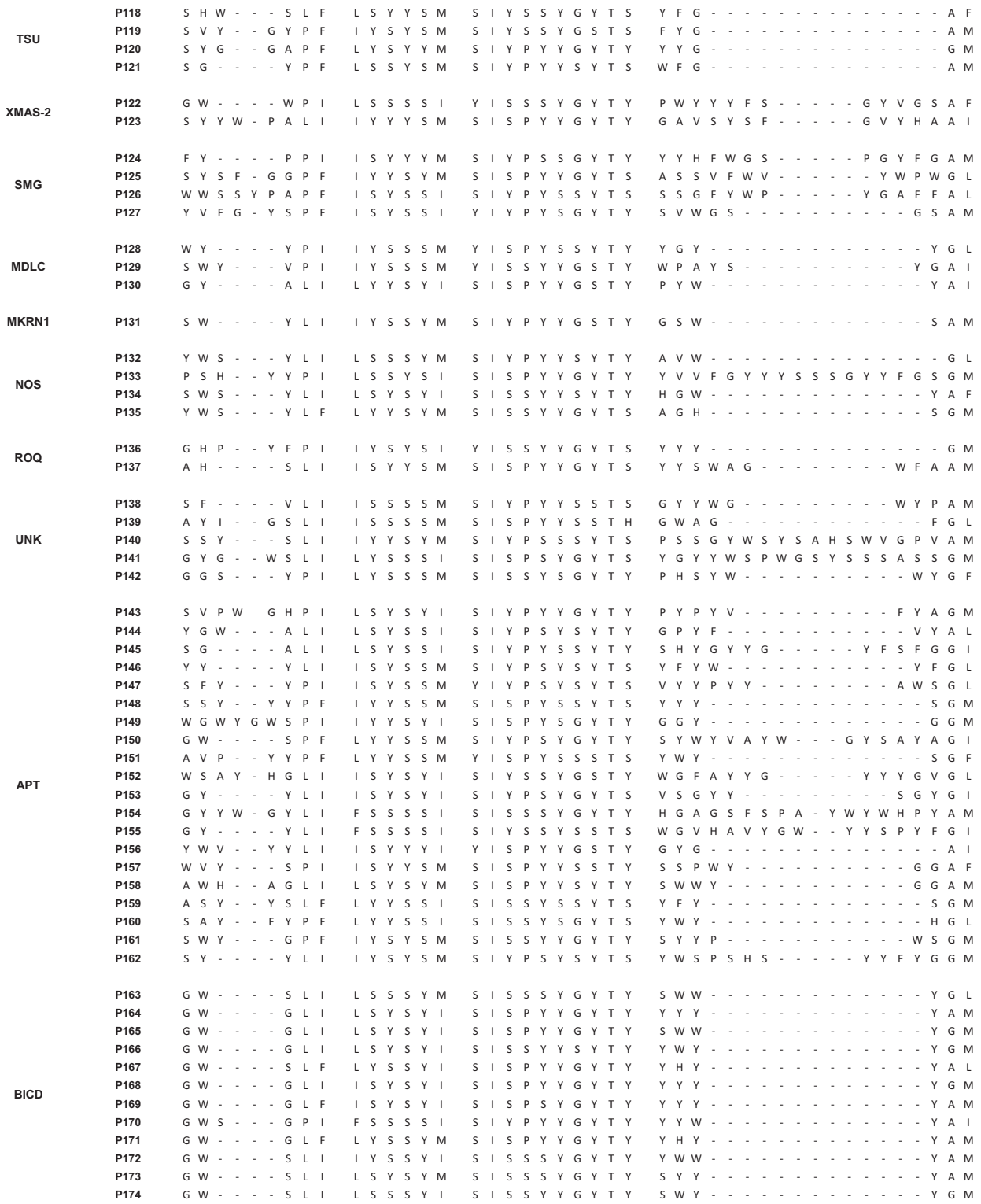

FIGURE 5. *Continued.*

target protein by probing Western blots of the IPs using conventional polyclonal or monoclonal antibodies raised against the same protein. In parallel, we used the same approach to test six additional synthetic Fabs that had been generated against three other RNP proteins (FMR1—two Fabs, HOW—one Fab, and ORB2—three Fabs) with low-throughput approaches prior to the establishment of the high-throughput pipeline.

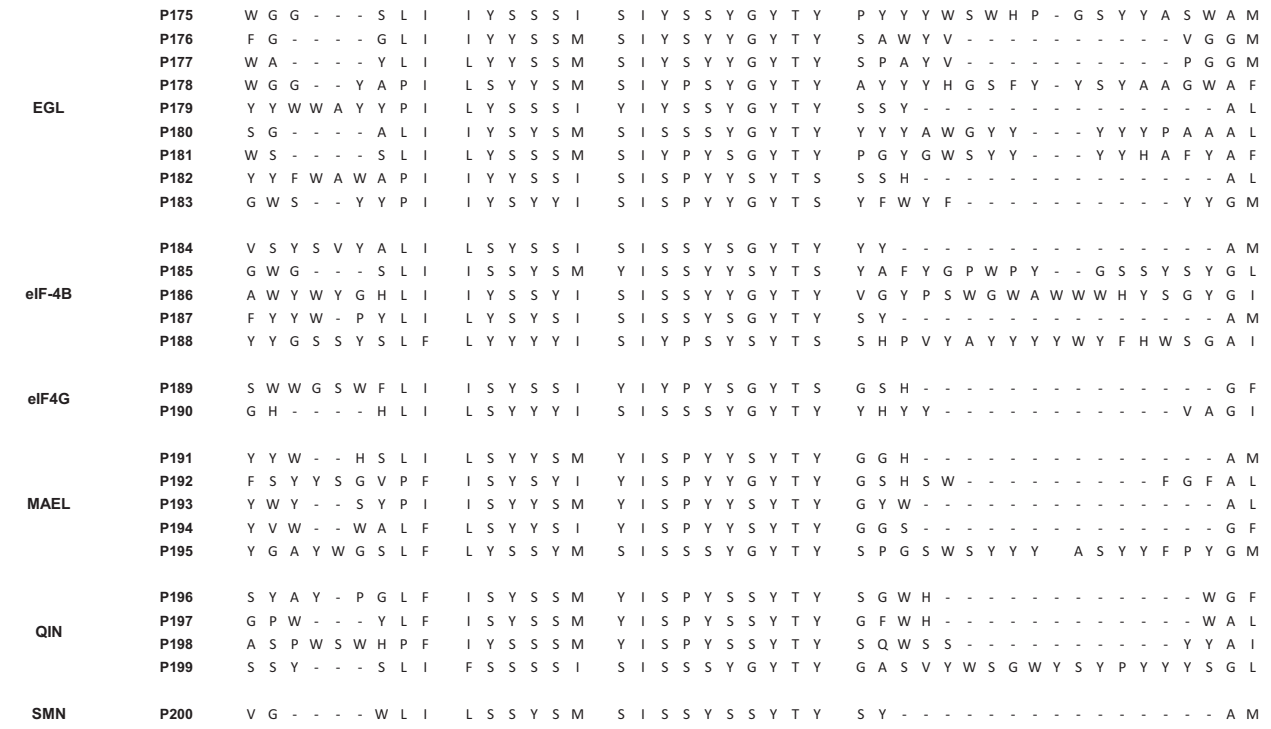

FIGURE 5. *Continued.*

In total, 12 of the 14 Fabs tested successfully immunoprecipitated their endogenous target proteins from *Drosophila* embryo extract, including all six of the Fabs produced by low-throughput approaches and six of the eight Fabs produced by our high-throughput pipeline (Fig. 8). Together with our published results on four additional Fabs produced by low-throughput methods—that recognize the RBPs STAU (two Fabs), BRAT (one Fab), and PUM (one Fab) (Laver et al. 2012, 2013, 2015)—a total of 16 out of 18 Fabs (89%) immunoprecipitated their endogenous target protein.

