2016

# Nanobody-Based Interactomic Studies of Single Transcripts During mRNA Maturation

Peter C. Fridy


# NANOBODY-BASED INTERACTOMIC STUDIES OF SINGLE TRANSCRIPTS DURING MRNA MATURATION

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Peter C. Fridy

June 2016

# NANOBODY-BASED INTERACTOMIC STUDIES OF SINGLE TRANSCRIPTS DURING MRNA MATURATION

Peter C. Fridy, Ph.D.

The Rockefeller University 2016

During and after transcription in the nucleus, messenger RNAs (mRNAs) undergo a variety of processing events before being exported to the cytoplasm through the nuclear pore complex. mRNA processing and nuclear export require a wide range of protein factors, which interact with maturing transcripts and each other to form dynamic mRNP complexes. While there are many core, essential mRNP factors, the pathways governing mRNA maturation are not uniform, and different transcripts can be associated with mRNP complexes of dramatically different composition or kinetics. To date though, it has been difficult to study RNP complexes specific to any single mRNA species, as each transcript is relatively unabundant in the cell, and few robust techniques exist to specifically purify a particular mRNP for proteomic analysis. We thus sought to develop a method to isolate mRNPs from a single transcript, allowing us to study the dynamic RNP compositions of individual mRNA maturation pathways.

To optimize purifications of the protein tags required for RNP isolations, we first generated high affinity reagents targeting key tags like GFP and mCherry. Instead of traditional antibodies, we chose to use nanobodies: recombinant single domain derivatives of a heavy chain-only antibody variant found in camelids. The recombinant nature and small size of nanobodies make them ideal reagents for affinity isolations. We developed an improved pipeline for the identification of nanobody repertoires against any antigen of interest, which provided us with 25 nanobodies against

GFP, the most common and robust protein tag in use. This pipeline has also allowed us to develop nanobodies against a variety of other antigens of biomedical interest.

With the help of optimized reagents, we developed a two-step purification method allowing highly targeted isolations of mRNPs, starting in a budding yeast model system. In our approach, a single target transcript is tagged with MS2 hairpin sequences – these hairpins are bound specifically and with high affinity by the bacteriophage MS2 coat protein (MS2CP). In the first purification step, a chosen RNP protein known to be associated with a particular mRNA processing step of interest is Protein A-tagged and affinity isolated. From this material, anti-GFP nanobodies are used in the second step to isolate the MS2-tagged transcript of interest, through purification of MS2CP-GFP fusion proteins bound to the tag. This approach is able to efficiently and cleanly isolate a particular transcript at a chosen step of mRNP maturation. The use of an RNP factor as a separate purification target both improves overall purity and simplifies analysis by limiting heterogeneity of the mRNP mixture.

Using this novel method for single mRNP isolations, we have performed a preliminary survey of transcripts with distinct sequence elements suspected to be associated with unique processing machinery. Mass spectrometric (MS) analysis of RNPs co-purified with these transcripts revealed several RNA-specific changes in composition. Most notably, introns from either a house keeping ACT1 gene or the RPS30b ribosomal protein gene led to dramatically different levels of various splicing-related proteins. These differences provide mechanistic insight into changes in the kinetics of spliceosome assembly determined by intron sequence.

# Acknowledgements

First and foremost, I thank my advisor, Dr. Michael Rout, for his input and inspiration over the course of my graduate studies. His insights and unrelenting optimism have been invaluable. I also thank Dr. Brian Chait and Dr. Marlene Oeffinger for their crucial contributions to the conceptual framework underlying our approaches to nanobody discovery and mRNP purifications, respectively. I must also acknowledge all members of the Rout lab, past and present, who have combined to produce such a congenial, supportive, and intellectually stimulating environment. In particular, I owe a great deal to the technical and moral support of the technicians I've worked with, Mary Thompson and Jill Trivedi, who were integral to all stages of our nanobody studies.

I am also grateful to the members of my thesis committee, Dr. Brian Chait, Dr. Robert Darnell, and Dr. Daniel Zenklusen. Their regular advice and encouragement have always been valuable, and have greatly contributed to the direction of my work.

I am deeply indebted to the various collaborators who have made so many aspects of my work possible. Dr. Yinyin Li, Dr. Erica Jacobs, and Kelly Molloy from the Chait lab were enormously helpful in running and analyzing all mass spectrometry experiments. Dr. David Fenyö and Sarah Keegan were similarly essential to the computational work behind nanobody identification. I also thank Dr. Ilona Nudelman and Dr. Natalia Ketaren for their structural biology contributions, and Dr. Junjie Wang from the Chait lab for his insights into nanobody analysis.

# Table of Contents

# List of Figures

# List of Tables

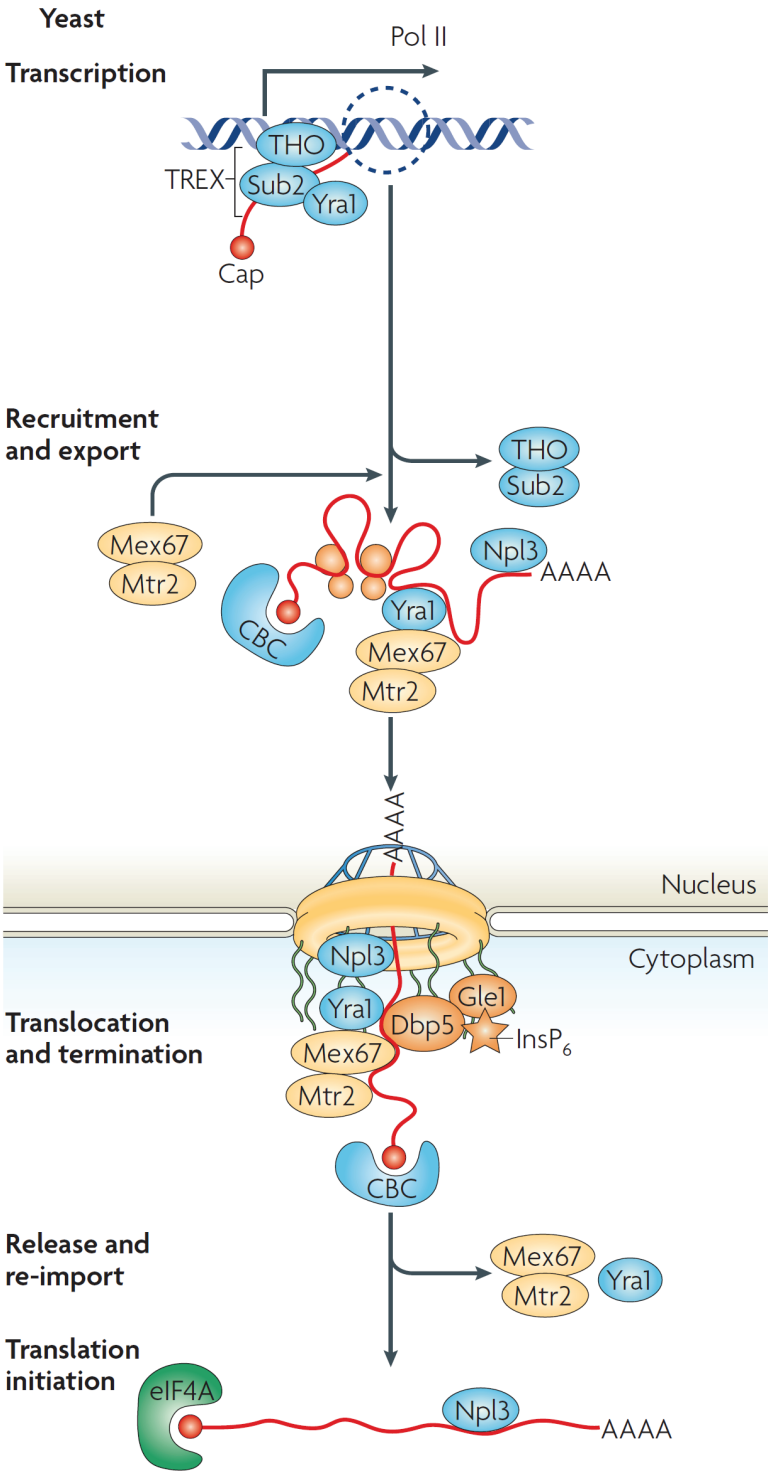
# Chapter 1: Introduction

### *Maturation of mRNP complexes in the nucleus*

The development of a cell nucleus that separates genetic information from other cellular processes was a key evolutionary transition, forcing a multitude of adaptations to allow information from the nucleus to be transmitted to and expressed in the cytoplasm. Messenger RNAs (mRNAs) transcribed from DNA in the nucleus must be exported through the nuclear envelope, by way of the nuclear pore complex (NPC), for subsequent cytoplasmic translation. Within this genetic information pathway, post-transcriptional processing of RNA is a complex step critical to the proper propagation and regulation of such genetic information. Beyond the transcription of a gene's sequence to RNA, distinct combinations of modifications, splicing events, and interactions with export and localization factors can be required for normal gene expression. These nuclear maturation and export processes are mediated through an elaborate, hierarchical network of mRNP complexes, comprised of dynamically associating RNA processing factors. While many of the essential processing steps along this pathway have been well-characterized, along with their key components, many questions remain regarding the timing and ordered assembly of mRNP complexes. Furthermore, while there is evidence that different mRNA transcripts follow differential processing and export pathways, the difficulties of analyzing individual subsets of mRNPs have made identification of alternative mRNA processing machinery difficult. However, recent methodological advances that, at least potentially, allow high quality isolations of individual classes of RNP complexes are permitting new insight into the dynamic organization of maturing mRNPs (Tackett et al., 2005; Gavin et al., 2006; Krogan et al., 2006; Oeffinger et al., 2007).

The basic framework for the biogenesis of a typical mRNP in the model organism *Saccharomyces cerevisiae* (budding yeast) is relatively well-established (Figure 1.1). During transcription by

Figure 1.1. Overview of mRNP maturation pathway. RNP factors in involved in mRNA transcription, processing, and export are shown.




From Köhler and Hurt, 2007

RNA polymerase II, the conserved THO/TREX complex associates with the nascent RNA transcript, along with splicing factors in the case of intron-containing RNAs (Strasser et al., 2002; Gornemann et al., 2005; Lacadie and Rosbash, 2005). RNA polymerase II also recruits the enzyme responsible for adding a 5' 7-methylguanylate cap, which is then bound by the cap binding complex (Lewis et al., 1996; Cho et al., 1997; Lewis and Izaurralde, 1997). While yeast TREX is primarily associated with the transcription machinery, in mammals, TREX is recruited in a splicing and CBC-dependent manner to the 5' end of mRNAs (Masuda et al., 2005; Cheng et al., 2006).

Components of the THO/TREX complex are subsequently able to recruit additional RNA-binding proteins, most notably the yeast Mex67-Mtr2 nuclear export factor, or its TAP-p15 homolog in metazoans (Gruter et al., 1998; Strasser and Hurt, 2000; Zenklusen et al., 2002; Abruzzi et al., 2004; Hurt et al., 2004). The Mex67-Mtr2 complex is primarily recruited by Yra1p (metazoan ALY or REF), an RNA-binding adaptor protein in the TREX complex, and mediates export through the nuclear pore complex (NPC) by direct interactions with FG-nucleoporins (Segref et al., 1997; Hurt et al., 2000; Strasser et al., 2000; Zhou et al., 2000). In contrast to typical karyopherin nuclear export receptors which derive the directionality of their cargo transport from how they interact with the GTPase Ran, Mex67-mediated export is Ran independent, and directionality is instead believed to be derived from cytoplasmic mRNP remodeling factors. This remodeling relies primarily on the ATP-driven RNA helicase activity of Dbp5p, an mRNP associating enzyme required for the removal of Mex67-Mtr2 (Snay-Hodge et al., 1998; Tseng et al., 1998; Lund and Guthrie, 2005). Dbp5p is not fully active until stimulated by Gle1p, which is localized to cytoplasmic NPC filaments, in conjunction with inositol hexakisphosphate ($IP_6$), a primarily cytosolic signaling small molecule (Alcazar-Roman et al., 2006; Weirich et al., 2006). According to this model, as the mRNP molecule passes through the NPC, this cytoplasmic

remodeling-dependent removal of Mex67-Mtr2 prevents transport back into the nucleus. There is also evidence that other factors contribute to this export termination, such as the regulated recruitment of Mex67-Mtr2 by Npl3p, driven by differential nuclear and cytoplasmic protein phosphorylation. When dephosphorylated in the nucleus, Npl3p can mediate interactions between mRNA and Mex67; after export however, it is phosphorylated by Sky1p, destabilizing the Mex67 interaction and promoting Npl3p import (Gilbert et al., 2001; Gilbert and Guthrie, 2004). Once in the cytoplasm, eIF4A and other translation initiation factors are recruited to the mRNP by the CBC, allowing translation to commence (Lewis and Izaurralde, 1997; Gingras et al., 1999).

Anchoring of a subset of active genes and their transcripts to the NPC is also an important, if relatively poorly understood element of RNA processing (Casolari et al., 2004; Iglesias and Stutz, 2008; Dieppois and Stutz, 2010). This association appears to involve many different interactions, such as binding of the SAGA transcription initiation complex by Sus1p, a member of the mRNP- and NPC-associated TREX-2 complex, and interactions between other NPC and mRNP proteins, such as Mex67p and Mlp1p (Galy et al., 2004; Rodriguez-Navarro et al., 2004; Dieppois et al., 2006). However, how certain genes and their transcripts are selected to undergo anchoring, the relative timing of this anchoring, and its relationship with various RNA processing events remain unclear. Several quality control mechanisms are also required for proper mRNA processing. This task first depends on the nuclear exosome, which is responsible for the sequestration and degradation of unspliced, unadenylated, or misassembled mRNPs (Bousquet-Antonelli et al., 2000; Hilleren et al., 2001; Zenklusen et al., 2002; Hieronymus and Silver, 2003). Surveillance of transcripts to determine which are faulty and thus targeted for degradation also occurs at the nuclear face of the NPC; there, it seems to chiefly require Mlp1p and Mlp2p, which are involved

in the retention of unspliced or improperly packaged mRNPs at the nuclear face of the NPC and away from the NPC's transport mechanism (Galy et al., 2004; Vinciguerra et al., 2005).

While most studies of mRNP maturation have taken a ground-up approach focusing on individual processing steps, there has been increasing interest in the characterization of the broader interactome of maturing RNP particles. Several proteomic studies have begun to comprehensively identify the factors comprising these dynamic complexes, revealing novel RNP binding proteins (Gavin et al., 2002; Gavin et al., 2006; Krogan et al., 2006; Oeffinger et al., 2007). However, as these studies have typically relied on the isolation of tagged RNP proteins rather than individual species of RNA, it has been difficult to survey anything other than mixtures of many different types of RNPs associating with a highly heterogeneous RNA population. Furthermore, while these types of pullouts have revealed discrete subsets of proteins dynamically interacting with and processing maturing RNA, many questions remain about the timing and order of assembly of such subcomplexes.

***Alternative mRNP maturation pathways***

There has been an increasing amount of evidence that certain mRNA processing factors associate preferentially with different subsets of mRNAs, indicating that there is not a single universal mRNP maturation pathway (Hieronymus and Silver, 2003; Kim Guisbert et al., 2005; Hogan et al., 2008; Tuck and Tollervey, 2013; Baejen et al., 2014). This seems to be in contrast to other classes of RNA, such as tRNAs, rRNAs, or miRNAs, which may have more defined processing mechanisms, consistent with their less variable structures (Venema and Tollervey, 1999; Kohler and Hurt, 2007). In addition to obvious differences such as the presence or absence of introns affecting interactions with splicing machinery, other more central RNP-binding proteins, such as Npl3p, display different binding preferences to different types of pre-mRNAs (Kim Guisbert et

al., 2005; Kress et al., 2008). Even Mex67-Mtr2 recruitment, central to mRNA export, can occur through different adaptor proteins in different transcripts, including Npl3p and the shuttling mRNA binding protein Nab2p, in addition to the canonical TREX proteins Yra1p and Sub2p (Gilbert and Guthrie, 2004; Iglesias et al., 2010). Certain transcripts may additionally require another export receptor, Crm1p, for proper nuclear export (Gallouzi and Steitz, 2001; Dong et al., 2007).