We note, however, that not all Fabs performed equally well in IP: Some immunoprecipitated their targets very efficiently (>10% of IP input was precipitated: anti-eIF4G, anti-PABP, anti-MEI-P26, anti-NOS, and anti-BRAT); others immunoprecipitated their targets with moderate efficiency (between 1% and 10% of IP input was precipitated: anti-IMP, anti-HOW, anti-ORB2 E7, anti-ORB2 E8, anti-ORB2 H4, anti-PUM, anti-STAU 2A5, and anti-STAU 2C5); and others immunoprecipitated their targets more weakly (<1% of IP input was precipitated: anti-ME31B, anti-FMR1 C7, and anti-FMR1 F12). Importantly, despite this range in IP efficiencies, we observed no bias between antibodies produced by high- versus low-throughput methods. In addition, we note that anti-STAU 2A5, one of the "moderate" Fabs, has been used successfully to identify STAU target mRNAs in RIP-Chip experiments (Laver et al. 2013). Moreover, when necessary, weaker Fabs can be affinity matured to improve their performance in IPs (Li et al. 2009; Huang et al. 2015) and/or can be converted to bivalent IgGs that exhibit enhanced effective affinities due to avidity effects.

In summary, our success in the design and expression of soluble antigens, many of which map outside of annotated domains (96%), and in obtaining one or more Fabs against a large majority of the antigens screened (66%) combined with the high success rate of our Fabs in IP of their of endogenous target proteins (89%), demonstrates the utility of our high-throughput pipeline for generating synthetic antibodies as tools for studies of RNPs.

DISCUSSION

Here, we have described a robust methodology for the high-throughput production of synthetic antibodies as a means to generate tools required for global studies of endogenous RNP complexes. Our high-throughput pipeline consists of three main steps: (i) computationally guided antigen design, (ii) high-throughput antigen expression and purification, and (iii) high-throughput antibody selection and screening. Using a panel of 93 antigens representing 90 RNP complex proteins from *D. melanogaster*, we have demonstrated that this pipeline can successfully produce antibodies against a majority of antigens screened, and that these antibodies can be used to immunoprecipitate their endogenous target proteins.

Designing antigens located outside of annotated domains

An important aspect of our pipeline is our strategy for designing antigens that are found outside of annotated domains. Given that independently folding, well-structured regions serve as the most effective antigens for synthetic

RBP name	Fab ID	CDR-L3									CDR-H1						CDR-H2										CDR-H3																		
		107	108	109	110	112	113	114	115	116	30	35	36	37	38	39	55	56	57	58	59	62	63	64	65	66	107	108	109	110	111	111.1	111.2	111.3	111.4	111.5	112.5	112.4	112.3	112.2	112.1	112	113	114	115
OSK	D001	G	G	-	-	-	-	Y	P	I	I	S	Y	Y	S	M	S	I	Y	P	Y	Y	G	S	T	Y	F	S	Y	S	G	-	-	-	-	-	-	-	-	-	H	Y	W	G	L
	D002	Y	G	Y	P	G	H	A	L	I	L	Y	Y	S	S	I	S	I	Y	P	S	Y	G	Y	T	S	S	Y	A	H	Y	F	S	Y	-	-	-	Y	G	Y	V	S	G	A	L
AGO2	D003	S	S	Y	-	-	-	S	L	I	F	S	S	S	S	I	Y	I	S	S	Y	S	G	Y	T	S	H	V	Y	G	Y	S	A	W	-	-	-	P	Y	G	Y	Y	S	A	M
CSTF-64	D004	W	G	-	-	-	-	S	L	F	L	Y	Y	Y	S	I	Y	I	S	P	Y	Y	G	Y	T	Y	F	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	M
	D005	A	P	S	Y	Y	F	S	P	I	I	S	Y	S	S	I	S	I	Y	S	Y	Y	G	S	T	S	A	W	S	-	-	-	-	-	-	-	-	-	-	-	-	-	W	G	L
	D006	W	S	-	-	-	-	Y	P	I	I	Y	Y	Y	S	M	S	I	S	P	Y	Y	G	Y	T	Y	G	Y	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	L
	D007	W	A	P	-	-	-	Y	P	I	L	Y	Y	Y	S	M	S	I	S	S	Y	Y	G	Y	T	Y	P	G	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	Y	G	F
GW	D008	Y	G	-	-	-	-	A	P	F	L	Y	Y	S	Y	I	S	I	S	S	S	Y	G	Y	T	Y	W	H	G	W	G	Y	G	W	S	-	A	S	A	H	Y	Y	G	G	F
	D009	S	H	H	-	-	-	Y	P	I	L	Y	Y	S	Y	I	S	I	Y	S	S	Y	G	Y	T	Y	S	V	W	W	H	W	-	-	-	-	-	-	-	-	Y	T	S	A	F
	D010	Y	W	-	-	-	-	G	P	F	I	Y	S	S	Y	M	S	I	Y	S	S	Y	G	Y	T	S	G	W	W	A	Y	S	G	S	-	-	-	G	W	P	S	Y	S	G	M
	D011	W	G	Y	Y	W	A	G	L	I	I	Y	S	S	Y	M	S	I	S	S	S	Y	G	Y	T	Y	S	F	V	Y	W	S	-	-	-	-	-	-	-	G	W	G	A	G	L
	D012	A	A	S	S	-	S	F	L	I	L	Y	Y	S	S	M	S	I	Y	S	Y	Y	G	Y	T	S	S	F	V	W	Y	-	-	-	-	-	-	-	-	-	-	G	F	G	L
	D013	S	S	Y	-	-	-	S	L	I	I	Y	Y	Y	Y	M	S	I	S	P	Y	S	G	Y	T	Y	S	S	A	V	Y	S	A	G	Y	-	W	F	S	A	F	Y	A	A	M
	D014	S	Y	F	G	-	P	H	P	I	F	S	S	S	S	I	S	I	Y	S	Y	S	G	Y	T	S	G	Y	W	G	F	S	G	H	V	-	-	Y	W	G	A	W	V	G	M
	D015	S	Y	F	G	-	P	H	P	I	F	S	S	S	S	I	S	I	Y	S	Y	S	G	Y	T	S	G	Y	W	G	F	S	G	H	-	-	-	V	Y	W	G	A	W	V	G
	D016	Y	G	Y	W	-	G	G	L	I	I	S	Y	Y	Y	I	Y	I	S	P	Y	S	S	Y	T	Y	S	V	W	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	M
	D017	S	S	Y	-	-	-	S	L	I	I	S	Y	Y	S	M	Y	I	S	P	Y	S	S	Y	T	Y	S	G	W	S	W	-	-	-	-	-	-	-	-	-	-	Y	S	G	M
	D018	W	H	-	-	-	-	S	L	I	I	Y	S	Y	S	M	S	I	Y	P	Y	S	G	Y	T	Y	G	W	W	Y	-	-	-	-	-	-	-	-	-	-	-	W	P	G	L
	D019	G	W	W	G	-	A	G	L	I	I	Y	S	Y	S	M	S	I	S	S	Y	S	S	S	T	Y	P	V	G	W	W	-	-	-	-	-	-	-	-	-	G	G	S	G	F
	D020	S	S	V	W	-	A	S	P	F	I	Y	S	Y	Y	M	S	I	S	P	S	Y	S	S	T	S	S	Y	H	Y	P	S	V	G	-	-	-	W	Y	V	Y	A	H	A	L
	D021	S	W	-	-	-	-	Y	L	F	L	Y	Y	Y	Y	M	S	I	Y	P	S	Y	G	Y	T	Y	S	P	W	Y	P	S	G	S	S	-	W	A	F	Y	G	S	F	A	F
PABP	D022	S	A	W	-	-	W	Y	P	I	L	S	Y	S	S	I	Y	I	S	S	Y	Y	G	Y	T	S	G	S	A	P	W	G	V	P	G		W	G	S	S	F	S	P	A	F
	D023	G	G	W	S	W	Y	A	P	I	L	S	Y	S	S	M	S	I	S	S	Y	Y	G	Y	T	Y	S	W	G	-	-	-	-	-	-	-	-	-	-	-	-	-	V	G	M
	D024	S	W	S	-	-	W	Y	P	I	I	S	S	Y	S	M	S	I	S	S	Y	S	S	Y	T	Y	S	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	M
	D025	S	Y	Y	-	-	A	H	L	F	I	Y	S	S	S	I	Y	I	Y	P	S	Y	G	Y	T	Y	Y	W	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	G	A	F
	D026	A	F	W	P	Y	S	W	P	I	I	Y	S	S	S	I	S	I	S	P	Y	S	G	Y	T	S	W	S	Y	H	-	-	-	-	-	-	-	-	-	-	-	G	A	G	M
	D027	G	G	H	-	-	-	F	L	I	I	S	S	S	S	M	S	I	S	P	Y	Y	G	Y	T	S	S	G	S	W	V	-	-	-	-	-	-	-	-	-	G	Y	G	G	M
	D028	S	Y	G	-	-	G	Y	L	I	I	S	Y	S	Y	I	S	I	Y	S	Y	Y	G	S	T	Y	A	Y	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	M
	D029	S	W	Y	-	-	A	V	L	F	I	S	Y	S	Y	M	S	I	Y	S	Y	Y	G	Y	T	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	M
	D030	S	S	P	-	-	Y	G	L	I	I	S	Y	S	Y	M	Y	I	Y	P	Y	S	G	Y	T	Y	G	P	W	A	S	W	Y	P	-	-	-	G	W	S	Y	H	P	A	F
	D031	S	Y		-	-		Y	P	F	I	S	Y	S	S	I	Y	I	Y	S	S	S	G	Y	T	Y	S	Y	V	G	W	F	G	S	P	A	W	S	G	S	Y	W	A	G	L
	D032	G	W	Y	-	-	S	Y	P	I	L	Y	S	S	S	I	S	I	S	P	Y	S	G	S	T	S	W	Y	G	H	A	-	-	-	-	-	-	-	-	-	Y	G	Y	A	L
	D033	S	W	S	-	-	W	Y	P	I	I	Y	S	Y	S	M	S	I	S	S	S	Y	G	Y	T	Y	F	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	M
	D034	G	P	G	G	S	S	S	L	I	I	Y	Y	Y	S	M	S	I	Y	S	S	Y	G	S	T	Y	G	A	W	Y	Y	Y	P	-	-	-	-	-	-	Y	S	P	G	W	G	I
	D035	S	W	-	-	-	-	Y	L	I	I	Y	Y	Y	S	M	S	I	Y	S	S	S	G	S	T	Y	P	Y	Y	W	H	S	W	-	-	-	-	-	-	G	W	S	G	S	G	M
	D036	S	W	S	-	-	W	Y	P	I	I	S	Y	Y	S	I	S	I	Y	S	Y	S	S	Y	T	Y	S	P	Y	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	M
eIF-2alpha	D037	S	Y	Y	-	-	-	H	P	I	I	Y	S	S	Y	I	Y	I	S	P	Y	Y	G	Y	T	Y	W	G	Y	P	G	S	G	Y	-	-	-	W	Y	A	W	G	Y	G	F
	D038	W	G	-	-	-	-	H	L	I	I	Y	Y	Y	S	M	S	I	Y	S	Y	Y	G	Y	T	Y	A	S	G	G	H	-	-	-	-	-	-	-	-	-	-	Y	P	G	F
MEI-P26	D039	S	A	-	-	-	-	W	L	F	I	S	Y	S	S	M	S	I	S	P	Y	Y	G	S	T	Y	A	S	W	S	G	-	-	-	-	-	-	-	-	-	-	V	Y	A	M
	D040	S	H	-	-	-	-	Y	L	F	I	S	Y	S	S	M	S	I	S	P	Y	Y	G	S	T	Y	A	S	Y	S	G	-	-	-	-	-	-	-	-	-	-	P	S	A	F
	D041	Y	Y	-	-	-	-	Y	L	F	I	S	Y	S	S	M	S	I	S	P	Y	Y	S	S	T	Y	A	S	Y	S	A	-	-	-	-	-	-	-	-	-	A	S	Y	G	M
	D042	V	W	-	-	-	-	G	P	I	I	Y	S	S	S	M	S	I	S	P	Y	Y	G	S	T	Y	P	Y	G	-	-	-	-	-	-	-	-	-	-	-	-	-	F	G	L
	D043	Y	S	H	-	-	-	S	P	I	I	Y	S	S	S	M	S	I	S	S	Y	Y	G	S	T	Y	H	W	Y	Y	G	-	-	-	-	-	-	-	-	-	G	V	V	G	M
	D044	S	Y	-	-	-	-	W	L	I	I	S	S	S	S	M	S	I	Y	P	Y	Y	S	Y	T	S	Y	P	W	G	S	-	-	-	-	-	-	-	-	-	-	P	Y	A	M
	D045	A	W	Y	Y	P	W	P	L	I	I	Y	Y	Y	S	M	S	I	Y	S	Y	Y	G	Y	T	Y	H	G	Y	P	-	-	-	-	-	-	-	-	-	-	-	G	S	A	M
	D046	Y	H	-	-	-	-	Y	L	F	L	H	S	S	S	M	S	I	S	S	S	Y	S	S	T	Y	A	W	Y	G	G	-	-	-	-	-	-	-	-	-	P	A	Y	G	F
ZN72D	D047	W	F	-	-	-	-	S	L	F	L	S	S	Y	S	I	Y	I	Y	P	Y	S	G	Y	T	Y	Y	S	W	W	Y	P	S	G	W		S	Y	Y	A	A	S	G	A	L
	D048	G	F	H	S	Y	W	S	L	F	L	Y	Y	S	Y	I	S	I	S	P	Y	Y	S	S	T	Y	S	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	V	W	G	I
CBP80	D049	Y	P	H	-	-	-	W	P	F	I	Y	Y	S	S	M	S	I	Y	S	Y	Y	G	S	T	Y	S	Y	W	F	P	A	-	-	-	-	-	-	-	S	Y	A	Y	A	L
	D050	S	S	Y	-	-	-	S	L	I	L	Y	S	S	S	M	Y	I	S	P	Y	Y	G	Y	T	S	H	H	S	V	W	Y	Y	S	G	Y	Y	Y	A	Y	G	P	V	A	M
	D051	G	S	Y	S	-	S	S	L	I	I	Y	Y	S	S	I	S	I	Y	P	Y	S	G	Y	T	Y	S	G	Y	G	-	-	-	-	-	-	-	-	-	-	-	-	V	A	I
	D052	S	W	G	W	G	G	W	L	I	L	S	S	Y	Y	I	S	I	S	P	Y	Y	S	Y	T	Y	F	V	S	-	-	-	-	-	-	-	-	-	-	-	-	-	W	A	L
	D053	Y	F	S	H	-	H	A	L	I	I	Y	S	Y	Y	I	S	I	S	S	Y	Y	G	Y	T	Y	A	P	G	W	S	F	A	V	G	S	V	S	Y	G	W	A	V	G	L
	D054	S	V	C	P	V	G	S	M	I	I	Y	S	Y	Y	I	S	I	S	S	Y	Y	G	Y	T	Y	S	P	G	W	S	F	C	G	G	S	V	S	Y	C	W	A	V	G	L
	D055	A	A	W	S	V	G	S	L	I	I	Y	Y	Y	Y	I	S	I	S	P	Y	S	G	S	T	Y	S	S	W	F	S	Y	-	-	-	-	-	-	-	W	W	G	W	A	F
DIS3	D056	Y	H	H	G	S	G	S	L	F	I	S	Y	Y	Y	I	S	I	S	S	S	Y	G	Y	T	Y	Y	S	Y	F	-	-	-	-	-	-	-	-	-	-	-	G	H	G	F