A variety of transcript-specific RNP interactions have also been identified that depend on sequences in the 3' UTR. One of the most well-studied examples is the 3' UTR of the ASH1 transcript in budding yeast. This gene encodes a transcription factor involved in mating type switching of daughter cells, and its mRNA localizes to the yeast bud during cell division (Bobola et al., 1996; Jansen et al., 1996). This localization is mediated by the interaction of She2p with sequence elements in the 3' UTR – She2p then recruits She3p and Myo4p, which transports the mRNA to the bud along actin filaments. This localization activity relies solely on stem-loop sequence elements in the transcript, which directly encode the alternative cytoplasmic transport through interactions with SHE proteins (Long et al., 2000; Olivier et al., 2005).

Surprisingly, promoter sequences for a number of genes have been implicated in the regulation of alternative post-transcriptional processing of their mRNA products, independent of the transcribed RNA sequence. The promoters of several genes have been shown to confer decreased stability of their mRNA products, including CLB2 and SWI5, whose transcripts become rapidly degraded during mitosis (Trcek et al., 2011). Several genes, such as RPL30, whose promoters include sequence elements recognized by the Rap1p transcription factor, also produce mRNAs subject to rapid degradation in the cytoplasm by the Xrn1 exonuclease (Bregman et al., 2011). This decrease in stability is encoded solely by the promoter elements recognized by Rap1p, not the transcript sequence itself. A similar phenomenon was identified in the case of a number of genes responsive

to glucose starvation in yeast, such as heat shock genes (Zid and O'Shea, 2014). For a number of transcripts upregulated during glucose starvation, it was found that promoter sequences alone encoded divergent fates for these mRNAs in the cytoplasm: either sequestration to foci and inhibition translation, or diffusion and rapid translation. In all of these cases, the mechanism allowing the promoter to determine its transcript's fate remains largely unclear.

***Surveys of mRNA pools associated with specific RNP factors***

While it appears clear that multiple mRNP maturation pathways exist, it has been difficult to comprehensively define any single mRNA's unique assortment of processing factors or intermediate complexes. The most common approach to this question has been analysis of mRNAs associated with a particular tagged RNP factor isolated from cells. A number of studies have used this approach in yeast, surveying co-purifying mRNAs by microarray or RNA-Seq (Hieronymus and Silver, 2003; Kim Guisbert et al., 2005; Hogan et al., 2008; Tuck and Tollervey, 2013; Baejen et al., 2014). Fundamentally, these studies used similar straightforward techniques, namely Protein A tagging and affinity isolation of major RNP proteins, followed by either RNA-seq or microarray analysis of purified RNA. In some cases, crosslinking was used to allow mapping of protein-bound RNA sites. To address these questions in mammalian cells, a more common approach has been to identify RNA binding proteins isolated with bulk RNA or synthetic RNA probes, and correlate RNPs with different transcript sequences (Butter et al., 2009; Castello et al., 2012).

Surveys of this kind have generated a great deal of data, but a few general patterns have been consistently found in multiple studies, all consistent with the view that mRNAs commonly have different preferences in their RNP interactions. For instance, different members of the shuttling hnRNP family, such as Nab2, Npl3, and Hrp1, have strikingly different mRNA specificities. Ribosomal protein transcripts in particular (in which introns are most common) have a significant

preference for Npl3. While Mex67 exhibits general affinity for all mRNAs, its adaptor proteins, Nab2, Npl3, and Yra1, each have distinct preferences for different groups of mRNAs. These preferences also exhibit certain patterns; longer transcripts, for instance, are preferentially bound to Nab2, while shorter ones tend to be preferred by Yra1.

While this transcriptomic approach has produced much insight on overall trends in RNA preferences of many RNPs, it is inherently limited in its ability to correlate such differences to higher order mRNP assemblies. That is, any particular RNP factor will likely be purified from a heterogenous mixture of different complexes, with different overall compositions. Even an mRNP protein's direct interactions with RNAs may be influenced by other components in the broader complex. Such complications could in part be addressed by performing concurrent proteomic analysis of such RNP purifications, allowing any identified mRNA enrichments to be correlated to overall complex protein composition. An alternative, however, would be mRNP purifications designed to specifically isolate a specific RNA transcript. By this approach, the association of an RNA sequence for any RNP factor could be directly observed in the context of the overall mRNP complex, as other proteins present can also be detected.

***Tools for isolating RNP complexes***

A number of labs have made efforts to tag specific RNAs with unique sequences recognized by a binding partner and adaptable to a pullout system (Bardwell and Wickens, 1990; Hogg and Collins, 2007; Said et al., 2009; Hogg and Goff, 2010; Slobodin and Gerst, 2010). To date, these methodologies have shown relatively low purity and yield. This has allowed detection of major cytoplasmic RNA-associated factors, but typically not lower-abundance factors found in the nucleus. In theory though, this type of system can allow the characterization of copurifying RNP

complexes by mass spectrometry and other analytical techniques, providing for a more direct study of an individual RNA processing pathway.

***The use of bacteriophage coat proteins in RNA tagging***

A variety of different RNA tagging systems have been developed over the years, with various advantages and disadvantages. The most common approach has proven to be the use of bacteriophage coat proteins binding a cognate RNA hairpin, specifically the MS2 and PP7 systems (Kozak and Nathans, 1972; Bertrand et al., 1998; Lim et al., 2001; Chao et al., 2008). Short 19-25 bp RNA hairpins found in these bacteriophages bind with high specificity and affinity ($K_D$ ~ $4x10^{-10}$ M) to their corresponding MS2 or PP7 binding proteins (Figure 1.2a). By fusing these bacteriophage coat proteins to a tag such as GFP and co-expressing with transcripts genetically tagged with the corresponding RNA hairpins, GFP can be targeted to nascent transcripts through the coat protein-hairpin interaction (Figure 1.2b). This technique has been extensively used for fluorescent visualization of mRNA transcripts in living cells, as well as in tethering proteins to RNAs *in vivo* (SenGupta et al., 1996; Bertrand et al., 1998; Rodriguez et al., 2007; Hocine et al., 2013).

In a few cases, MS2 or PP7 have been used with varying levels of success to purify tagged RNAs from cells, along with associated RNP proteins. PP7 hairpins, for instance, were used to tag and purify RNP complexes of 7SK non-coding RNA (ncRNA), an abundant snRNP regulating transcription (Hogg and Collins, 2007). The yield in this approach was limited to 20-30% however, and further limited by non-specific contaminants. Related approaches with either PP7 or MS2 have been used to identify abundant factors from a primarily cytoplasmic mRNA pool, with similar limitations (Said et al., 2009; Hogg and Goff, 2010; Tsai et al., 2011). To date, these

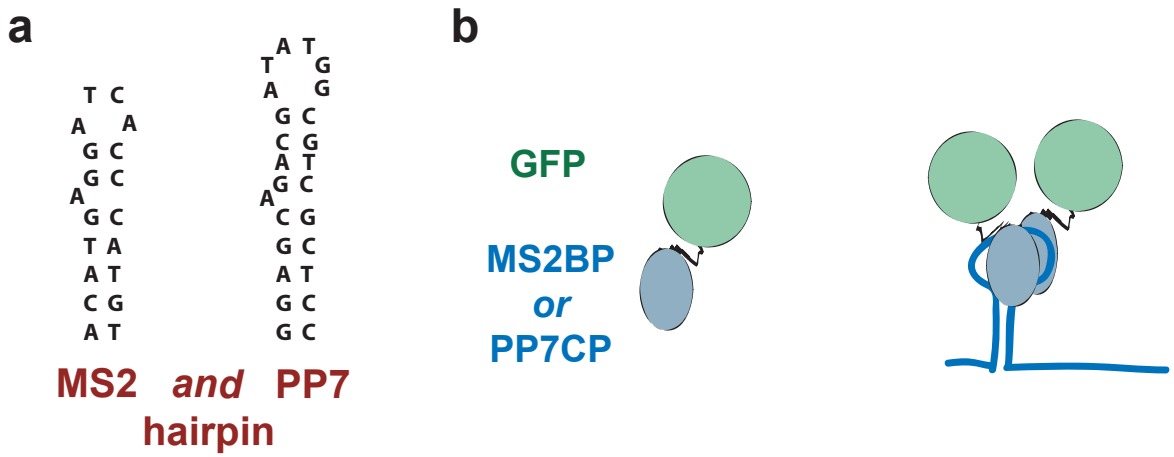



Figure 1.2. Bacteriophage hairpins and coat proteins. (a) RNA hairpin structures bound by MS2 or PP7 coat proteins. (b) Binding of GFP-tagged MS2 or PP7 coat proteins as dimers to cognate hairpins.

techniques have not been applied to the study of lower abundance mRNP factors typically associated with nuclear mRNAs.

Other bacteriophage proteins have also been successfully used in these approaches. Effective systems have been developed from the N protein from bacteriophage λ, known as $\lambda_N$. This protein recognizes a short RNA hairpin structure, BoxB, with binding activity determined by a 22 amino acid region that can be independently synthesized or expressed (Franklin, 1985; Baron-Benhamou et al., 2004; Daigle and Ellenberg, 2007). Similar to MS2 and PP7, this $\lambda_N$ peptide specifically binds its cognate 15 bp hairpin structure with high affinity, and has been successfully used in RNA visualization and protein tethering.

***Alternative approaches to RNA affinity capture***

Beyond the widely-adopted bacteriophage-based systems, alternative RNA purification techniques have also been attempted, including the use of antisense oligomers complementary to unique RNA sequences. Immobilized oligomers can potentially recognize complementary RNA sequences in cell lysate, particularly using synthetic oligomers such as 2'O-methyl-RNA, LNA (locked nucleic acids), or PNA (peptide nucleic acids), which are able to form more stable hybrid structures with complementary RNA sequences (Cotten et al., 1991; Kumar et al., 1998; Larsen et al., 1999). While these types of reagents have been extensively used for various hybridization techniques, few studies have succeeded in adapting this technology to isolating intact RNPs from target RNA, and typically only in the case of stable, highly abundant RNA-protein complexes (Grunweller et al., 2003; Upadhyay et al., 2013). More typically, these types of approaches have been used to isolate direct RNA-binding proteins under stringent, denaturing conditions (Castello et al., 2012). This relies on covalent crosslinking between RNAs and interacting proteins, but the conditions

required for hybridization and washing of the RNAs to complementary oligomers do not allow for recovery of additional complex components.

Similar techniques have been developed through the use of RNA aptamers, short RNA secondary structure-forming sequences that are designed to bind to substrates such as streptavidin, streptomycin, and tobramycin. These aptamers have been used to tag RNAs of interest, allowing affinity capture by the immobilized ligand (Bachler et al., 1999; Srisawat and Engelke, 2001; Hartmuth et al., 2002). This technology has been used, for example, to isolate transcripts containing AU-rich elements (AREs) from HEK293 cells, copurifying major ARE-binding factors like HuR (Vasudevan and Steitz, 2007). However, this approach has been limited to some degree by slow kinetics of binding to streptavidin, low yield (~20%), and non-specific binders (Bachler et al., 1999; Srisawat and Engelke, 2001). Aptamer tags have also been used in complementary techniques, such as immobilization of *in vitro*-transcribed RNAs in order to capture RNA binding proteins from cell lysate (Butter et al., 2009).

***Reagents for affinity isolations***

The most promising methods for tagging and isolating RNA transcripts have tended to rely on MS2 or GFP bacteriophage coat proteins fused to GFP or Protein A (Hogg and Goff, 2010; Oeffinger, 2012). As efficient purification of these fusion proteins is critical for a successful isolation of their RNA targets, any such technique must rely on high quality reagents to purify GFP or Protein A.

In the case of common protein tags like GFP, mCherry, FLAG, myc, or HA, the vast majority of studies make use of high quality commercial antibodies to target the tag, particularly when affinity isolation is required (Rigaut et al., 1999; Ho et al., 2002; Cristea et al., 2005; Domanski et al.,

2012). Although monoclonal or polyclonal antibodies remain the primary bait reagents available for these purposes, they have a variety of limitations that can sometimes prove problematic. They are large multimeric proteins, at approximately 150 kDa, and this can result in steric hindrances or limit density when attached to surfaces, such as resins used for immobilization. Furthermore, the high cost of commercial reagents, limited availability, and batch to batch variation have often proved problematic for providing the large amounts of consistently high quality affinity capture reagents needed for such biochemical or proteomic studies (Gingras et al., 2005).

***Single domain antibodies***

Due to the limitations of traditional IgG antibodies, a number of efforts have been made to generate alternatives of smaller size, preferably suitable to production as single recombinant proteins. Some of the earliest and most successful such reagents are synthetic single-chain antibodies, often termed scFvs, created by tethering variable heavy ($V_H$) and light ($V_L$) chains from monoclonal antibodies (Figure 1.3) (Bird et al., 1988; Huston et al., 1988; Skerra and Pluckthun, 1988). While there have been many highly successful examples of such scFvs, whose affinities can rival their monoclonal IgG progenitors, they can also prove problematic in terms of stability (Dumoulin et al., 2002). Furthermore, development of a traditional scFv first requires generation of a high quality monoclonal antibody, followed by engineering of a recombinant sequence able to properly fold into its native $V_H/V_L$ conformation, making this a labor- and time-intensive process.

The disadvantages of scFvs are largely due to their two-domain structure, so an antibody fragment containing only a single variable domain is an appealing alternative. However, individual $V_H$ or $V_L$ domains are typically not sufficiently stable or soluble on their own, and may not retain high antigen affinity when unpaired (Ward et al., 1989). A promising answer to this problem came with the discovery in camels of a novel class of IgG molecules made only of a heavy chain homodimer,

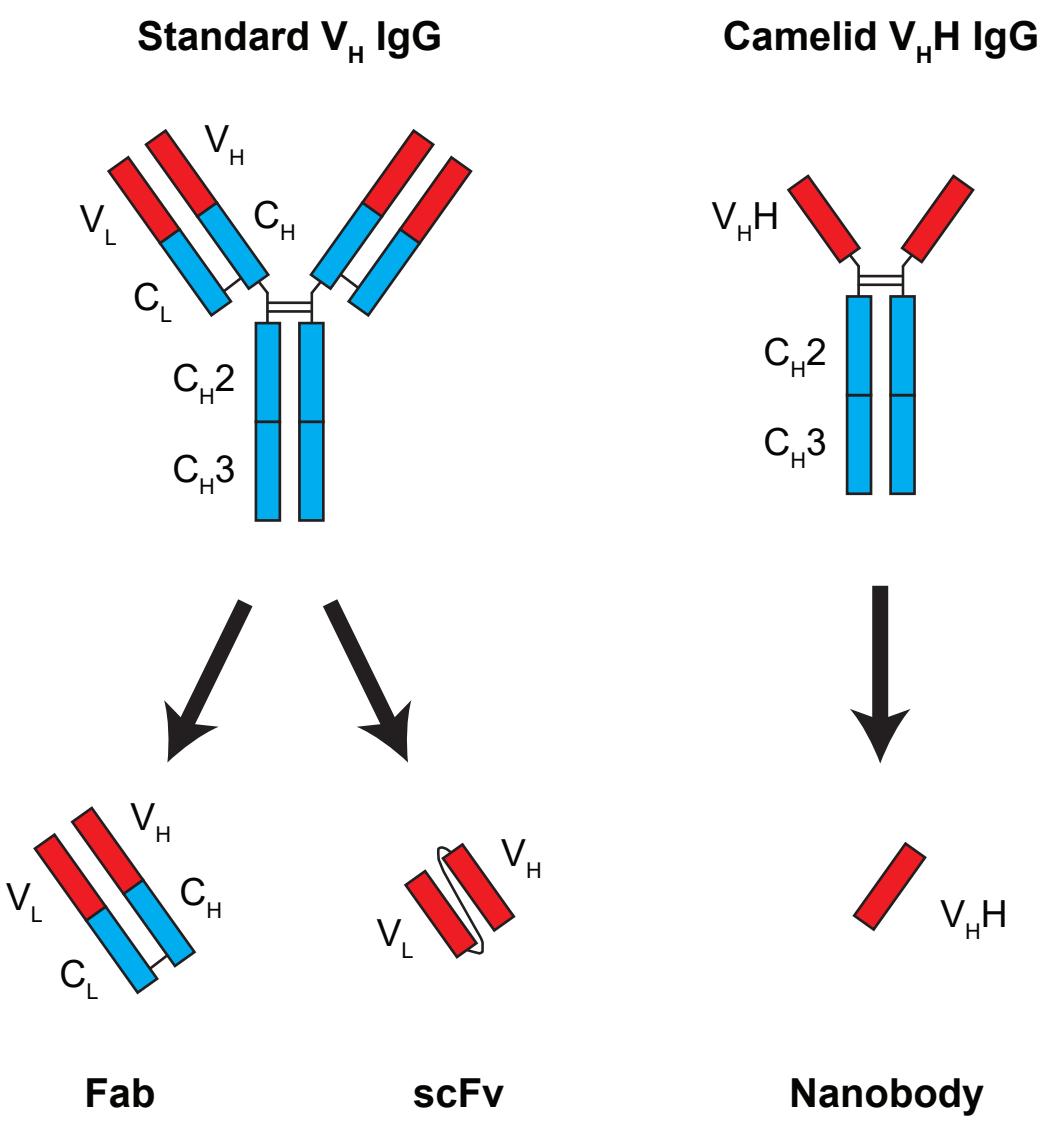



Figure 1.3. IgG variants and their derivatives. Standard IgG structures contain both heavy and light chains. Variable regions from both chains must be combined for use as an Fab fragment or for recombinant expression as a linked scFv. The camelid $V_HH$ IgG variant has only heavy chains. The single variable region can be cloned and expressed independently as a nanobody.

lacking any associated light chains (Figure 1.3) (Hamers-Casterman et al., 1993). This unusual antibody variant, termed $V_HH$, has only been observed in camelid species, including llamas and alpacas, though an analogous heavy chain-only Ig (IgNAR), has been identified in cartilaginous fish (Greenberg et al., 1995). The $V_HH$ variable region is thus the smallest antigen-binding single polypeptide chain naturally found in the antibody world (Arbabi-Ghahroudi et al., 1997; Dumoulin et al., 2002; Harmsen and De Haard, 2007; Romer et al., 2011; Muyldermans, 2013).

As these $V_HH$ antibodies retain strong affinity for their antigens, the variable regions responsible for binding activity can be cloned out and expressed as independent antigen-binding proteins known as nanobodies (Arbabi-Ghahroudi et al., 1997; Cortez-Retamozo et al., 2004; Harmsen and De Haard, 2007; Muyldermans, 2013). Unlike monoclonal antibodies or many other antibody fragments, they can be readily produced in large, soluble amounts in bacterial expression systems (Arbabi-Ghahroudi et al., 2005; Harmsen and De Haard, 2007). Moreover, nanobodies are usually extremely stable, can bind antigens with affinities in the nanomolar or sub-nanomolar range, and are much smaller in size (approximately 15 kDa) than other antibody constructs (Bird et al., 1988; Skerra and Pluckthun, 1988; Worn and Pluckthun, 2001; Dumoulin et al., 2002; Rothbauer et al., 2006; Muyldermans et al., 2009).

However, rapid and robust techniques for the isolation of extensive repertoires of high affinity nanobodies have proven elusive – the labor-intensive nature and poor efficiency of current approaches have been a major bottleneck for the widespread implementation of these reagents (Arbabi-Ghahroudi et al., 1997; Rothbauer et al., 2006; Muyldermans, 2013), explaining why demand for these reagents greatly exceeds supply (Muyldermans, 2013).

***Tools for targeted affinity isolation of single mRNP populations***

To allow closer examination of the multitude of likely alternative mRNP processing pathways, we have aimed to develop techniques and reagents for the isolation and analysis of individual subsets of mRNPs. This has involved adapting and optimizing the most promising approaches to tagging and isolating mRNAs and their associated mRNP factors. A key component of these techniques is highly efficient protein affinity capture reagents, for which we turned to nanobodies. Given the limitations of existing nanobody development methods, we generated a new pipeline for nanobody discovery based on a combination of mass spectrometry (MS) and high-throughput sequencing. This has provided optimized nanobodies for the high efficiency anti-GFP isolations required for our mRNP purification method. Secondarily, this pipeline has allowed us to generate repertoires of nanobodies against a variety of additional antigens of biomedical interest.

# Chapter 2: A new pipeline to generate anti-GFP nanobodies

### *Strategies for identifying high-affinity nanobody clones*

Given the importance of highly efficient isolations of GFP for an RNA purification scheme making use of MS2CP-GFP, we sought optimized reagents targeting this protein tag. As recombinant nanobodies are among the most efficient tools available for protein affinity isolation, and are not currently available against GFP in the quantities required, we sought to generate a new repertoire of anti-GFP nanobody clones.

In order to identify camelid $V_HH$ clones that can be expressed as a recombinant nanobody against an antigen of interest, such sequences must be cloned out and selected from immunized animals. Traditionally, this has been done by phage display performed with material from llamas (Arbabi-Ghahroudi et al., 1997; Conrath et al., 2001; Rothbauer et al., 2006; Alvarez-Rueda et al., 2007; Pardon et al., 2014). In this method, llama lymphocytes are taken from peripheral blood samples, and a cDNA library is generated from total mRNA. DNA primers against conserved constant region sequences flanking the $V_HH$ region are then used to PCR amplify this variable domain. These amplicons are then cloned into a vector that expresses them fused to a bacteriophage coat protein, producing a library that can be transformed into *E. coli* to release phages displaying $V_HH$ domains on their surface. These phages can then be bound to the immobilized antigen of interest, and target-specific $V_HH$-expressing phages can be collected after washing. This panning process is typically repeated multiple times with the enriched phage populations, after which individual clones can be selected and screened to confirm binding activity.

While this phage display approach has successfully been used to produce nanobodies against a variety of targets, it has practical disadvantages. It is a time and labor intensive process, often taking a total of 3-6 months from beginning to end (Pardon et al., 2014). It carries risk of false

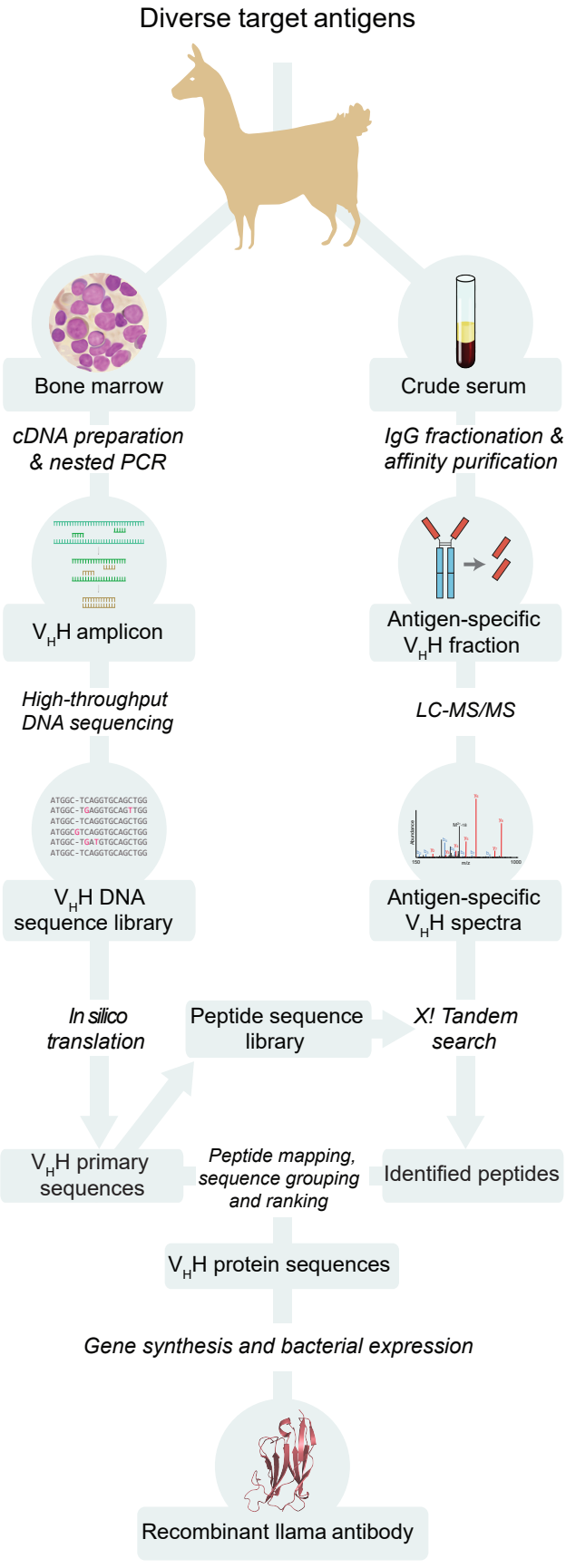
negatives, as $V_HH$ sequences may not fold correctly when displayed on a phage surface, as well as false positives, since increased avidity of multiple displayed proteins on a phage can amplify apparent binding of weak or non-specific clones. Furthermore, this approach has often only identified only a few stable, high affinity nanobodies, even for well-behaved protein antigens such as GFP (Rothbauer et al., 2006). To overcome these limitations, we sought to develop an orthogonal approach to nanobody discovery that avoids all these issues and allows simpler, more rapid identification of larger nanobody repertoires. We therefore worked to design a highly optimized pipeline that allows the rapid production of repertoires of high affinity nanobodies against selected proteins. This concept is inspired by previous efforts that identified circulating neutralizing HIV antibodies in humans using MS in conjunction with patient-specific antibody cDNA databases (Scheid et al., 2011). The approach is thus based on high-throughput DNA sequencing of a marrow lymphocyte $V_HH$ cDNA library from an immunized llama combined with mass spectrometric (MS) identification of high affinity $V_HH$ regions derived from serum of the same animal.

***Strategy for nanobody identification***

Our approach to nanobody discovery centers on MS identification of affinity-purified heavy-chain antibodies isolated from an individual llama, in correlation with a DNA sequence database generated from the same animal (Figure 2.1). Our approach represents a novel pipeline for nanobody production where each stage has been highly optimized.

To generate nanobody repertoires of maximal utility, we chose the GFP (green fluorescent protein) and mCherry tags for our first target antigens, due to their central roles in cell biological studies (Shaner et al., 2005). GFP is perhaps the most universally used tag, presenting many advantages over other tags. Most obviously, it permits both localization and interactomic studies in parallel

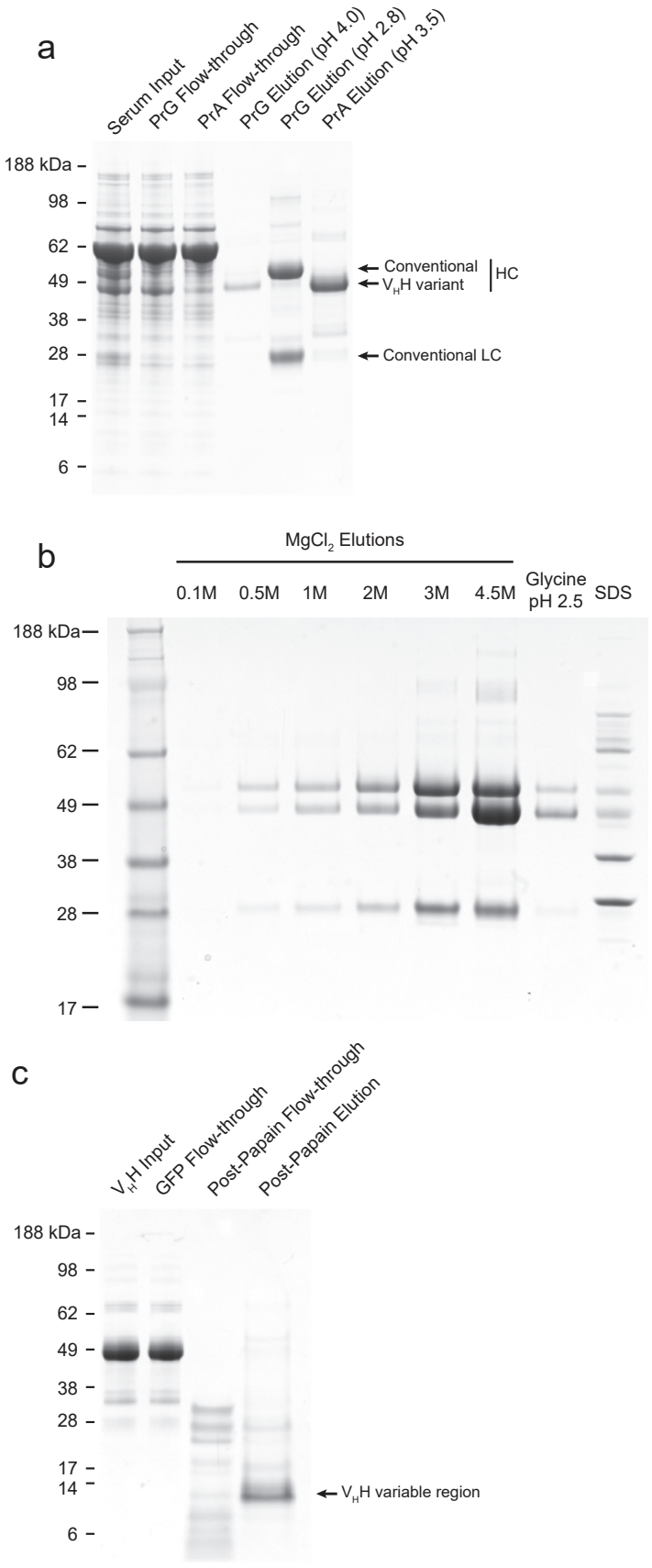
Figure 2.1. Overview of nanobody identification and production pipeline. The example nanobody structure shown was obtained from PDB 3K1K (Kirchhofer et al., 2010). LC-MS/MS, liquid chromatography–tandem mass spectrometry.


Diverse target antigens
Bone marrow
Crude serum
cDNA preparation & nested PCR
IgG fractionation & affinity purification
V_HH amplicon
Antigen-specific V_HH fraction
High-throughput DNA sequencing
LC-MS/MS
V_HH DNA sequence library
Antigen-specific V_HH spectra
In silico translation
Peptide sequence library
X! Tandem search
V_HH primary sequences
Peptide mapping, sequence grouping and ranking
Identified peptides
V_HH protein sequences
Gene synthesis and bacterial expression
Recombinant llama antibody

(Poser et al., 2008; Hutchins et al., 2010), and many thousands of proteins in many model organisms have already been tagged with GFP for visualization, at an enormous cost in both research time and expense. Based on PubMed searches, the Rout and Chait groups have conservatively estimated that in excess of $1 billion has been spent on GFP-tagged reagents. Similarly, mCherry is a very widely used red-fluorescing alternative to GFP, often used as its complement.

Further, while these fluorescent proteins have a broadly similar beta barrel structure, they are significantly evolutionarily divergent, making for very distinct immunogens (Shagin et al., 2004). After immunization of individual llamas with these antigens and confirmation of an immune response, we serially fractionated serum bleeds to obtain exclusively $V_HH$-containing heavy chain antibodies (Figure 2.2a), taking advantage of the differential specificity of Protein A and Protein G for $V_HH$ versus conventional antibodies (Hamers-Casterman et al., 1993). The $V_HH$-containing fraction was affinity purified over antigen-coupled resin, washed with $MgCl_2$ at various stringencies to remove non-specific and weakly specific binders (Figure 2.2b), and digested with papain on-resin to cleave away the constant regions and leave behind the desired minimal $V_HH$ variable region fragments from the remaining high-affinity binders. Finally, the antigen-bound $V_HH$ fragments were eluted and separated by SDS-PAGE, allowing the purification of the ~15 kDa $V_HH$ fragments away from residual conventional $F_{ab}$ fragments and $F_c$ fragments (both ~25 kDa), and undigested antibodies (~50 kDa) (Figure 2.2c). The gel-purified bands were trypsin-digested and analyzed by liquid chromatography-MS and MS/MS (with Yinyin Li, B. Chait lab) (Figure 2.3). We recovered the highest affinity $V_HH$ fragments by using the highest stringency washes, which also decreased the complexity of the eluted sample, aiding MS analysis.