FIGURE 6. CDR sequences of Fabs produced against antigens that represent annotated non-RNA-binding domains. Positions randomized within each of the CDRs are shown at the *top* of each column, numbered according to IMGT standards (Lefranc et al. 2003). Each Fab is named with a unique identifier, and antigens are listed by the FlyBase gene symbol of the parent protein.

antibody selections (Hornsby et al. 2015; Marcon et al. 2015), annotated domains or full-length proteins have previously been the most commonly used antigens for synthetic antibody production (Schofield et al. 2007; Colwill and Graslund 2011; Hornsby et al. 2015; Huang et al. 2015). However, although this provides a straightforward approach to choosing antigens, for our purposes, we sought to exclude annotated RNA-binding domains as antigens in order to prevent the production of antibodies that would disrupt RBP–RNA interactions and, for 75% of the RNA-binding RNP complex proteins, were left with no other annotated domains from which to choose. Indeed, more generally, it often may be desirable to exclude particular domains as antigens in order to avoid production of antibodies that might disrupt specific protein functions and interactions.

The computationally guided methodology that we developed provides a powerful approach for designing structured antigens that lie outside of annotated domains. The success of this approach is highlighted both by the extremely high percentage of antigens (96%) for which we were able to obtain at least partially intact protein upon purification from *E. coli* and by the fact that we obtained antibodies against two-thirds of these antigens, with identical success rates for antigens derived from annotated domains and for those lying outside of annotated domains. Our success rate in obtaining antibodies is comparable with that of high-throughput selections performed against annotated SH3 domains (Huang et al. 2015).

In summary, our approach should be widely applicable for designing antigens for synthetic antibody selections for proteins where no annotated domains exist, or where particular annotated domains must be excluded as antigens, thus expanding the repertoire of potential synthetic antibody targets.

Validation and uses of antibodies generated by the high-throughput pipeline

When we began this project, we did not know how useful our Fabs would be for IP of endogenous RNP complex proteins. We have now shown that, of 18 low- and high-throughput generated Fabs, 89% immunoprecipitated their target proteins from early embryo extracts. This was determined using IP-westerns with available conventional antibodies against the same RNP complex proteins. The use of conventional antibodies for Western blots was necessitated as synthetic antibodies typically recognize native rather than denatured antigen and thus may not work well on westerns when detecting their endogenous target protein (Marcon et al. 2015).

To validate all synthetic antibodies generated by the high-throughput pipeline, IP followed by mass spectrometry (IP-MS) will be the tool of choice. First, conventional antibodies do not exist for most *Drosophila* RNP complex proteins (or, for that matter, for most proteins encoded by any organism's genome). Second, IP-MS is a quantitative method for determining the ability of synthetic antibodies to immunoprecipitate their target proteins, as has been confirmed recently for a panel of 1124 synthetic antibodies against 152 chromatin-associated proteins (Marcon et al. 2015). Finally, IP-MS will be useful, not just for validation, but also for identification of protein–protein interactions that occur in each RNP complex.

We also anticipate that our synthetic antibodies will be useful for immunofluorescence (IF) detection to determine the localization of their target proteins in tissues and cells. An analysis of 381 synthetic antibodies that recognize human proteins revealed that 37% worked for IF of the endogenous protein on tissue microarrays (Schofield et al. 2007). The abovementioned study of synthetic antibodies for chromatin-associated proteins found that 50 of 66 (76%) of the "gold-standard" antibodies were useful for IF of the endogenous protein in HEK293 cells (Marcon et al. 2015).

Global characterization of RNP components and functions

Our previous low-throughput and the currently reported high-throughput efforts for synthetic antibody production have generated antibodies against 67 *Drosophila* RNP complex

eIF4G2	D057	AYW---SLI	LSYSSM	SIYSYYGYTS	HA----------AF
	D058	HWW---YPI	IYYSYM	SISSYYGYTY	HS----------AM
	D059	WVY---YPI	FSSSSI	SISPSYGYTS	WWY----------GM
	D060	WVY---YPI	ISSSSI	SISSSYGYTY	WYY----------GF
	D061	WWW---YPI	LYSSYM	SIYPYYGYTY	VHG----------GGM
	D062	GSY--GSPF	IYSYSM	SISSSYGYTY	AHFS----------AYGM
eIF-5A	D063	SSY---SLI	LSYSYI	SIYPYSGYTY	YGWYPYGGY-FYSVSPYAF
HRG	D064	YG----GLI	LSYSSM	SIYPYYGSTY	YWYYP----------VWSGM
	D065	YVW---GLF	FSSSSI	YIYPYYGYTY	SSWYAYYAY-SAYSYAPAM
LOST	D066	SYY--WPLI	LYYSYI	SISPYYSYTY	SGWV----------SAL
	D067	SYY--WPLI	LYYSYI	SIFPFYSHTF	S----------GW
	D068	PHY--GPLF	LSSSSI	YISPSYSYTS	YW----------AM
MSL-3	D069	GFP---WLF	ISYYYM	SISPSYGSTS	VWFVGY-------GFVYGM
	D070	YHW--YSPI	ISSSYM	SIYSYYGSTY	SYA----------AL
	D071	YYWWSYSLF	ISYYSM	YISSYYGSTY	SWY----------AL

FIGURE 6. *Continued.*

CDR-L3

RBP name	Fab ID	107	108	109	110	112	113	114	115	116
eIF-4A	R001	W	W	-	-	-	-	G	P	I
eIF4AIII	R002	Y	H	F	-	-	-	S	L	I
	R003	A	W	V	S	S	S	Y	P	F
AGO1	R004	W	S	Y	S	-	F	Y	P	F
	R005	Y	Y	S	Y	-	S	P	P	I
	R006	Y	W	G	F	W	S	S	L	I
	R007	Y	S	Y	Y	-	S	Y	P	F
CBP20	R008	Y	Y	V	-	-	-	W	P	I

CDR-H1

RBP name	Fab ID	30	35	36	37	38	39
eIF-4A	R001	I	Y	Y	S	Y	I
eIF4AIII	R002	L	S	S	Y	S	I
	R003	I	S	Y	Y	Y	M
AGO1	R004	L	S	S	S	Y	M
	R005	L	Y	Y	S	Y	M
	R006	L	S	Y	S	S	I
	R007	I	Y	Y	S	Y	I
CBP20	R008	I	S	Y	Y	Y	M