Figure 2.2. Purification of an antigen-specific $V_HH$ fraction. (a) Llama serum fractionated by successively binding to Protein G and Protein A agarose resin. $V_HH$ variant heavy chains were specifically eluted in pH 4.0 (Protein G) and pH 3.5 (Protein A) buffers. (b) The GFP affinity of serum VH and $V_HH$ IgG was assessed by resistance to high stringency $MgCl_2$ washes. A llama serum sample was bound to GFP-sepharose resin, which was then serially washed with increasing $MgCl_2$ concentrations, followed by elutions with 0.1M glycine-HCl, pH 2.5 and boiling SDS. (c) Pooled $V_HH$ fraction was bound to GFP-Sepharose resin, and washed with 3.5 M $MgCl_2$. Bound IgG was digested on the beads with papain, releasing Fc fragments. The remaining GFP-bound $V_HH$ fragments were eluted with ammonium hydroxide.


a
Serum Input
PrG Flow-through
PrA Flow-through
PrG Elution (pH 4.0)
PrG Elution (pH 2.8)
PrA Elution (pH 3.5)
188 kDa
98
62
49
38
28
17
14
6
Conventional
VHH variant
HC
Conventional LC
b
MgCl2 Elutions
0.1M
0.5M
1M
2M
3M
4.5M
Glycine pH 2.5
SDS
188 kDa
98
62
49
38
28
17
c
VHH Input
GFP Flow-through
Post-Papain Flow-through
Post-Papain Elution
188 kDa
98
62
49
38
28
17
14
6
VHH variable region

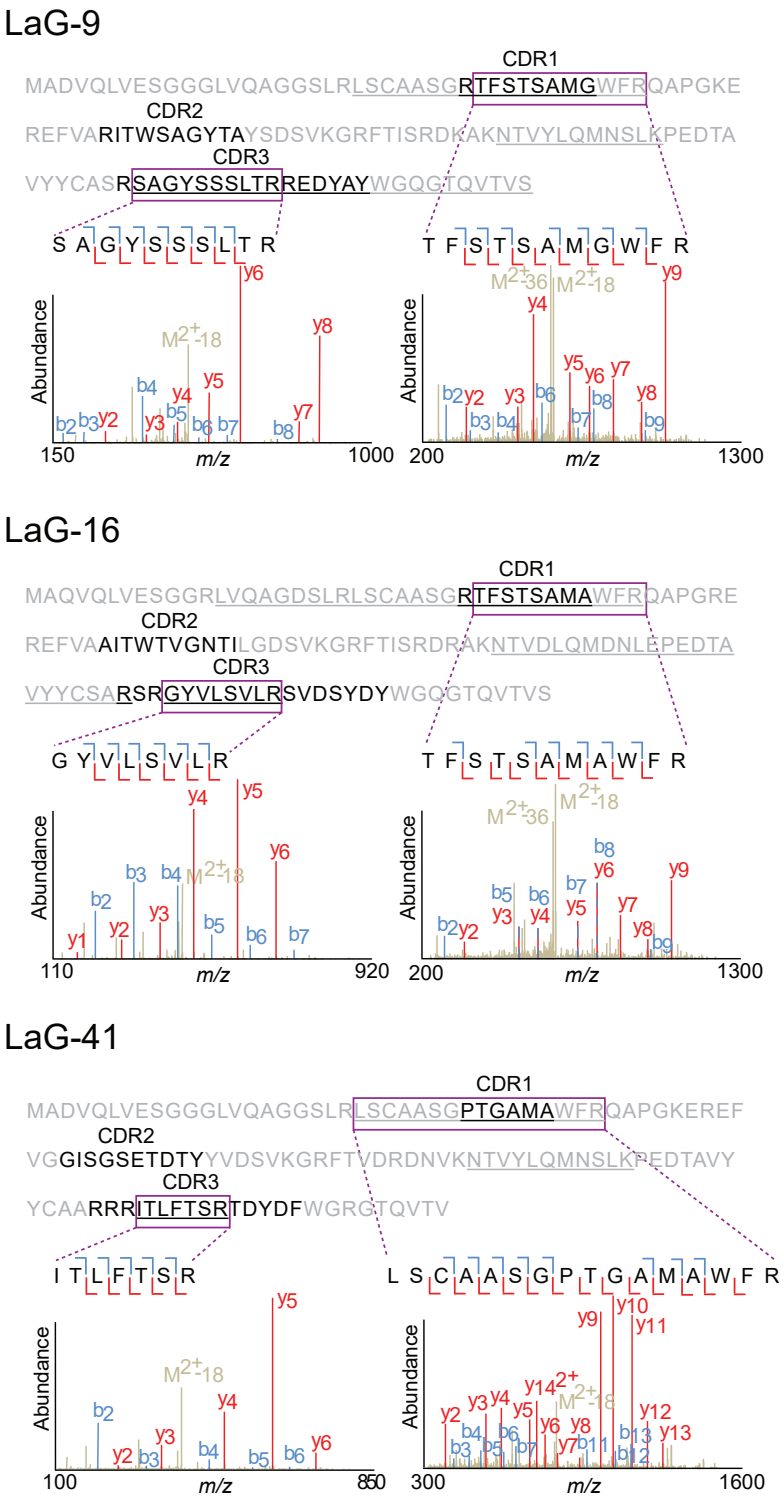



Figure 2.3. $V_HH$ peptide identification of mass spectrometry. Tandem mass spectra of identified peptides (shown boxed), mapped to the CDR regions of three candidate $V_HH$ sequences. The MS-covered regions of these sequences are underlined. Dashed lines indicate overlapping peaks. MS analysis was performed by Yinyin Li.

To create an animal-specific antibody sequence database, lymphocyte mRNA samples from individual immunized llamas were obtained for high-throughput sequencing. Mononuclear cells were isolated from bone marrow aspirates, enriching for long-lived antibody secreting plasma cells, which express elevated levels of immunoglobulin RNA (Benner et al., 1981; Dorner and Radbruch, 2007; Scheid et al., 2011). Importantly, we do not create expression libraries, and thus remove the need for efficient exogenous expression, folding, and presentation of the clones, and avoiding issues of codon adaptation incompatibility.

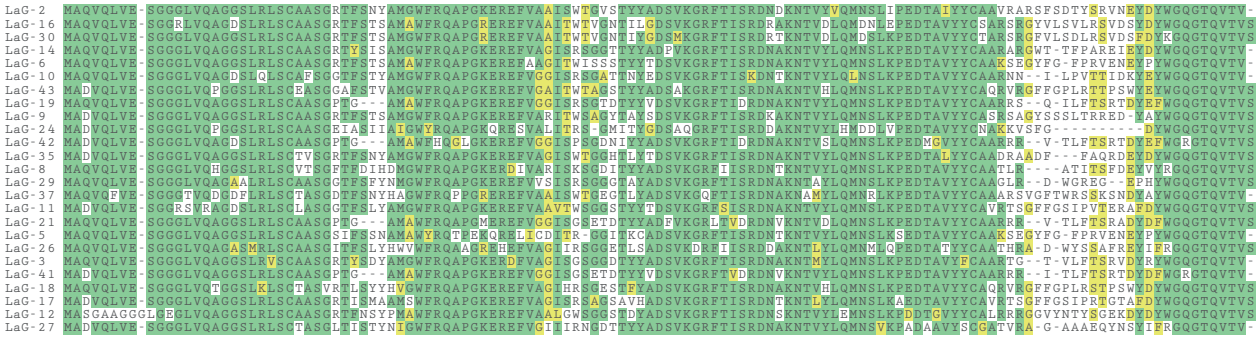
Total RNA from these lymphocytes was reverse transcribed, and a nested PCR was performed to specifically amplify sequences encoding the $V_HH$ variable regions (Rothbauer et al., 2006). This PCR product was sequenced using high-throughput 454 (GFP) or MiSeq (mCherry), resulting in ~800,000 and ~3,000,000 unique reads, respectively. These reads were translated, filtered and trypsin-digested *in silico* to create a searchable peptide database for MS analysis (Figure 2.1).

The identification of specific $V_HH$ sequences is more challenging than typical proteins, as they consist in large part of highly conserved framework regions that are less easily distinguished by MS. Moreover, rather than searching well-established databases, a $V_HH$ cDNA database must be generated for each immunized animal. To deal with both challenges, we developed (with Sarah Keegan and David Fenyö, NYU) a bioinformatic pipeline that is able to identify the highest probability matches from a large pool of related $V_HH$ sequences (Llama Magic; http://www.llamamagic.org). In this pipeline, $V_HH$ sequences were ranked by a metric based on MS/MS sequence coverage of complementarity determining region 3 (CDR3, the most diverse $V_HH$ region) as well as CDR1 and CDR2 coverage, total $V_HH$ coverage, sequencing counts, mass spectral counts, and the expectation values of matched peptides. Preliminary attempts to identify $V_HH$ sequences solely by their CDR3 regions revealed that identical CDR3 sequences are

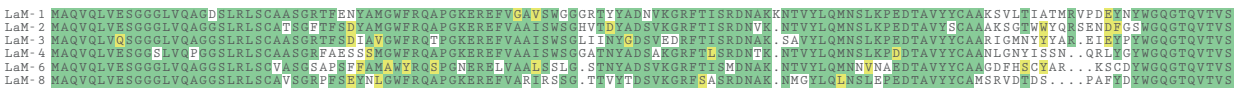
frequently shared between multiple distinct $V_HH$ sequences, with diverse CDR1 and CDR2 sequences. It is possible that this is a result of somatic gene conversion, in which, after V(D)J recombination, secondary recombination occurs between upstream V gene segments and already rearranged V(D)J genes (Becker and Knight, 1990; Knight, 1992). Alternatively, PCR amplification could possibly produce apparent recombination of this type, if a conserved region on an incomplete PCR product serves to prime an unrelated sequence, thus fusing two different $V_HH$ sequences. Our automatic ranking pipeline, coupled with careful manual inspection, overcame these issues and provided us 44 high-probability hits against GFP, classified as LaG (Llama antibody against GFP) 1-44, which we subjected to further screening. A smaller subset of eight clones was similarly chosen for follow up (LaM 1-8) for mCherry.

Codon optimized genes for these hits were synthesized and cloned into a bacterial expression vector. After expression, lysates were passed over antigen-coupled resin to identify nanobodies that displayed both robust expression as well as high and specific affinity. From these screens, we found 25 specific nanobodies against GFP (LaGs) and 6 against mCherry (LaMs). Phylogenetic analysis of the verified nanobodies revealed substantial sequence diversity among clones (Figure 2.4). While not directly analogous, the high success rate of this single screening step (57-75%) is favorable in comparison to the final panning and selection steps of phage display, in which batches of up to $10^7$ clones are screened repeatedly to identify even a few positive clones (Arbabi-Ghahroudi et al., 1997; Conrath et al., 2001; Rothbauer et al., 2006; Alvarez-Rueda et al., 2007). The affinity of these nanobodies was further assessed by either surface plasmon resonance (SPR) or *in vitro* binding assays with immobilized nanobodies. For the larger repertoire of LaGs, these experiments revealed a wide range of affinities, with $K_d$s from 0.5 nM to over 20 µM (Table 1),

a



b



c

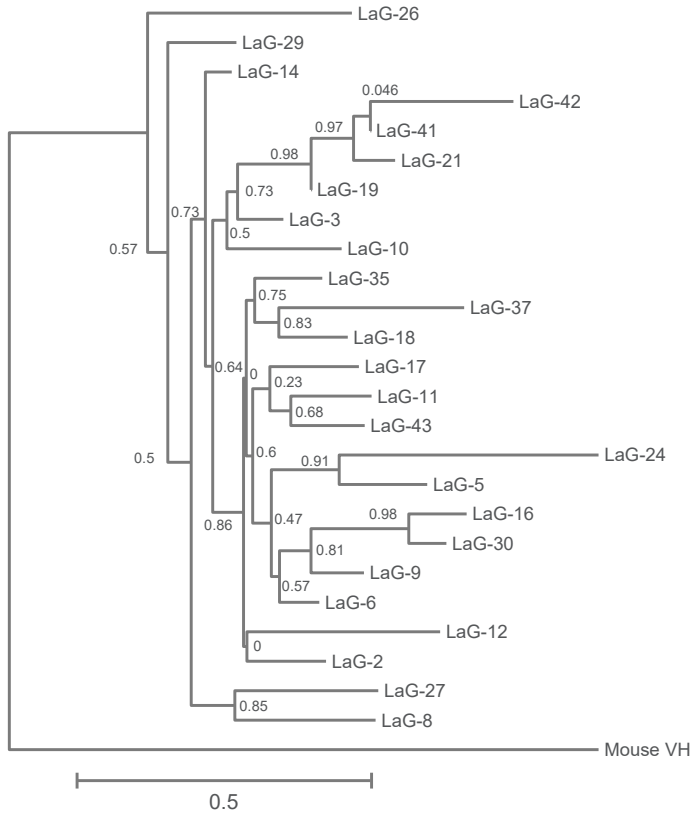



Figure 2.4. Phylogenetic analysis of nanobody sequences. (a) LaG and (b) LaM sequences were aligned, and (c) a phylogenetic tree was generated from LaG sequences, using the Phylogeny.fr web service.

Table 1. Characteristics of LaG, LaG Dimer, and LaM proteins. $K_d$ values for GFP and mCherry binding were determined by SPR unless otherwise noted. $K_d$ values are also listed for LaG dimers fused by a glycine-rich peptide linker (three repeats of GGGGS, or $G_4S$) or by a 3xFlag linker. For yeast Nup84-GFP and mammalian RBM7-GFP affinity isolations using LaGs, Coomassie-stained bands from elutions separated by SDS-PAGE were quantified, and known specific and nonspecific bands were used to calculate signal-to-noise (S:N) ratios. Bead binding assays were used to determine affinity for fluorescent proteins, and the table highlights differences in specificity for *A. macrodactyla* CFP among LaGs and for DsRed among LaMs. GFP epitopes for LaGs were determined by NMR and classified into three groups according to their location (I–III). Also shown are the number of residues identified in the binding site and the site's calculated accessible surface area (ASA). MW, molecular weight; nd, not determined; N/A, not applicable.

[a]$K_d$ determined by bead binding assay.