CDR-H2

RBP name	Fab ID	55	56	57	58	59	62	63	64	65	66
eIF-4A	R001	S	I	Y	P	Y	S	G	S	T	Y
eIF4AIII	R002	S	I	Y	P	Y	Y	G	Y	T	S
	R003	S	I	S	P	Y	Y	G	Y	T	S
AGO1	R004	S	I	Y	P	S	S	S	Y	T	Y
	R005	S	I	S	P	Y	Y	G	Y	T	S
	R006	S	I	S	S	S	S	S	Y	T	Y
	R007	S	I	Y	P	S	S	G	Y	T	Y
CBP20	R008	S	I	Y	S	Y	Y	G	Y	T	Y

CDR-H3

RBP name	Fab ID	107	108	109	110	111	111.1	111.2	111.3	111.4	111.5	112.5	112.4	112.3	112.2	112.1	112	113	114	115
eIF-4A	R001	W	P	G	W	A	S	Y	W	G	-	S	W	H	W	P	W	S	A	L
eIF4AIII	R002	F	S	V	S	A	Y	A	V	A	-	Y	S	F	G	W	Y	F	A	M
	R003	S	Y	G	Y	Y	Y	-	-	-	-	-	-	-	-	S	Y	H	A	L
AGO1	R004	Y	G	S	G	-	-	-	-	-	-	-	-	-	-	-	Y	Y	A	M
	R005	S	P	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	L
	R006	G	P	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	L
	R007	Y	G	Y	S	-	-	-	-	-	-	-	-	-	-	-	Y	W	A	M
CBP20	R008	S	F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	F

FIGURE 7. CDR sequences of Fabs produced against antigens that represent RNA-binding domains. Positions randomized within each of the CDRs are shown at the *top* of each column, numbered according to IMGT standards (Lefranc et al. 2003). Each Fab is named with a unique identifier, and antigens are listed by the FlyBase gene symbol of the parent protein.

proteins, most of which are known or predicted to be expressed in early embryos. Early embryos provide a particularly attractive system for global studies of RNPs, as they are an established model for studies of post-transcriptional regulation, and there are a variety of genome-wide descriptions of mRNA behavior available that can facilitate analysis of such data. These include descriptions of mRNA stability, translation, and subcellular localization (De Renzis et al. 2007; Lecuyer et al. 2007; Qin et al. 2007; Tadros et al. 2007; Thomsen et al. 2010; Dunn et al. 2013; Chen et al. 2014), which allow one to infer how RNP complexes function in post-transcriptional control based simply on the behavior of their component mRNAs (for examples, see Laver et al. 2013, 2015).

Extrapolating from the results obtained from the high-throughput pipeline reported here, it should be feasible to produce antibodies against two-thirds of *Drosophila* RNP complex proteins or, indeed, of the RNP complex proteins from any organism, in a single high-throughput screen. This fraction could be significantly increased by the screening of more than one antigen per protein. Thus, high-throughput production of synthetic antibodies using this pipeline could provide, in a rapid and cost-effective manner, the tools required to gain a global view of the composition of endogenous RNP complexes. Moreover, the pipeline could also be effectively applied to generate a large-scale set of synthetic antibodies against any collection of proteins from any organism.

MATERIALS AND METHODS

Computationally guided antigen design

To aid in the selection of structured protein regions as antigens, we developed an algorithm that compiles predicted structural information based on protein sequences, using the following tools: PSIPRED to predict regions of secondary structure (Buchan et al. 2010), DISOPRED2 to predict regions of disorder (Ward et al. 2004), InterProScan to find known annotated protein domains (Zdobnov and Apweiler 2001), and Antigenic and Hmoment from the EMBOSS package to predict antigenicity and hydrophobicity (Rice et al. 2000). In addition, we used Rate4Site (Mayrose et al. 2004) to calculate evolutionary conservation among *Drosophila* orthologs from 12 species. To predict disorder and secondary structure, homologous sequences were selected from Swiss-Prot/TrEMBL: sequences whose length was 0.7–1.4 times the query sequence length and had <90% similarity to other sequences were considered. For the estimation of evolutionary conservation, orthologous sequences from 12 *Drosophila* species were downloaded from the FlyBase database (http://flybase.org). All selected homologous and orthologous sequences were aligned using the MUSCLE sequence alignment tool (Edgar 2004). All tools that were integrated in the algorithm were applied with default options.

After compiling this information, the algorithm assigned scores that represent the following structural properties: secondary structure (0: secondary structure absent, 1: secondary structure present); residue-specific evolutionary conservation (0–1 in steps of size 0.2; 0 and 1 indicate low and high conservation, respectively); disorder (0: disordered, 1: ordered residues); antigenicity (0: nonantigenic site, 1: antigenic site); and hydrophobicity (0: < top 50%, 0.5: top 50% ≤ hydrophobic score ≤ top 75%, 1: ≥ top 75% of hydrophobic scores). Known RNA-binding domains were excluded from further analysis.

Using these scores, we selected regions that have a high predicted propensity for forming secondary structures, are conserved among *Drosophila* orthologs, have high predicted antigenicity, have low predicted hydrophobicity, and/or have low predicted disorder. To optimize the boundaries of our antigens, boundaries were chosen to correspond precisely to predicted regions of secondary structure whenever possible or, for antigens where secondary structure was not predicted, boundaries were based on regions of conservation. Only regions predicted to be present in all protein isoforms of a given RNP complex protein (i.e., present in all splice variants of the mRNA) were selected.

High-throughput antigen expression and purification

DNA encoding each of the 93 RNP complex antigens was synthesized and cloned into an IPTG-inducible expression vector with amino-terminal hexaHis- and GST-tags (Huang et al. 2015). Plasmids were transformed into *E. coli* BL21 cells, and high-throughput antigen expression and purification was carried out as previously described (Huang and Sidhu 2011) with minor modifications as follows: 10 μL of glycerol stocks of BL21 cells harboring each antigen expression plasmid was inoculated into individual wells of a