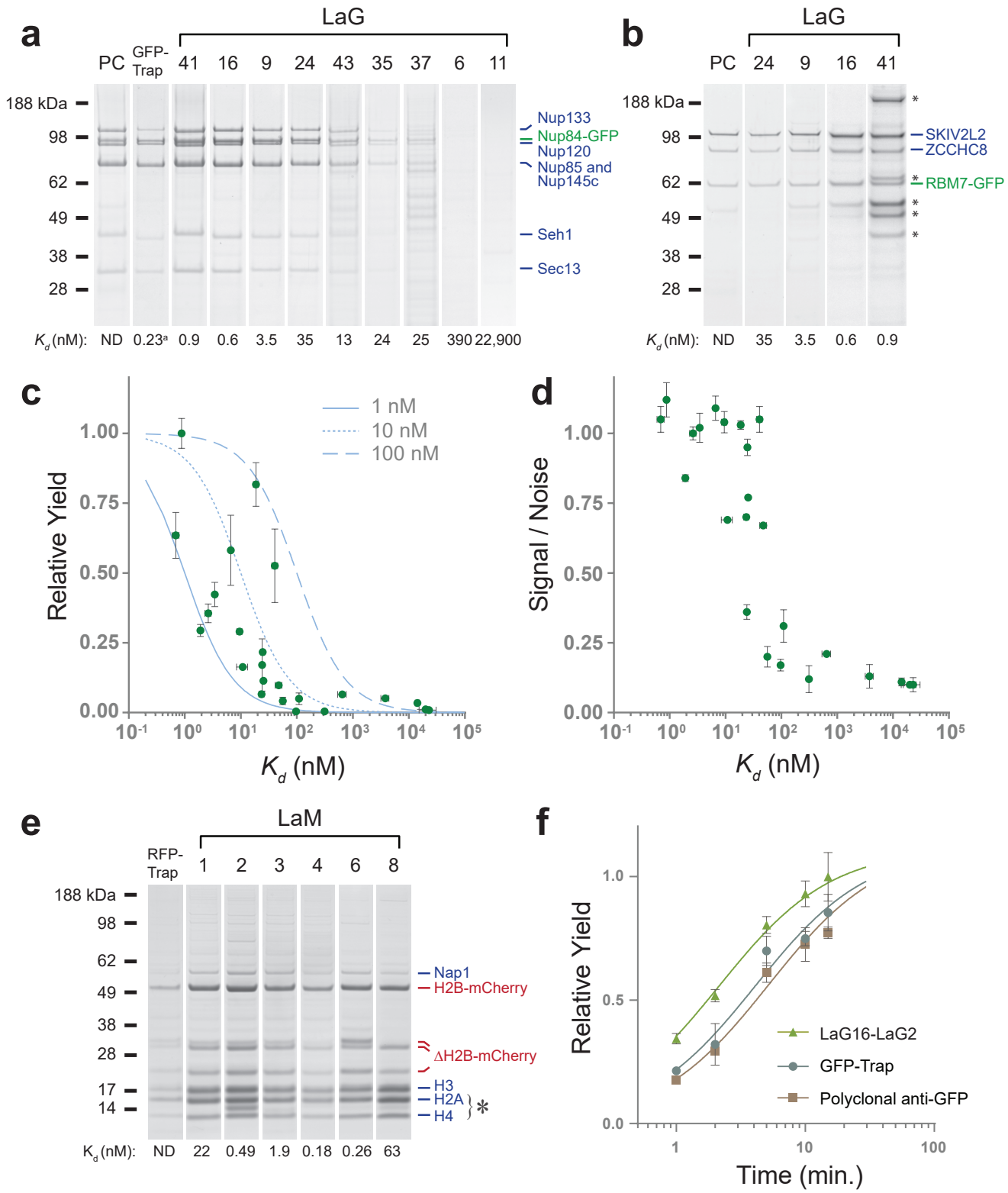
| Clone ID | MW (Da) | $K_d$ (nM) | Nup84 -GFP S:N | RBM7-LAP S:N | Binds *Am*CFP (LaG) / DsRed (LaM) | GFP Epitope | No. of binding site residues | ASA of binding site residues ($A^2$) |
|---|---|---|---|---|---|---|---|---|
| LaG-2 | 15,919 | 19[a], 16 | 1.03 | 0.42 | – | III | 55 | 2,204 |
| LaG-3 | 15,329 | 25 | 0.77 | 1.13 | + | nd | nd | nd |
| LaG-6 | 15,700 | 310 | 0.12 | nd | + | nd | nd | nd |
| LaG-9 | 16,062 | 3.5 | 1.02 | 1.04 | + | I | 62 | 2,551 |
| LaG-10 | 15,748 | 97 | 0.17 | nd | + | nd | nd | nd |
| LaG-12 | 16,090 | 56 | 0.20 | nd | + | nd | nd | nd |
| LaG-14 | 16,002 | 1.9 | 0.84 | 0.58 | + | I | 66 | 2,519 |
| LaG-16 | 16,306 | 0.7 | 1.05 | 0.92 | + | I | 60 | 2,605 |
| LaG-17 | 15,823 | 50 | 0.67 | nd | + | I | 60 | 2,543 |
| LaG-19 | 15,528 | 24.6[a] | 0.95 | 1.06 | + | II | 54 | 2,404 |
| LaG-21 | 15,452 | 7 | 1.09 | nd | + | II | 56 | 2,340 |
| LaG-24 | 14,763 | 41 | 1.05 | 1.09 | – | III[2] | nd | nd |
| LaG-26 | 16,221 | 2.6 | 1.00 | nd | + | II | 53 | 2,070 |
| LaG-27 | 15,565 | 9.5 | 1.04 | nd | + | II | 57 | 2,216 |
| LaG-29 | 15,449 | 110 | 0.31 | nd | + | nd | nd | nd |
| LaG-30 | 16,159 | 0.5 | 1.04 | nd | + | nd | nd | nd |
| LaG-35 | 16,010 | 23.5[a] | 0.70 | nd | + | nd | nd | nd |
| LaG-37 | 16,329 | 24 | 0.36 | nd | + | nd | nd | nd |
| LaG-41 | 15,471 | 0.9 | 1.12 | 0.41 | + | II | 53 | 2,091 |
| LaG-42 | 15,490 | 600 | 0.21 | nd | + | nd | nd | nd |
| LaG-43 | 16,167 | 11 | 0.69 | nd | + | I | 55 | 2,381 |
| LaG-5 | 15,589 | 14,200[a] | 0.11 | nd | nd | nd | nd | nd |
| LaG-8 | 15,953 | 20,000[a] | 0.10 | nd | nd | nd | nd | nd |
| LaG-11 | 16,221 | 22,900[a] | 0.10 | nd | nd | nd | nd | nd |
| LaG-18 | 16,459 | 3,800[a] | 0.13 | nd | nd | nd | nd | nd |
| LaG16-$G_4S$-2 | 30,791 | 0.036 | nd | nd | nd | nd | nd | nd |
| LaG16-3xFLAG-2 | 32,972 | 0.268 | nd | nd | nd | nd | nd | nd |
| LaG41-$G_4S$-2 | 29,956 | 0.150 | nd | nd | nd | nd | nd | nd |
| LaM-1 | 15,380 | 22 | N/A | N/A | – | N/A | N/A | N/A |
| LaM-2 | 15,151 | 0.49 | N/A | N/A | – | N/A | N/A | N/A |
| LaM-3 | 15,196 | 1.9 | N/A | N/A | + | N/A | N/A | N/A |
| LaM-4 | 14,866 | 0.18 | N/A | N/A | + | N/A | N/A | N/A |
| LaM-6 | 14,428 | 0.26 | N/A | N/A | – | N/A | N/A | N/A |
| LaM-8 | 14,666 | 63 | N/A | N/A | – | N/A | N/A | N/A |

and identified 16 nanobodies with very high affinity binding (≤50 nM). The $K_d$s of the six LaMs were consistently strong, ranging from 0.18 nM to 63 nM (Table 1).

***Specificity and efficacy of recombinant nanobodies***

We performed a variety of experiments to assess our nanobodies. Affinity isolations were performed on endogenous GFP- and mCherry-tagged proteins in yeast and human cells. All 25 positive LaGs were used for the isolation of GFP-tagged Nup84, a structural nuclear pore complex component, in budding yeast (Figure 2.5a) (Brohawn et al., 2009; Fernandez-Martinez et al., 2012). We plotted each LaG's observed $K_d$ against a quantification of either signal to background or yield from a Nup84-GFP affinity capture (Figure 2.5c,d and Table 1). Almost all LaGs were able to pull down detectable amounts of Nup84-GFP and its associated proteins, and many performed as well or better than either our best affinity-purified polyclonal antibodies (Cristea et al., 2005), or than the single commercially available GFP-Trap® anti-GFP nanobody (ChromoTek GmbH), which has a $K_d$ of 0.6 nM (Figure 2.5a,f) (Kirchhofer et al., 2010). When determining depletion of Nup84-GFP by Western blot, LaG-16, for instance, displays slightly higher yields than GFP-Trap®. Generally speaking, a strong correlation is seen between low $K_d$ and both high signal to background and high yield. This correlation is consistent with the relationship theoretically predicted for the percentage of the low abundance yeast target proteins bound in solution (Ghaemmaghami et al., 2003) (Figure 2.5c). Our ability to compare structurally similar nanobodies raised against a single antigen provides a unique opportunity to demonstrate the importance of very low $K_d$ to high quality antibody performance in this type of application. Even nanobodies with $K_d$s around 10 nM, typically considered high affinity for an antibody, start displaying a precipitous decline in affinity purification performance. These findings highlight the

Figure 2.5. Characterization of $V_HH$ IgG and recombinant nanobodies. (a,b) Affinity isolations of Nup84-GFP from *S. cerevisiae* (a) or RBM7-GFP from HeLa cells (b) using LaGs, GFP-Trap or polyclonal anti-GFP llama antibody (PC). Eluted proteins were analyzed by SDS-PAGE followed by MS. $K_d$ values for GFP are listed. MW, molecular weight. ND, not determined. Contaminant bands are identified by asterisks. [a]$K_d$ taken from published value (Kirchhofer et al., 2010). (c) Relative yields of affinity-isolated Nup84-GFP protein plotted against LaG *in vitro* affinities for GFP (green dots). Theoretical curves of the expected fraction of ligand bound to an immobilized binding partner at various $K_d$ values are shown for three hypothetical ligand concentrations (blue lines). (d) Signal-to-noise ratio of three Nup84-complex components plotted against each LaG's $K_d$. (e) Affinity isolation of mCherry-tagged histone H2B (Htb2) from S. cerevisiae by LaMs or RFP-Trap. Eluted proteins were analyzed by SDS-PAGE and identified by MS. Breakdown products of H2B are labeled ΔH2B-mCherry. Asterisk indicates LaM nanobody that leaked during the affinity-purification procedure. LaM lanes are labeled with the $K_d$ for mCherry. (f) Affinity isolations of yeast Nup84-GFP using the commercial nanobody GFP-Trap, polyclonal anti-GFP or a LaG-16–LaG-2 dimer with a glycine-rich peptide linker. The complex was isolated at various time points, and relative yield was determined by quantification of Coomassie-stained bands of known Nup84-complex components. Data are representative of two experiments (a–b,e) or are means from two experiments ± s.e.m. (c,d,f).

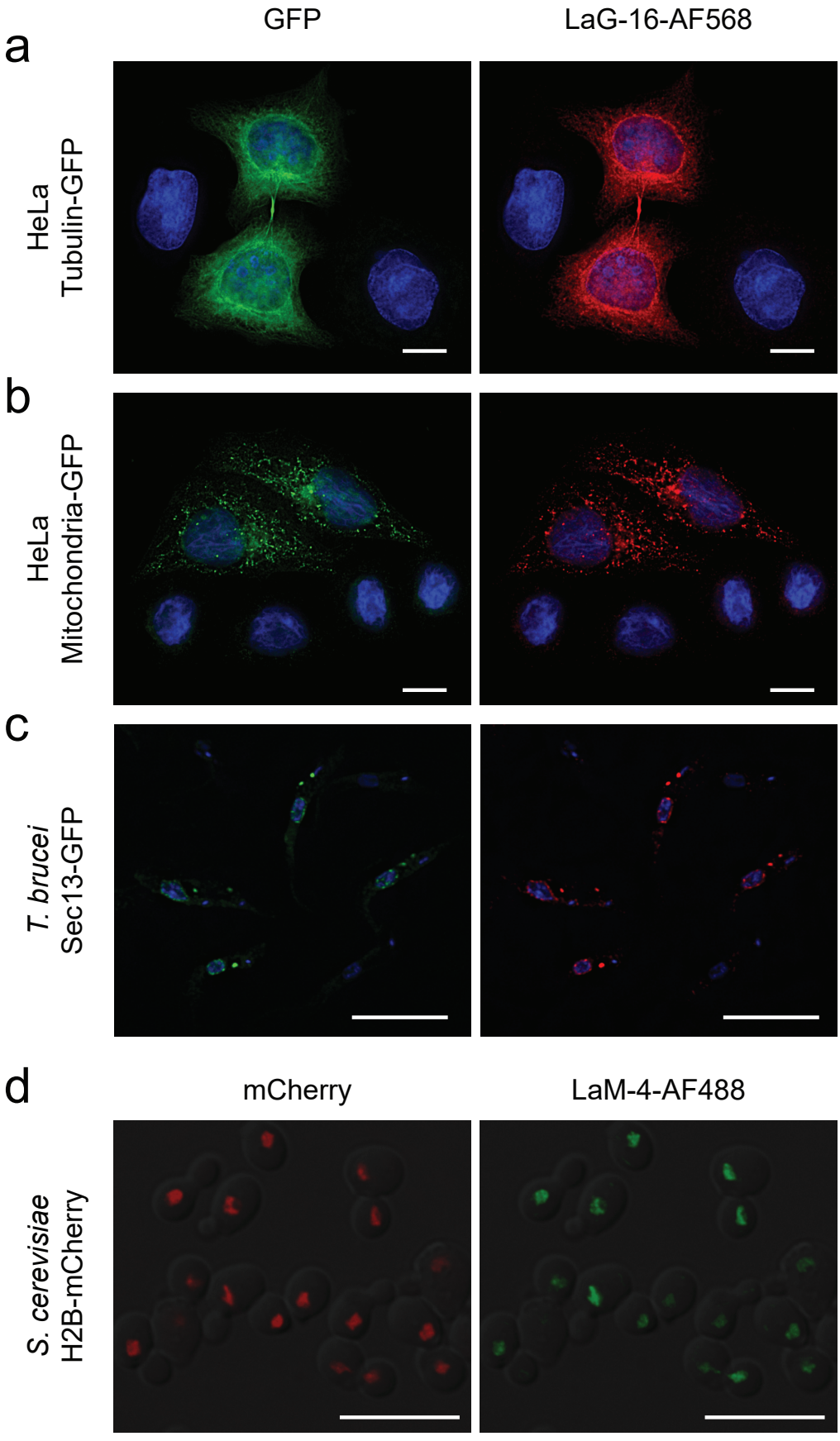

a
LaG
PC
GFP-Trap
41
16
9
24
43
35
37
6
11
188 kDa
98
62
49
38
28
Nup133
Nup84-GFP
Nup120
Nup85 and Nup145c
Seh1
Sec13
$K_d$ (nM): ND 0.23[a] 0.9 0.6 3.5 35 13 24 25 390 22,900
b
LaG
PC
24
9
16
41
188 kDa
98
62
49
38
28
SKIV2L2
ZCCHC8
RBM7-GFP
$K_d$ (nM): ND 35 3.5 0.6 0.9
c
1 nM
10 nM
100 nM
Relative Yield
1.00
0.75
0.50
0.25
0.00
$10^{-1}$ $10^{0}$ $10^{1}$ $10^{2}$ $10^{3}$ $10^{4}$ $10^{5}$
$K_d$ (nM)
d
Signal / Noise
1.00
0.75
0.50
0.25
0.00
$10^{-1}$ $10^{0}$ $10^{1}$ $10^{2}$ $10^{3}$ $10^{4}$ $10^{5}$
$K_d$ (nM)
e
LaM
RFP-Trap
1
2
3
4
6
8
188 kDa
98
62
49
38
28
17
14
Nap1
H2B-mCherry
ΔH2B-mCherry
H3
H2A
H4
$K_d$ (nM): ND 22 0.49 1.9 0.18 0.26 63
f
Relative Yield
1.0
0.5
0.0
1
10
100
Time (min.)
LaG16-LaG2
GFP-Trap
Polyclonal anti-GFP

importance of ultra-high affinity reagents, such as the nanobodies described here, for proteomic and interactomic studies.

Affinity capture experiments were also performed on GFP-tagged Rbm7, a component of the human nuclear exosome, from HeLa cells (Figure 2.5b) (Domanski et al., 2012), yielding performances comparable to those seen with Nup84-GFP. However, differences in the amount of contaminants were seen for certain LaGs, notably LaG-41, from purifications in yeast versus HeLa cells (Figure 2.5a,b). These results underscore how even high affinity reagents can give unpredictable background in certain cell types, demonstrating the utility of obtaining and testing large repertoires of such affinity reagents to improve the chances that at least one is likely to be optimal for any particular application. Similarly, Dynabead-conjugated LaMs were used to isolate mCherry-tagged histone H2B from yeast (Figure 2.5e). For all six LaMs tested, the core nucleosome complex was efficiently isolated, demonstrating the affinity and specificity of this second group of nanobodies. Consistent with the low $K_d$s of all the identified LaMs, the yield and specificity of all affinity isolations were similarly high. Commercial RFP-Trap® nanobody (ChromoTek GmbH) was tested in parallel, giving consistently lower yields.

Nanobodies are powerful new tools for fluorescence microscopy, both standard and super-resolution(Ries et al., 2012). We therefore tested the effectiveness of a selection of the LaG and LaM repertoire for immunofluorescence microscopy (Figure 2.6). As target proteins, we first made use of emGFP-tagged tubulin and mitochondria-targeted emGFP, transiently transfected into HeLa cells (Dolman et al., 2013). Fixed cells were stained with LaG-16 conjugated to Alexa Fluor® 568, giving specific and strong staining of GFP-tagged microtubule or mitochondrial structures respectively, with negligible non-specific staining of untransfected cells (Figure 2.6a,b). To demonstrate the versatility of these reagents, we also used them for immunofluorescence in a

Figure 2.6. Efficacy of LaG and LaM nanobodies in immunofluorescence microscopy. (a,b) HeLa cells transiently transfected with tubulin-EmGFP or an EmGFP-tagged mitochondrial marker (in green) were fixed and immunostained with LaG-16 conjugated to Alexa Fluor 568 (AF568, in red). Nuclei were counterstained with DAPI (blue). (**c**) *T. brucei* cells expressing EGFP-tagged Sec13 were mixed 1:1 with wild-type cells, fixed and stained with LaG-16–AF568, with DAPI counterstaining. (d) An *S. cerevisiae* strain with mCherry-tagged histone H2B was fixed, permeabilized and then directly stained with LaM-4 conjugated to Alexa Fluor 488 (AF488). Scale bars, 10 μm. Images are representative of at least three experiments.