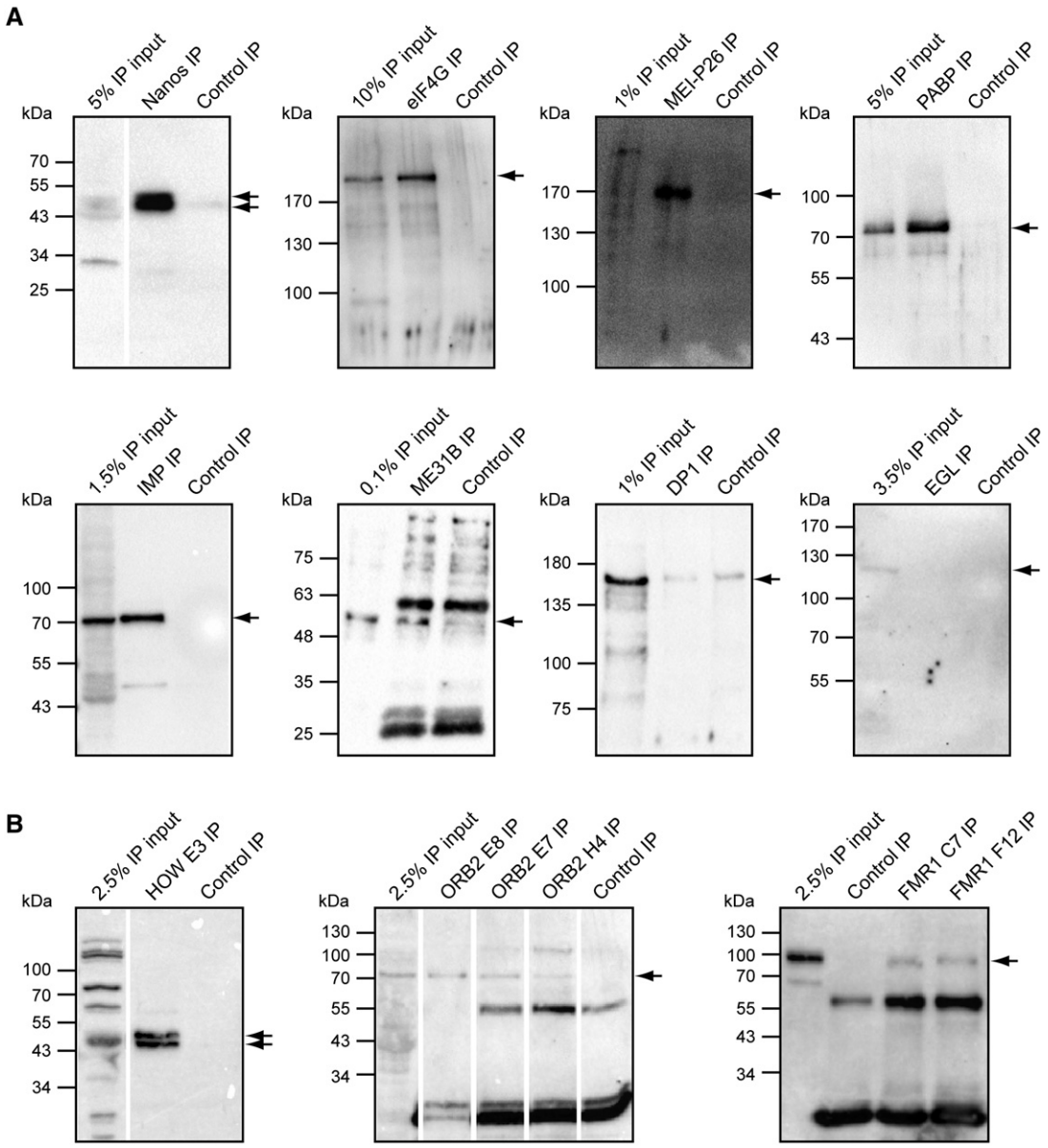

FIGURE 8. Validation of synthetic antibodies in the immunoprecipitation of endogenous target proteins. (*A*) One synthetic Fab against each of eight different RNP complex proteins, produced by the high-throughput pipeline, and (*B*) synthetic Fabs produced against HOW, FMR1, and ORB2 by low-throughput approaches, were used to carry out IPs from *Drosophila* embryo extract collected 0- to 3-h post-egg-laying. The success of the IPs was assessed by Western blots probed with conventional polyclonal or monoclonal antibodies against the same proteins. The percentage of the IP input loaded is indicated, and comparison of the IP input with the amount of protein present in the IP provides an estimate of the efficiency of the IP for each antibody. In all cases, negative control IPs were carried out with the control Fab C1 (Laver et al. 2012). White space *between* lanes indicates where lanes from the same blot were reordered.

96-well Mini tube system 0.6-mL (Axygen) containing 400 µL of 2YT media supplemented with carbenicillin (100 µg/mL) and grown overnight at 37°C with shaking at 200 rpm. The next day, two 96 deep-well plates (Whatman) containing 1.2 mL per well of MagicMedia (Invitrogen) supplemented with carbenicillin (100 µg/mL) were inoculated with 50 µL of overnight culture, such that each overnight antigen culture was added into the corresponding wells of each of the two deep-well plates. Cells were grown at 37°C for 6 h with shaking at 200 rpm, until OD 600 nm was ~0.6–0.9. The culture was then incubated at 18°C for 22 h with shaking at 200 rpm. The cells were harvested by centrifuging one of the plates at 4000 rpm for 10 min at 4°C, decanting the supernatant, transferring the corresponding culture from the duplicate plate to the first plate, and centrifuging again at 4000 rpm for 10 min at 4°C. Pellets were stored at −20°C overnight.

The following day, pellets were thawed and lysed by the addition of 1 mL of freshly prepared lysis buffer to each well, followed by shaking at 200 rpm for 40 min at room temperature. One hundred milliliters of lysis buffer were prepared by mixing 98 mL binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole,

5% glycerol), 1 mL Triton X-100 detergent, two protease inhibitor cocktail tablets (Roche), 16.5 µL Benzonase (EMD Millipore), and 100 mg Lysozyme (Sigma); for purifications where sarkosyl (*N*-lauroylsarcosine sodium salt; Sigma) was included, it was added to a final concentration of 1%. Seven milliliters of Ni-NTA agarose beads were prewashed three times by adding 15 mL of prewash buffer (binding buffer with 1% Triton X-100), spinning at 500 rpm for 5 min, and decanting supernatant. Seventeen milliliters of prewash buffer were added to the resin after the last wash. One hundred and fifty microliters of the prewashed Ni-NTA agarose beads were added to each well of a filter-bottom microplate (Seahorse Bioscience) sealed with four layers of Parafilm. After centrifuging the lysis plate at 2000*g* for 10 min at 4°C, 850 µL of supernatant was transferred from the lysis plate to corresponding wells in the filter-bottom microplate, followed by shaking the filter-bottom microplate at 170 rpm for 1 h at room temperature. Subsequently, Parafilm was removed from the filter-bottom microplate and the beads in each well were washed three times with 1 mL of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM imidazole, 5% glycerol). A multiwell vacuum manifold (Pall Corporation) was used to remove flow through and wash buffer. Subsequently, 100 µL of elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole, 5% glycerol) was added to each well, and the plate was incubated at room temperature for 5 min. A 96-well PCR plate was placed under the filter plate and the protein was eluted by centrifuging at 500 rpm for 5 min at 4°C.

Protein concentrations were determined by Bradford assay using bovine serum albumin (BSA) as a standard. The yield of protein from 2.4 mL of MagicMedia usually ranged from 5 µg to 100 µg in 100 µL eluent. For SDS-PAGE analysis, 5 µL of each purified antigen sample was run, and the gels were stained with InstantBlue (Expedeon Inc.).

High-throughput Fab-phage selections against RNP complex antigens

Four rounds of binding selection with phage-displayed Fab Library F (Persson et al. 2013) were performed to select for binders against the 93 antigens, in a high-throughput, 96-well format, as previously described (Huang and Sidhu 2011; Huang et al. 2015), with the following minor modifications. Four 96-well Maxisorp immunoplates (NUNC) were coated with antigens that were purified in the presence (two plates) or absence (two plates) of sarkosyl as capture targets for each round, by adding to each well, 90 µL of PBS followed by 10 µL of purified antigen (one antigen per well). In parallel, 96-well Maxisorp immunoplates were coated with 100 µL of GST (10 µg/mL) per well for negative selection at each round to remove GST-binding and nonspecific Fab-phage from the phage pool prior to incubation with the antigen-coated plates.

To determine whether reusing a library might be a cost-effective method to increase the coverage of library diversity for each antigen, in the first round of selection, naive Library F was applied to two antigen-coated plates—one coated with sarkosyl-purified antigens and one coated with antigens purified in the absence of sarkosyl—for selection of specific antigen binders. After incubation with these plates, the used library aliquots from each plate were collected, pooled, and used again by application to the remaining two antigen-coated plates (again one coated with sarkosyl-purified antigens and one coated with antigens purified in the absence of sarkosyl).

A total of four rounds of selections was performed for each of the different selection conditions (antigens purified with or without sarkosyl, and selections performed with naive or reused library for each), after which the pH-adjusted phage supernatants from the fourth round for each selection condition were used in pool-phage ELISAs to monitor the efficacy of the binding selections in the enrichment of specific binding clones, as previously described (Huang and Sidhu 2011).