GFP
LaG-16-AF568
a
HeLa
Tubulin-GFP
b
HeLa
Mitochondria-GFP
c
T. brucei
Sec13-GFP
mCherry
LaM-4-AF488
d
S. cerevisiae
H2B-mCherry

*Trypanosoma brucei* strain with eGFP-tagged Sec13. This protein localizes to both the nuclear pore complex and COPII-coated vesicles, and indeed the AF568-nanobody signal colocalized with GFP to give the expected nuclear rim and endoplasmic reticulum staining (DeGrasse et al., 2009) (Figure 2.6c). To determine if our anti-mCherry nanobodies were similarly well-suited for immunofluorescence microscopy, we conjugated LaM-4 to Alexa Fluor® 488 and stained *S. cerevisiae* expressing mCherry-tagged histone H2B; this also showed specific, colocalized nuclear staining (Figure 2.6d).

We also compared the fluorescence spectra of GFP in the presence or absence of various LaGs to look for spectral shifts upon binding, as have previously been reported, and observed moderate increases in fluorescence for several LaGs, with a maximum increase in fluorescence intensity of approximately 60% (Kirchhofer et al., 2010).

One additional question of specificity we sought to address was the ability of our nanobodies to recognize other fluorescent homologs of *Aequorea victoria* GFP and *Discosoma* mCherry. We tested the 13 highest affinity LaGs against a variety of fluorescent proteins: eGFP, two YFP variants, two CFP variants, BFP, mCherry, and DsRed (Figure 2.7a). None of these nanobodies bound DsRed or mCherry, two *Discosoma sp.*-derived proteins with low sequence identity to eGFP (<30%), or TurboYFP, derived from *Phialidium sp.*, which has 53% sequence identity to eGFP(Matz et al., 1999; Shagin et al., 2004; Shu et al., 2006). All bound standard *Aequorea victoria*-derived CFP, YFP, and BFP variants (>96% eGFP identity). Two LaGs did not bind a moderately divergent (78% eGFP identity) CFP sequence from *Aequorea macrodactyla*, while all others did (Xia et al., 2002). These results indicate that while identified LaGs bind specifically to fluorescent proteins with high identity to eGFP, differential binding activities can be obtained through selection of variants from other species. Our anti-mCherry LaM nanobodies bound to

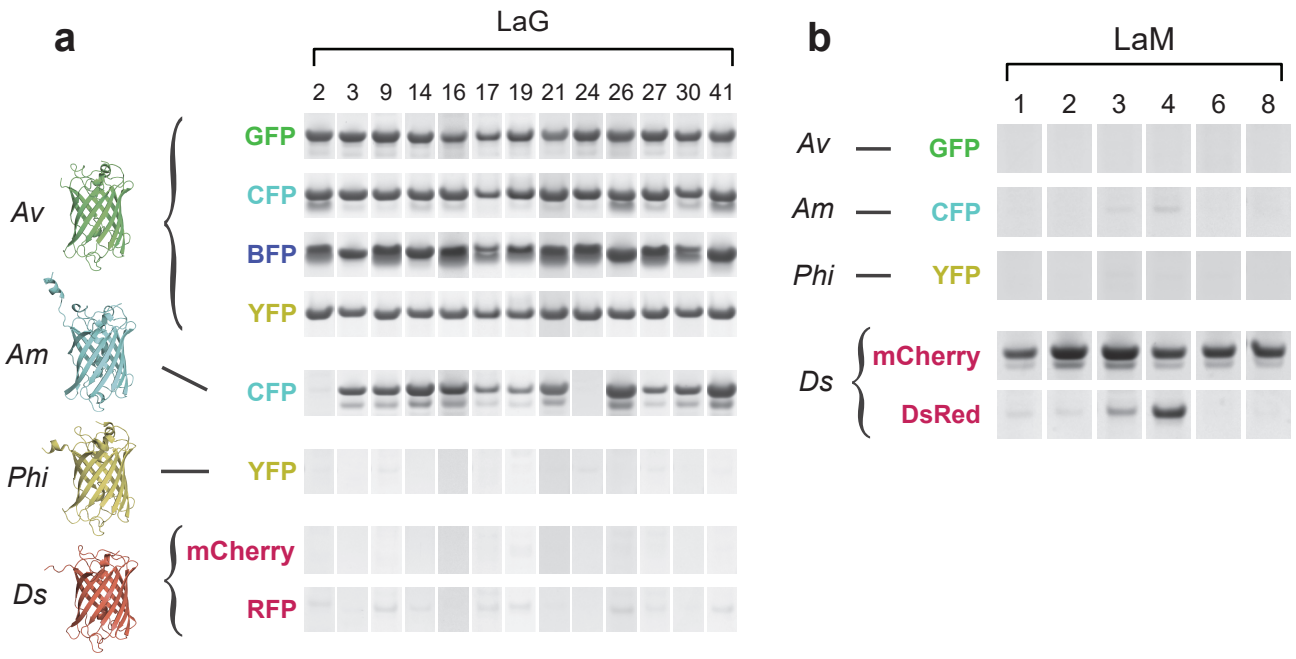



Figure 2.7. Nanobody fluorescent protein binding. (a,b) SDS-PAGE analysis of high-affinity LaGs (a) or LaMs (b) that were conjugated to magnetic beads and incubated with various recombinant fluorescent proteins: *A. victoria* (*Av*) GFP and its variants CFP, BFP and YFP; a cyan fluorescent protein derived from *A. macrodactyla* (*Am* CFP); a yellow fluorescent protein from *Phialidium* (*Phi* YFP); and mCherry and DsRed from *Discosoma* (*Ds*). Structural models were obtained from PDB 1EMA (*Av*) (Ormo et al., 1996), PDB 4HE4 (*Phi*) (Pletneva et al., 2013) and PDB 1GGX (*Ds*) (Wall et al., 2000); the *Am* CFP model is a Phyre server prediction (Xia et al., 2002; Kelley and Sternberg, 2009). Gels are representative of at least two experiments.

mCherry, but not to any form of GFP, YFP, or CFP tested (Figure 2.7b). Interestingly, two LaMs (LaM-3 and LaM-4) bound to standard DsRed, which has approximately 80% sequence identity to mCherry. Given the different fluorescent protein affinities observed with the LaG and LaM nanobodies, including specificity for *Am*CFP and DsRed, these reagents have diverse potential uses in differential labeling and affinity capture experiments from cells simultaneously expressing different fluorescently-tagged proteins.

***Mapping of the nanobody epitopes on GFP***

We identified the epitopes on GFP recognized by the twelve highest affinity LaGs using chemical shift perturbation, a well-established nuclear magnetic resonance (NMR) technique (with Ilona Nudelman). This method allows the mapping of binding sites on a protein by following changes in its characteristic "fingerprint" spectrum (typically the $^{15}$N-$^{1}$H HSQC) occurring as a result of adding an unlabeled ligand into a $^{15}$N-labeled protein sample (Goldflam et al., 2012).

Because previous studies have already made backbone $^{15}$N-$^{1}$H chemical shift assignments of the GFPuv variant (Georgescu et al., 2003; Khan et al., 2003) (closely related to standard eGFP with 97% sequence identity), we prepared $^{15}$N-labeled GFPuv, measured its $^{15}$N-$^{1}$H HSQC spectrum and obtained chemical shift assignments based on those published (Georgescu et al., 2003; Khan et al., 2003). We then tested complexes between 12 high affinity LaGs and $^{15}$N-labeled GFPuv and measured their $^{15}$N-$^{1}$H HSQC spectra. For 11 out of the 12 cases, we observed clear and specific changes in chemical shifts of a large percentage of cross-peaks compared to the $^{15}$N-$^{1}$H HSQC spectrum of GFPuv alone. In the 12$^{th}$ case, LaG-24, the nanobody did not bind the GFPuv variant.

A chemical shift difference was calculated for all spectra, and residues exhibiting a difference higher than 0.03 ppm were judged to be in the binding interface (Zuiderweg, 2002; Goldflam et

al., 2012). All the identified epitopes corresponded to large interfaces comprising more than 50 amino acids, consistent with the high affinity binding observed (Figure 2.8 and Table 1). The binding epitopes of the nanobodies can be divided into 3 distinct groups. The binding site of group I, containing 5 nanobodies (LaG-16, LaG-9, LaG-14, LaG-43 and LaG-17) overlaps with the binding site of group II, also containing 5 nanobodies (LaG-19, LaG-21, LaG-26, LaG-27 and LaG-41), whereas the two group III nanobodies (LaG-2 and LaG-24) exhibit a binding epitope on the opposite side of the GFP molecule compared to groups I and II. As a control, we also used this NMR approach to determine the GFPuv binding site of the commercial GFP-Trap® nanobody, the structure of whose complex with GFP has been crystallographically determined (PDB ID 3K1K)(Kirchhofer et al., 2010), and showed that the NMR-mapped epitope matched the published results (Battistutta et al., 2000; Kirchhofer et al., 2010). Comparing the binding epitopes of our nanobodies with that of GFP-Trap®, groups I and II show little or no overlap with the GFP-Trap® binding site, while group III, which binds on the same face of GFP, shows significant overlap (Figure 2.8).

***Generalizability of the pipeline***

Our optimized pipeline for the production and generation of nanobodies allows for the rapid generation of a large antibody repertoire against multiple epitopes in a chosen antigen. Notably, this approach identifies high affinity nanobody sequences directly from animal serum, taking advantage of the complex selection and maturation processes occurring in the animal's immune system, avoiding intermediary expression systems. The pipeline allows for the rapid production of a comprehensive repertoire of specific high affinity nanobodies for use in the characterization of target macromolecules, such as the GFP- and mCherry-tagged proteins shown here. The laboratory effort required after the collection of samples from llamas (50-70 days after initial immunization,

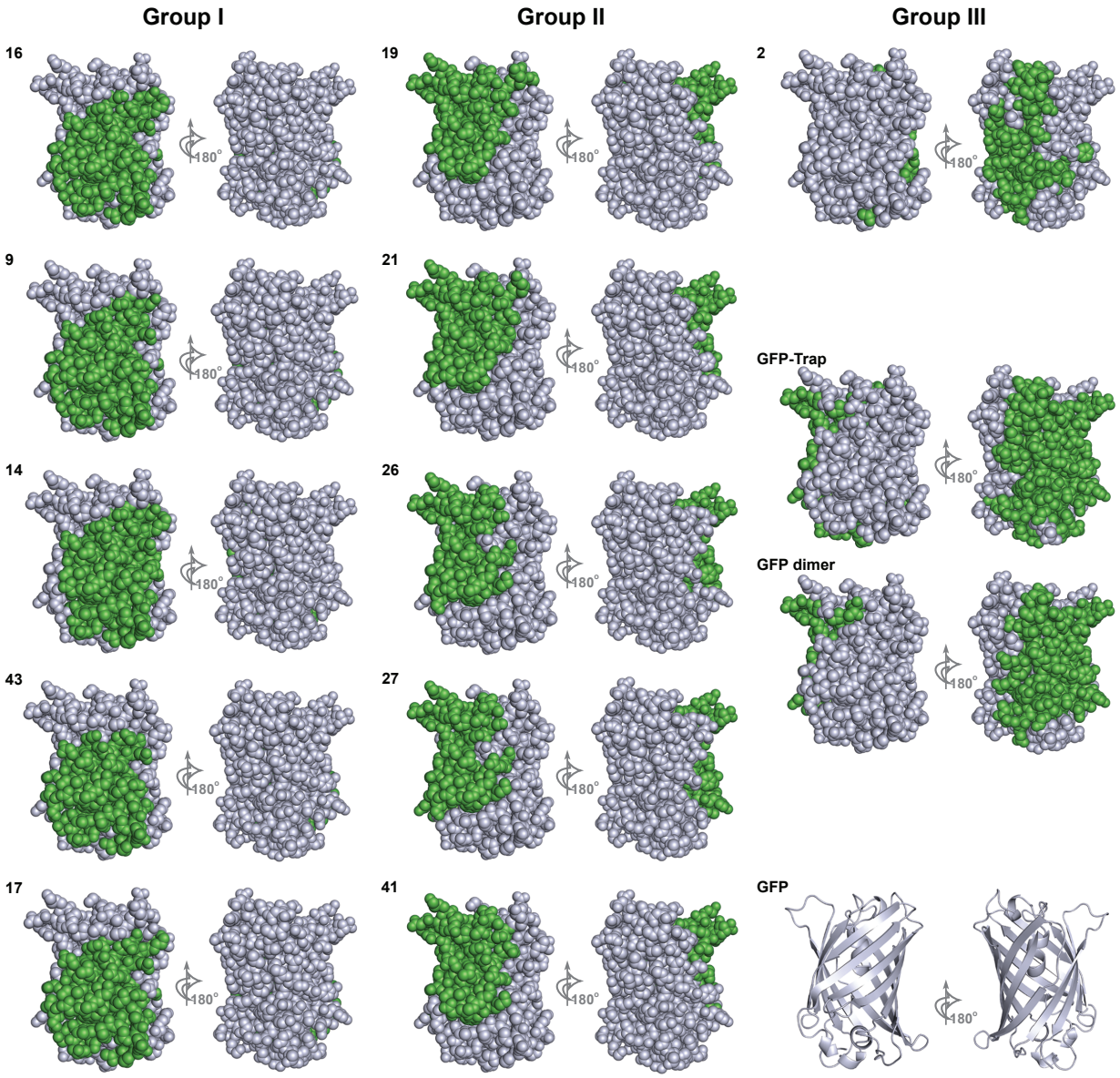



Figure 2.8. Mapping of nanobody binding epitopes on GFP by NMR. Binding epitopes of the 11 highest-affinity nanobodies on GFPuv are shown in three groups according to their location. For each nanobody, two opposite sides of GFPuv are shown (via a 180° rotation along a vertical axis), with the binding site of the respective nanobody colored green. All GFPuv molecules are represented in space-filling mode and have the same orientation in all panels. Maps below Group III in the right column show the GFP-Trap nanobody's binding epitope (top) and GFPuv's homodimerization interface (center). For reference, the ribbon diagram at bottom right depicts secondary structure elements of GFPuv, in the same orientation as other panels. NMR experiments and analysis were performed by Ilona Nudelman.

once an immune response is generated) is modest. The direct work required, including IgG purification (2 days), MS (2 days), cDNA generation and PCR (2 days), and final cloning and screening (3-6 days), can be performed over approximately 10 days. High-throughput sequencing and gene synthesis can be readily outsourced (as can animal handling and MS), and depending on turnaround times, each can typically be carried out in 1-2 weeks. The entire process can take as little as 4-6 weeks after an immune response is generated, with only standard techniques required in the primary laboratory. This is faster and more direct than other approaches available, which often require specialized high-throughput capability. Our approach puts the generation of large repertoires and quantities of high affinity single chain antibodies into the hands of the average researcher. Our LaM and LaG reagents, generated against the widely used GFP and mCherry tags, will be of immediate general use for the affinity isolation and enhanced visualization of these tags.

***Applications of anti-GFP nanobodies***

For the purposes of proteomics studies, the most immediate application of our highest performing nanobodies is affinity isolations of GFP-tagged proteins. Our lab has transitioned to anti-GFP nanobodies for most GFP-based experiments, with results equivalent to or better than those with polyclonal antibodies. Several collaborations have also led to improved affinity isolations using these nanobodies, for instance allowing improved analysis of the interactomes of supervillin and anillin, proteins regulating the actin and myosin cytoskeleton during cytokinesis (Smith et al., 2013).