Subcloning of enriched Fab-phage pools into an IPTG-inducible Fab expression vector

To subclone the entire Fab-phage pool after selection against each antigen, from phagemids into an IPTG-inducible expression vector, the phage supernatant from the fourth round of binding selection (either from an individual selection condition or combined pools from different selection conditions, as described in the Results) was used as a template for PCR to amplify from the pool the Fab-encoding DNA, using the primers HN4MF (5′-GCGGCCCATGCATCC ATGGCATCCGATATCCAGATGACCCAGTCCCC) and HN4MR (5′-GCGGCCCTCTAGATGTGTGAGTTTTGTCACAAGATTT GGG). The amplified DNA harboring the Fab-encoding regions was purified using a 96-well PCR purification kit (QIAGEN), digested with NcoI and XbaI restriction enzymes, gel-purified using a gel extraction kit (QIAGEN), and ligated into an IPTG-inducible Fab expression vector (RH2.2) digested with the same enzymes. The ligation reactions were transformed into *E. coli* DH5α competent cells and the transformation for each antigen was inoculated into individual wells of a 96 deep-well plate (Whatman) containing 800 µL LB medium supplemented with carbenicillin (100 µg/mL) per well, and cells were grown at 37°C overnight with shaking at 200 rpm. The next day, DNA isolation was carried out using a QIAGEN 96-plasmid DNA purification kit and DNA was eluted in 30 µL USP H_2O per well. These Fab expression plasmid pools were transformed into BL21 cells, which were plated to produce single clones. All of the above steps, except gel purifications, were performed in 96-well high-throughput format.

Screening and identification of unique antigen-binding Fabs

After subcloning the Fab pools resulting from Fab-phage selections for each antigen into expression vectors and transforming into *E. coli* BL21 cells, 24–48 colonies for each antigen were randomly picked and Fabs were screened by ELISA for binding to their cognate antigen. Individual colonies were grown in 96 deep-well plates containing 800 µL MagicMedia (Invitrogen) supplemented with carbenicillin (100 µg/mL) per well at 37°C for 24 h. One hundred microliters of culture for each clone were then used to make a glycerol stock by adding glycerol to a 20% final concentration; the remaining cells were pelleted by centrifuging at 4000 rpm for 20 min and pellets were stored at −20°C overnight.

To test each Fab clone for binding to its cognate antigen by ELISA, crude preparations of Fab proteins were produced from each clone by thawing the pelleted cells and lysing as follows: 150 µL of freshly prepared single-colony ELISA lysis buffer was added to each well (for 100 mL of lysis buffer: 98.6 mL of PBS, 1 mL Triton X-100, 1.5 µL of benzonase nuclease [250 U/µL], 200 µL of 1 M $MgCl_2$, 100 µL of 200 mM PMSF, 100 mg of lysozyme), and plates were

incubated with shaking at 70 rpm for 1 h at 4°C. Lysate was cleared by centrifugation at 4000 rpm for 20 min at 4°C, and supernatant, which contained Fab protein, was used to perform ELISAs comparing Fab binding to antigen- and GST-coated wells.

To perform ELISAs for each Fab, two wells of a 384-well Maxisorp plate (NUNC) were coated, one with 30 µL of antigen (2–5 µg/mL) and one with 30 µL of GST (5 µg/mL), overnight at 4°C with shaking, followed by blocking with 0.5% BSA in PBS for 1 h at room temperature, and washing four times with PBS supplemented with 0.05% Tween (PBS-T). Thirty microliters of crude Fab supernatant diluted 1:5 in PBS were added to each well; plates were incubated for 30 min at room temperature and washed eight times with PBS-T. To detect Fab binding via the Flag tag fused to the Fabs, 30 µL of HRP-conjugated anti-Flag antibody (1:5000 in cold PBS-T) was added to each well and incubated for 30 min at room temperature. Following eight washes with PBS-T, bound antibody was detected by adding 30 µL 3,3,5,5′ tetramethylbenzidine (TMB) peroxidase substrate (Kierkegaard and Perry Labs, Inc.) to each well, developing for 5 min, stopping the reaction by adding 30 µL of 1 M H_3PO_4, and measuring absorbance at 450 nm. Fabs exhibiting at least fivefold greater signal for the antigen-coated versus GST-coated wells were defined as positive antigen-binding clones.

The sequences of unique antigen-binding Fabs were determined by PCR amplifying and sequencing the encoding DNA of all positive antigen-binding clones for each antigen, using the glycerol stocks of BL21 cells containing each clone, which were prepared as described above, as template for the PCR reactions.

Immunoprecipitations and Western blots

To test a subset of Fabs generated by the high-throughput pipeline in IPs, Fabs were purified as previously described (Laver et al. 2012), and IPs for Western blots were carried out using the purified Fabs by pulling down the Fabs via their Flag-tag using anti-Flag-conjugated agarose beads (Sigma), as previously described (Laver et al. 2015). The anti-RBP Fabs produced by the high-throughput pipeline that were used in the IPs were anti-NOS (P134), anti-IMP (P054), anti-MEI-P26 (D046), anti-eIF4G (P190), anti-EGL (P177), anti-ME31B (P012), anti-PABP (D035), and anti-DP1 (P053) (see Figs. 5, 6 for Fab nomenclature). For the negative control, IPs were carried out with the control Fab C1 (Laver et al. 2012). To determine the efficacy of the IPs, IP samples were run on Western blots, which were probed with conventional polyclonal or monoclonal antibodies that had been previously produced against the RNP complex proteins of interest, as follows: rabbit anti-NOS (Hanyu-Nakamura et al. 2008), rabbit anti-IMP (Geng and Macdonald 2006), rabbit anti-MEI-P26 (Liu et al. 2009), rabbit anti-eIF4G (Zapata et al. 1994), rabbit anti-EGL (Mach and Lehmann 1997), rabbit anti-ME31B (Nakamura et al. 2001), rabbit anti-PABP (Zekri et al. 2013), guinea pig anti-DP1 (Nelson et al. 2007), rabbit anti-HOW (provided by Talila Volk), mouse anti-FMR1 5A11 (Okamura et al. 2004), and mouse anti-ORB2 4G8 (Hafer et al. 2011); the anti-FMR1 and anti-ORB2 antibodies were obtained from the Developmental Studies Hybridoma Bank at The University of Iowa.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank the following for providing the conventional antibodies used for our IP-westerns: Paul Macdonald (anti-IMP), Ruth Lehmann (anti-EGL), Paul Lasko (anti-MEI-P26), José Sierra (anti-eIF4G), Elisa Izaurralde (anti-PABP), Talila Volk (anti-HOW), and Akira Nakamura (anti-NOS and anti-ME31B). During the course of this research, we made extensive use of FlyBase and data from the Berkeley *Drosophila* Genome Project. This research was supported by grants to S.S.S. (Canadian Institutes of Health Research Operating Grant MOP-93725), to H.D.L. (Canadian Institutes of Health Research Operating Grant MOP-14409), and to C.A.S. (Natural Sciences and Engineering Research Council Discovery Grant RGPIN-435985). J.D.L. was supported in part by an Ontario Graduate Scholarship; J.D.L., K.A., W.X.C., K.N., and Z.Y. in part by University of Toronto Open Fellowships.

Received November 6, 2015; accepted December 12, 2015.

REFERENCES

Adams JJ, Sidhu SS. 2014. Synthetic antibody technologies. *Curr Opin Struct Biol* **24:** 1–9.

Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT. 2010. Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38:** W563–W568.

Chen L, Dumelie JG, Li X, Cheng MH, Yang Z, Laver JD, Siddiqui NU, Westwood JT, Morris Q, Lipshitz HD, et al. 2014. Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. *Genome Biol* **15:** R4.

Colwill K, Graslund S. 2011. A roadmap to generate renewable protein binders to the human proteome. *Nat Methods* **8:** 551–558.

De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. 2007. Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biol* **5:** e117.

Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* **2:** e01179.

Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J. 2004. Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol* **4:** 32.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Frankel S, Sohn R, Leinwand L. 1991. The use of sarkosyl in generating soluble protein after bacterial expression. *Proc Natl Acad Sci* **88:** 1192–1196.

Geng C, Macdonald PM. 2006. Imp associates with squid and Hrp48 and contributes to localized expression of *gurken* in the oocyte. *Mol Cell Biol* **26:** 9508–9516.

Gerber AP, Herschlag D, Brown PO. 2004. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* **2:** E79.

Hafer N, Xu S, Bhat KM, Schedl P. 2011. The *Drosophila* CPEB protein Orb2 has a novel expression pattern and is important for asymmetric cell division and nervous system function. *Genetics* **189:** 907–921.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.

Hanyu-Nakamura K, Sonobe-Nojima H, Tanigawa A, Lasko P, Nakamura A. 2008. *Drosophila* Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* **451:** 730–733.

Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. 2008. Diverse RNA-binding proteins interact with functionally related

sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **6:** e255.

Hornsby M, Paduch M, Miersch S, Saaf A, Matsuguchi T, Lee B, Wypisniak K, Doak A, King D, Usatyuk S, et al. 2015. A high through-put platform for recombinant antibodies to folded proteins. *Mol Cell Proteomics* **14:** 2833–2847.

Huang H, Sidhu SS. 2011. Studying binding specificities of peptide recognition modules by high-throughput phage display selections. *Methods Mol Biol* **781:** 87–97.

Huang H, Economopoulos NO, Liu BA, Uetrecht A, Gu J, Jarvik N, Nadeem V, Pawson T, Moffat J, Miersch S, et al. 2015. Selection of recombinant anti-SH3 domain antibodies by high-throughput phage display. *Protein Sci* **24:** 1890–1900.

Keene JD. 2007. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8:** 533–543.

Keene JD, Komisarow JM, Friedersdorf MB. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **1:** 302–307.

Laver JD, Ancevicius K, Sollazzo P, Westwood JT, Sidhu SS, Lipshitz HD, Smibert CA. 2012. Synthetic antibodies as tools to probe RNA-binding protein function. *Mol Biosyst* **8:** 1650–1657.

Laver JD, Li X, Ancevicius K, Westwood JT, Smibert CA, Morris QD, Lipshitz HD. 2013. Genome-wide analysis of Staufen-associated mRNAs identifies secondary structures that confer target specificity. *Nucleic Acids Res* **41:** 9438–9460.

Laver JD, Li X, Ray D, Cook KB, Hahn NA, Nabeel-Shah S, Kekis M, Luo H, Marsolais AJ, Fung KY, et al. 2015. Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the *Drosophila* maternal-to-zygotic transition. *Genome Biol* **16:** 94.

Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM. 2007. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131:** 174–187.

Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G. 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* **27:** 55–77.

Li B, Xi H, Diehl L, Lee WP, Sturgeon L, Chinn J, Deforge L, Kelley RF, Wiesmann C, van Lookeren Campagne M, et al. 2009. Improving therapeutic efficacy of a complement receptor by structure-based affinity maturation. *J Biol Chem* **284:** 35605–35611.

Liu N, Han H, Lasko P. 2009. Vasa promotes *Drosophila* germline stem cell differentiation by activating *mei-P26* translation by directly interacting with a (U)-rich motif in its 3′ UTR. *Genes Dev* **23:** 2742–2752.

Mach JM, Lehmann R. 1997. An Egalitarian–BicaudalD complex is essential for oocyte specification and axis determination in *Drosophila*. *Genes Dev* **11:** 423–435.

Marcon E, Jain H, Bhattacharya A, Guo H, Phanse S, Pu S, Byram G, Collins BC, Dowdell E, Fenner M, et al. 2015. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat Methods* **12:** 725–731.

Mayrose I, Graur D, Ben-Tal N, Pupko T. 2004. Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. *Mol Biol Evol* **21:** 1781–1791.

Morris AR, Mukherjee N, Keene JD. 2010. Systematic analysis of post-transcriptional gene expression. *Wiley Interdiscip Rev Syst Biol Med* **2:** 162–180.

Nakamura A, Amikura R, Hanyu K, Kobayashi S. 2001. Me31B silences translation of oocyte-localizing RNAs through the formation of cytoplasmic RNP complex during *Drosophila* oogenesis. *Development* **128:** 3233–3242.

Nelson MR, Luo H, Vari HK, Cox BJ, Simmonds AJ, Krause HM, Lipshitz HD, Smibert CA. 2007. A multiprotein complex that mediates translational enhancement in *Drosophila*. *J Biol Chem* **282:** 34031–34038.

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18:** 1655–1666.

Persson H, Ye W, Wernimont A, Adams JJ, Koide A, Koide S, Lam R, Sidhu SS. 2013. CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. *J Mol Biol* **425:** 803–811.

Qin X, Ahn S, Speed TP, Rubin GM. 2007. Global analyses of mRNA translational control during early *Drosophila* embryogenesis. *Genome Biol* **8:** R63.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16:** 276–277.

Schofield DJ, Pope AR, Clementel V, Buckell J, Chapple S, Clarke KF, Conquer JS, Crofts AM, Crowther SR, Dyson MR, et al. 2007. Application of phage display to high throughput antibody generation and characterization. *Genome Biol* **8:** R254.

Sidhu SS. 2012. Antibodies for all: the case for genome-wide affinity reagents. *FEBS Lett* **586:** 2778–2779.

Spitzer J, Hafner M, Landthaler M, Ascano M, Farazi T, Wardle G, Nusbaum J, Khorshid M, Burger L, Zavolan M, et al. 2014. PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods Enzymol* **539:** 113–161.

Stoiber MH, Olson S, May GE, Duff MO, Manent J, Obar R, Guruharsha K, Artavanis-Tsakonas S, Brown JB, Graveley BR, et al. 2015. Extensive cross-regulation of post-transcriptional regulatory networks in *Drosophila*. *Genome Res* **25:** 1692–1702.

Tadros W, Goldman AL, Babak T, Menzies F, Vardy L, Orr-Weaver T, Hughes TR, Westwood JT, Smibert CA, Lipshitz HD. 2007. SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev Cell* **12:** 143–155.

Tenenbaum SA, Carson CC, Lager PJ, Keene JD. 2000. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci* **97:** 14085–14090.

Tenenbaum SA, Lager PJ, Carson CC, Keene JD. 2002. Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* **26:** 191–198.

Thomsen S, Anders S, Janga SC, Huber W, Alonso CR. 2010. Genome-wide analysis of mRNA decay patterns during early *Drosophila* development. *Genome Biol* **11:** R93.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302:** 1212–1215.

Ule J, Jensen K, Mele A, Darnell RB. 2005. CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods* **37:** 376–386.

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337:** 635–645.

Zapata JM, Martinez MA, Sierra JM. 1994. Purification and characterization of eukaryotic polypeptide chain initiation factor 4F from *Drosophila melanogaster* embryos. *J Biol Chem* **269:** 18047–18052.

Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17:** 847–848.

Zekri L, Kuzuoglu-Ozturk D, Izaurralde E. 2013. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *EMBO J* **32:** 1052–1065.