We have also worked to develop new techniques using engineered nanobodies, taking advantage of the flexibility of these recombinant proteins. Specifically, we have mutated the high affinity LaG16 nanobody to remove both lysines, changing these to arginines instead (Shi et al., 2015). These residues are located away from the antigen binding interface, and the mutations did not

significantly affect the affinity or solubility of the nanobody (Figure 2.9). Removing lysines makes these nanobodies insensitive to commonly used amine-targeting crosslinkers like 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) or disuccinimidyl suberate (DSS), after chemical blocking of the amino terminus. Among other applications, crosslinking protein complexes with these types of reagents has emerged as an increasingly useful tool for obtaining structural information. Crosslinked peptides can be identified by MS, providing information on distances between crosslinked residues within a protein complex, revealing connectivity and aiding modeling of protein assemblies when integrated with other experimental data (Walzthoeni et al., 2013).

By developing nanobodies insensitive to crosslinkers, we were able to isolate complexes of structural interest and crosslink them on nanobody-conjugated beads. As the complex is not crosslinked to the immobilized nanobody, it can then be easily eluted for MS analysis identifying crosslinks (Shi et al., 2015). In comparison to alternative approaches, this is a simpler and more efficient sample processing approach, allowing effective cross-linking/MS of even small amounts of endogenous material. As a result, in addition to complexes like the yeast exosome or low abundance anaphase promoting complex, we were able to dissect the structures of assemblies from highly limited material, including tissue specific Beclin 1 from mouse liver. Our approach has thus allowed efficient, streamlined cross-linking analysis of endogenous protein complexes through the use of an engineered anti-GFP nanobody.

***Dimerized nanobodies as ultra-high affinity reagents***

Because NMR identified multiple epitopes for the 12 LaGs we characterized, we engineered heterodimers of LaGs with non-overlapping binding sites on GFP that could potentially bind with higher affinity, an approach that has been successfully used in various applications to develop high

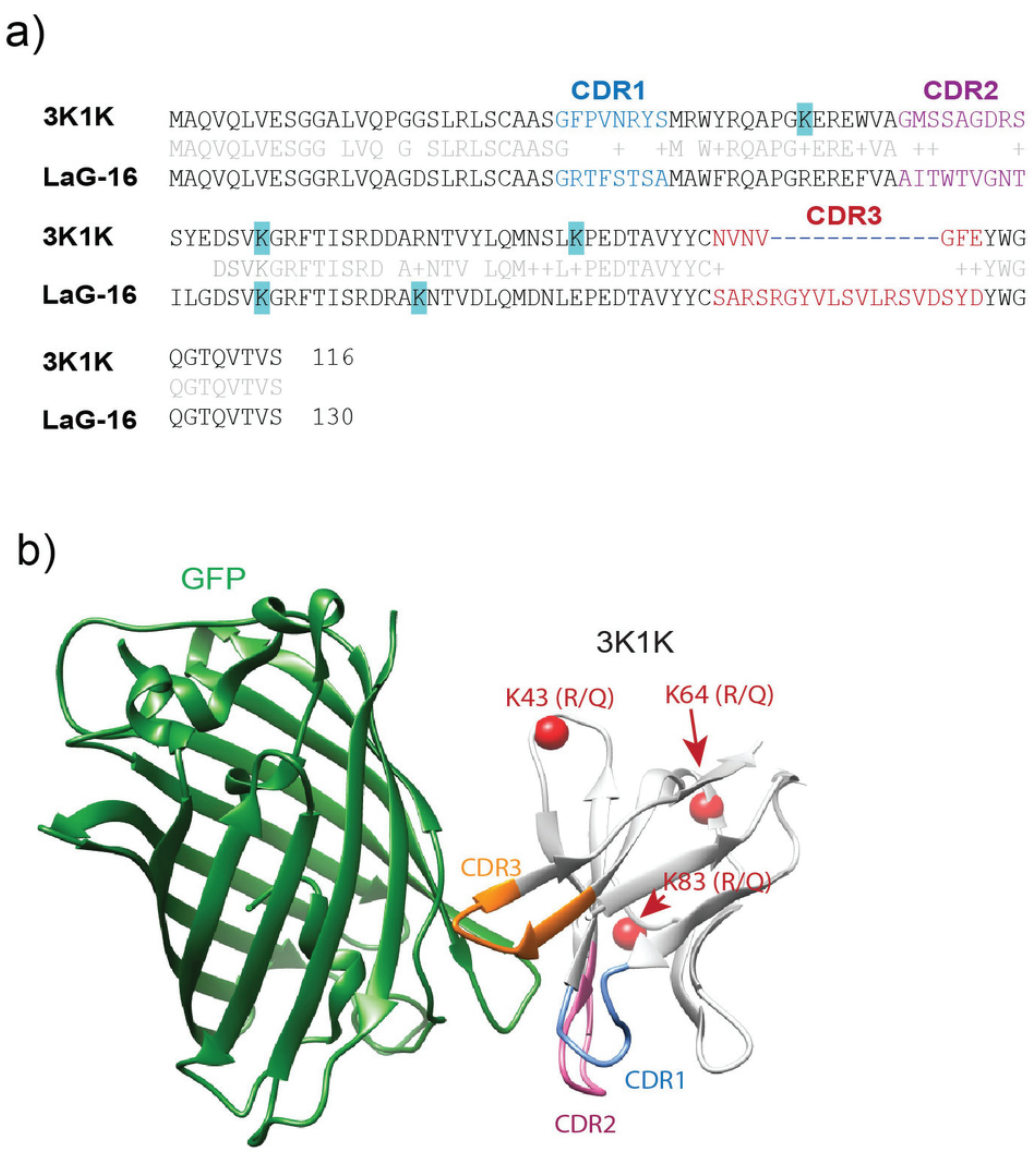



Figure 2.9. Lysineless mutants of anti-GFP nanobodies. (a) Protein sequence alignments of 3K1K and LaG-16. The lysine residues are highlighted in cyan. CDRs: complementarity determining regions. CDR1, CDR2 and CDR3 sequences are color coded in blue, pink and orange, respectively. (b) Localizations of the three lysine residues on the crystal structure of nanobody PDB 3K1K (Kirchhofer et al., 2010), which were mutated to arginine (R) or glutamine (Q) for the present work. The lysine residues are presented as red spheres.

avidity reagents (Neri et al., 1995; Silverman et al., 2005). A LaG16-LaG2 fusion with a flexible glycine-rich peptide linker (encoding three repeats of GGGGS) showed the highest affinity by SPR, with a $K_d$ of 36 pM. Dimers of other LaGs or with a different linker (a 3xFLAG tag), displayed $K_d$s in the range of 100-200 pM. We also sought to determine whether the higher affinity of these dimers could yield faster affinity isolations after conjugation to magnetic beads, compared to single nanobodies or polyclonal anti-GFP. We therefore performed time courses of yeast Nup84-GFP isolations and compared the relative yields of known Nup84 complex components. The LaG16-LaG2 dimer showed higher yields at earlier time points, reaching approximately 80% of maximum yield after only 5 minutes and 90% after 10 minutes (Figure 2.5f). These picomolar affinity reagents open the door for increasingly rapid affinity isolations, potentially allowing the capture of weakly or transiently associated complex components for interactome studies. In addition, their high avidity would allow for the detection of low abundance antigens, as is required for many diagnostic applications. As the LaG16-LaG2 dimer is the most effective affinity reagent we identified, we are now using this construct for the most demanding GFP isolations, such as those involved in mRNP purifications.

***Improvements to the nanobody pipeline***

Subsequent to our development of nanobodies against GFP and mCherry, we have made a variety of improvements to our general nanobody identification pipeline, and have succeeded in generating repertoires against several new antigens of biomedical interest. A full discussion of this additional work is included in the Appendix. In brief, we found that eliminating papain digestion of purified $V_HH$ IgG led to significantly less non-specific degradation of many antigens, and therefore provided much higher yields of material and more reliable MS results. As a

substitute, we determined that IdeS, a far more specific protease targeting IgG (von Pawel-Rammingen et al., 2002), could efficiently remove Fc fragments from isolated IgG.

A further improvement in the pipeline involved selection of llamas or alpacas more likely to produce the strong immune response we have found to be important in generating large nanobody repertoires. By pre-screening sera from candidate animals for activity against standard commercial vaccines, we determined that we could improve the chances of selecting immuno-responsive animals for antigen injections (Thompson et al., 2016). As immunization is the most costly step in antibody generation, this ability to screen out less productive animals early on can significantly improve the efficiency of our method.

We have also begun to apply these methodological improvements to the identification of new sets of nanobodies. Beginning with two cancer-associated proteins, CTLA4 (Leach et al., 1996) and EPHA2 (Oricchio et al., 2011), we have generated nanobodies against multiple novel antigens. By immunizing llamas with 10 or more antigens at once, it has also been possible to rapidly identify nanobodies against multiple targets in parallel, with no apparent reduction in efficiency. We are thus continuing to pursue nanobodies against a multitude of new proteins, including a large set of nucleoporins implicated in cancer (Simon and Rout, 2014).

Finally, in the process of our investigations, we identified strong affinity between a subset of our nanobodies and staphylococcal Protein A (PrA). This affinity was independent of CDR sequences, and appears to be a result of non-canonical binding of PrA to the backbone of certain Fab-like domains (Frenken et al., 2000). To take advantage of this, we engineered a minimal nanobody construct with strong PrA affinity, but with no other antigen-binding activity (Fridy et al., 2015).

This protein, along with an even higher affinity homodimer variant, has proven to be a highly effective PrA-targeting reagent, able to achieve results as good as the much larger and more heterogeneous IgG preparations typically used.

# Chapter 3: A method to tag and isolate single mRNP species from cells

### *Identifying a robust RNA tagging system*

As discussed above, a key element of this thesis is the development of an RNP purification technique capable of isolating a specific population of complexes associated with a single RNA. Many strategies for genetically tagging and purifying RNAs have been attempted, and we therefore first sought to compare a number of these established methods and determine the most promising approach. We began by surveying techniques based on anti-sense oligomers, as well as those involving bacteriophage coat protein / hairpin systems.

### *Anti-sense nucleic acids as capture reagents*

One of the simplest methods we assessed was the use of immobilized anti-sense oligomers to bind an exogenous, complementary RNA sequence. If this sequence were accessible in the context of a native mRNP complex, it should theoretically be able to bind virtually irreversibly to the antisense oligo by hybridization, allowing efficient capture on resin (Grunweller et al., 2003; Upadhyay et al., 2013). To assess this method, we designed a unique 18 bp sequence to be added to transcripts as either a series of repeats, or as a loop in a stem-loop structure, potentially aiding in the accessibility of the sequence. To maximize the stability of the RNA hybrid, we began with 2'O-methylated RNA as the immobilized oligomer, as this analogue forms significantly more stable hybrids with RNA than DNA (Inoue et al., 1987). These oligomers were synthesized with either a biotin or thiol group at the 5' end, allowing capture by streptavidin magnetic Dynabeads or covalent conjugation to epoxy-activated Dynabeads, respectively.

To test the binding ability of these immobilized oligomers, we first performed binding tests with *in vitro*-transcribed RNA transcripts and synthetic RNA oligomers. Using native buffer conditions routinely used for our RNP affinity purifications, these early tests showed positive but limited

RNA binding (Figure 3.1a). Notably, binding activity rapidly decreased as the target RNA length increased from an 18 bp complementary oligomer. A 110 bp tagged transcript showed very limited binding, while a similar 350 bp transcript had no detectable binding at all. Presumably, steric hindrance or interference from RNA secondary structure prevented efficient binding of longer RNAs in these conditions. Even heating binding reactions to 75°C, which would theoretically promote RNA denaturation and allow more efficient annealing for capture, did not allow binding of larger transcripts. Given the difficulties observed in these *in vitro* binding assays, it was assumed that this oligomer-based approach would not be readily useful in *ex vivo* RNP purifications, given the longer lengths of natural transcripts and the further complications likely in the presence of cell lysate. We also performed similar *in vitro* experiments with other types of nucleic acids to identify any oligomer-specific effects, but alternatives including RNA, DNA, or PNA all showed comparable results.

***Assessing the $\lambda_N$/BoxB system***

The most common approaches to tagging RNAs in cells make use of naturally occurring bacteriophage coat proteins that bind specifically to a cognate RNA hairpin structure. One example of this is the bacteriophage $\lambda_N$ protein, which includes a minimally necessary 22 amino acid peptide region that binds to the 15 nt BoxB hairpin sequence (Franklin, 1985). This small peptide has high affinity for its target RNA (about 22 nM) and has successfully been used in both *in vitro* and *in vivo* applications to tag or visualize RNA (Austin et al., 2002; Baron-Benhamou et al., 2004; Daigle and Ellenberg, 2007).

To determine if this system could be used for efficient mRNP affinity isolations, we first synthesized the $\lambda_N$ peptide and conjugated it to epoxy-activated Dynabeads. This immobilized peptide was capable of binding *in vitro*-transcribed RNAs with 4-16 copies of the BoxB hairpin

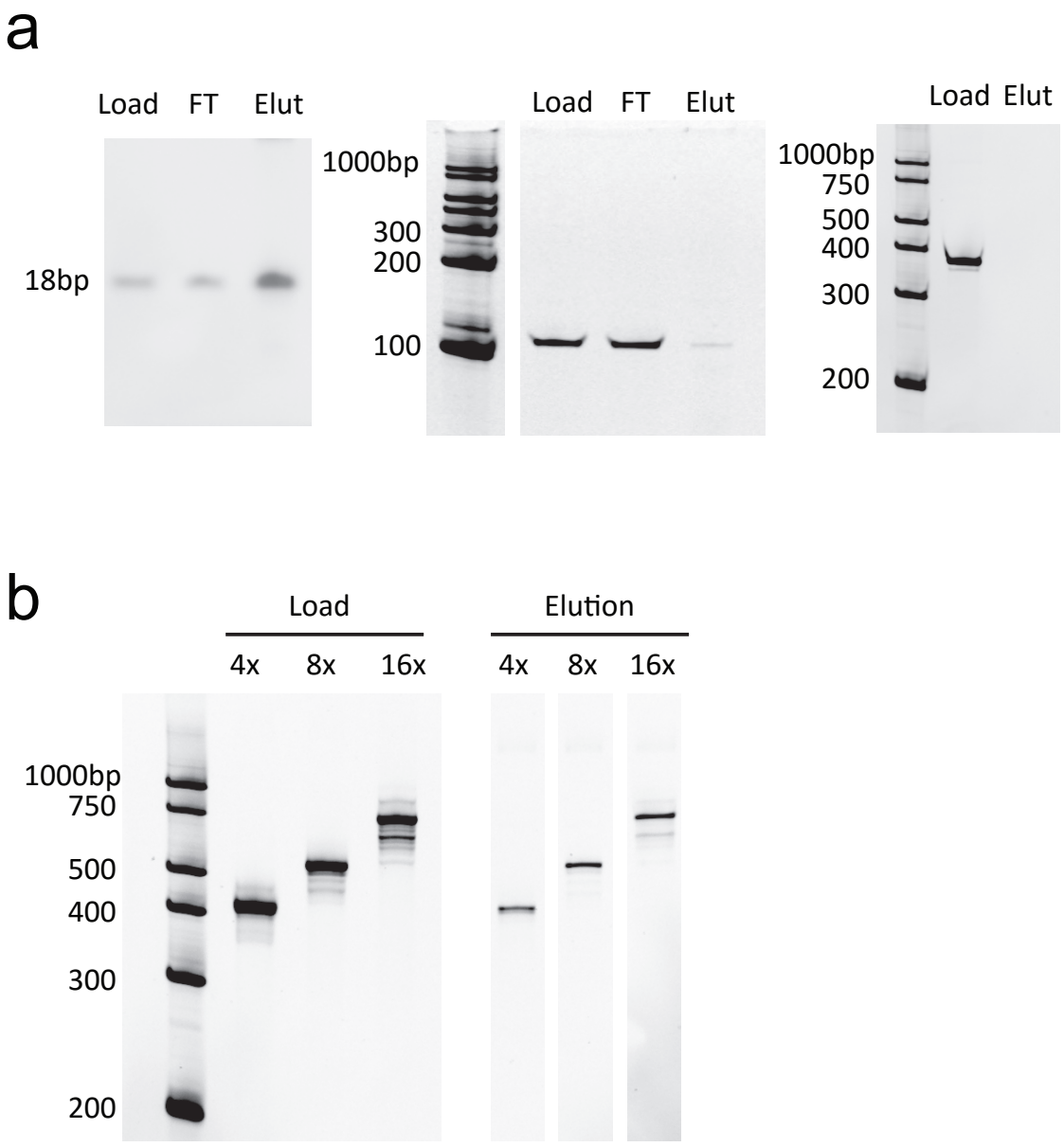



Figure 3.1. *In vitro* RNA binding to anti-sense oligos or BoxB peptide. (a) An 18 bp RNA oligomer (left), or 110 bp (middle) or 350 bp (right) *in vitro*-transcribed transcripts were incubated in excess with Dynabead-immobilized 2'O-methylated RNA oligomers with 18 bp of sequence complementarity. (b) *In vitro*-transcribed RNA transcripts with 4, 8, or 16 repeats of a BoxB hairpin were incubated in excess with Dynabeads coated with $\lambda_N$ peptide. In all experiments, bound RNA was eluted by heating in loading buffer, and run alongside load and flow-through (FT) samples on TBE-urea gels. RNA was visualized by SYBR gold staining.

sequence, even at transcript lengths up to 750 bp (Figure 3.1b). While the yield was not exceptional (<25%), these experiments did show substantially more promise for use in endogenous RNA isolations than immobilized oligomers, as transcripts of a typical mRNA's length could be recovered.

We next attempted to isolate RNA *ex vivo*, starting by tagging a model Protein A gene with BoxB hairpins (0, 4, 8, or 16 repeats) on a high copy plasmid, under a strong, constitutive TDH3 promoter, and transforming this into wild-type budding yeast. We then grew and prepared these cells for affinity isolations according to our lab's routine cryomilling strategy, and reconstituted frozen, milled cell material in a mild buffer known to stabilize mRNP complexes and that is compatible with their affinity capture (Oeffinger et al., 2007). After centrifuging to remove cell debris, soluble lysate was incubated with Dynabead-conjugated $\lambda_N$ peptide. These beads were then washed, and bound protein and RNA was eluted with hot LDS or phenol extraction, respectively. While some nonspecific protein binding was observed in the untagged negative control, enrichment of specific bands was in fact observed with 8xBoxB- or 16xBoxB-tagged transcripts (Figure 3.2a). Furthermore, qRT-PCR on input and elution samples revealed specific recovery of the PrA transcript when a BoxB tag was present. However, the yield of transcript was found to be at most 0.5%, even in the case of the 16xBoxB tag (Figure 3.2b). Thus, while the $\lambda_N$/BoxB system did seem capable of specific mRNP purification, the very low efficiency seen in these pilot experiments suggested that this would not be an immediately practical approach.

***Optimizing direct MS2 and PP7-based RNA isolations from yeast***

The most common approach to *in vivo* RNA visualization and targeting has been the MS2 bacteriophage coat protein system, or its close relative PP7 (Querido and Chartrand, 2008). This was therefore the most promising methodology for an RNA isolation technique, which we

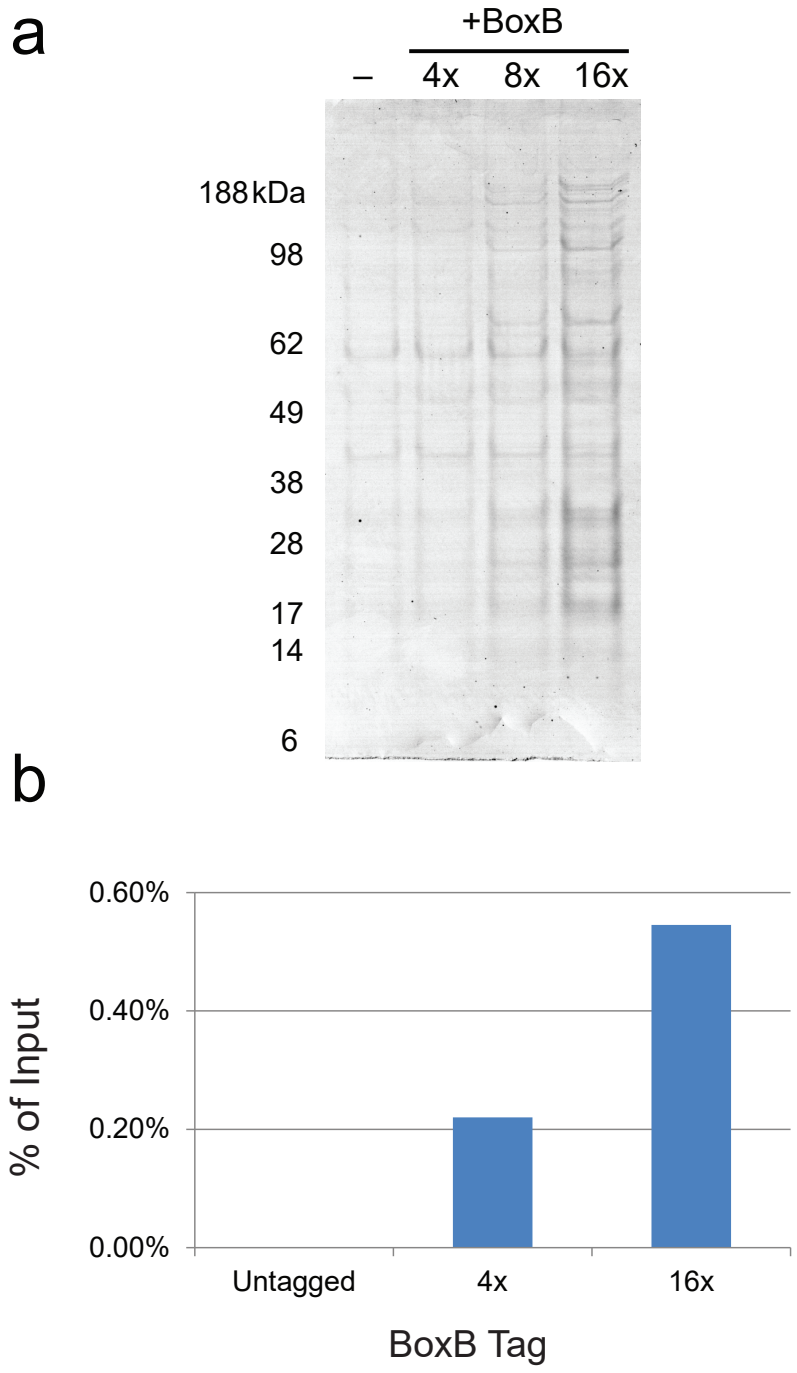



Figure 3.2. Affinity capture of BoxB-tagged mRNA from yeast. A Protein A transcript tagged with 0, 4, 8, or 16 repeats of the BoxB hairpin was expressed in yeast. Cells were cryomilled, and soluble lysate was incubated with Dynabeads conjugated to λN peptide. Dynabeads were washed and split; protein was eluted with hot SDS and RNA was eluted with hot phenol. (a) Protein elutions were separated by SDS-PAGE and silver-stained. (b) Levels of Protein A transcript in the RNA elutions were quantified by qRT-PCR, relative to the total amount in the input lysate.

investigated extensively. The MS2 and PP7 coat proteins (MS2CP or PP7CP) are both known to bind their corresponding ~20bp MS2/PP7 stem-loop RNA structures with high affinity (~0.5-1nM) and specificity ($10^6$-fold over other RNA sequences) (Carey et al., 1983; Lim et al., 2001). *In vivo*, this has allowed visualization of single RNAs through specific binding of hairpin tags by GFP-tagged MS2CP or PP7CP, as well as protein tethering to RNA via these same interactions (SenGupta et al., 1996; Bertrand et al., 1998; Hocine et al., 2013). To a limited extent, this has also been used to isolate abundant RNAs and their protein binding partners from cells (Hogg and Collins, 2007; Tsai et al., 2011). Given a high enough relative concentration of a hairpin-tagged RNA in cells, MS2CP or PP7CP should theoretically be able to provide a significant enrichment in a purification.

To assess this approach, we began by co-expressing MS2CP-PrA in yeast with a model transcript encoding GFP, with or without a tag of 6 MS2 hairpin repeats (6xMS2). We started with a GFP transcript as a simple, exogenous gene, one encoding an essentially inert protein to allow for nontoxic overexpression (Heim et al., 1994; Okabe et al., 1997). This is an important concern, as one of the main barriers to the isolation of single mRNAs is the low abundance of any one transcript. A number of microarray studies have suggested that transcript copy numbers are typically 1-2 per cell, going up to a maximum of a few hundred copies for the most highly expressed genes (Velculescu et al., 1997; Wodicka et al., 1997; Holstege et al., 1998; Arava et al., 2003). More recent single-cell studies, however, have suggested that true copy numbers may be as much as an order of magnitude higher, at least for weakly-expressed transcripts (Zenklusen et al., 2008; Larson et al., 2009). In either case, it can be conservatively estimated that 100 transcripts per cell may be observed for a strongly expressed gene. Identification of proteins by mass spectrometry requires a minimum protein concentration of 10-20 fmol present in the complex,

though in practice amounts in the hundreds of fmol or more are preferred (Geiger et al., 2010). Our cryolysis technique typically handles ~$1.2x10^{10}$ yeast cells/g powder, so if a model mRNA with ~100 copies per cell is selected, an associated protein would be present in a concentration of ~2 pmol/g. Thus, even with losses taken into account, a few grams of grindate should theoretically be sufficient to reliably detect most complex components.

In studies using similar approaches to directly isolate PrA-tagged RNPs, our lab has demonstrated the feasibility of isolating very low abundance proteins such as Sac3p, reported to have only a few hundred copies per cell (Ghaemmaghami et al., 2003; Oeffinger et al., 2007). Furthermore, expression levels of tagged RNAs can be significantly increased by using strong promoters like TDH3 (~400 transcripts/cell), and expressing off of high copy 2µ plasmids, which can themselves have 20-40 of copies per cell, or even 100-200 in some cases (Futcher and Cox, 1984; Lopes et al., 1989; Velculescu et al., 1997). As the total number of mRNA transcripts per cell is estimated to be between 15,000 and 60,000, and only makes up about 5% of all RNA transcripts, even this high level of overexpression should not significantly affect, much less overwhelm, the processing machinery (Hereford and Rosbash, 1977; Warner, 1999; Zenklusen et al., 2008).

Thus, our initial model transcript target was GFP on a 2µ plasmid (pRS426) under a TDH3 promoter, with or without a 6xMS2 tag (Figure 3.3a). MS2CP-PrA was co-expressed on a centromeric plasmid (pRS414) under a moderately strong inducible MET3 promoter, and with an SV40 nuclear localization signal (NLS). This NLS localizes MS2 to the nucleus to allow more efficient association with nascent transcripts, and has been shown to have no significant effect on subsequent nucleocytoplasmic transport of bound RNAs (Grunwald and Singer, 2010). Cells co-expressing these MS2CP-PrA and GFP constructs were grown and cryomilled, and we performed single step affinity purifications of MS2CP-PrA using IgG Dynabeads. With low level MS2CP

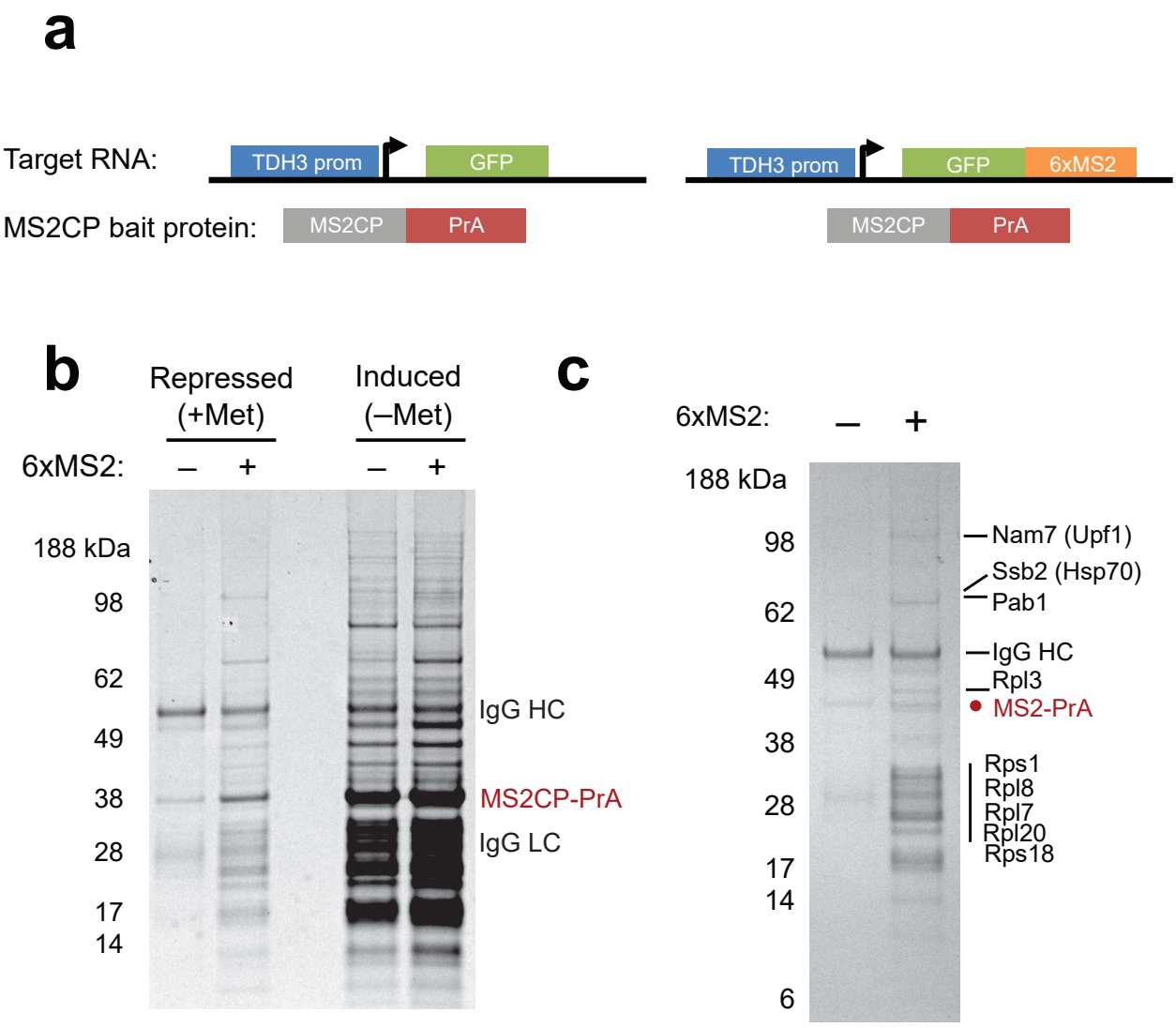



Figure 3.3. Affinity isolation of MS2-tagged RNA using MS2CP-PrA. (a) GFP transcripts with or without a six MS2 hairpin tag were co-expressed in yeast with an MS2CP-Protein A fusion protein. MS2CP-PrA was expressed under a MET3 promoter, either repressed (+MET) or induced (–MET). MS2CP-PrA was purified from cell lysate using Dynabeads coated with rabbit IgG. (b) Co-isolating mRNP proteins were eluted using ammonium hydroxide, separated by SDS-PAGE, and stained with Coomassie Blue. IgG heavy chain (HC) and light chain (LC) leakage is labeled. (c) Indicated bands from a MS2CP-GFP (+MET) purification were excised and analyzed by MALDI MS. Bands are labeled with the corresponding protein ID.

expression (MET3 repressed by growth in methionine), a number of protein bands copurified only in the presence of the tagged transcript, with little background observed in the untagged negative control (Figure 3.3b,c). By qRT-PCR analysis, 14% of the tagged GFP transcript was found in the final elution (Figure 3.4a). By contrast, under high level MS2CP expression (no methionine), this yield went up to 22%, but was also accompanied by a significant increase in protein contaminants in the untagged control sample (Figures 3.3b and 3.4a). Given potential inefficiencies in the recovery and quantification of the small amounts of RNA in these elution samples, assessing depletion of the GFP transcript from input to IgG bead flow-through may be a more accurate measure of yield; in this case yields could be as high as 30% or 65% with high or low MS2CP expression, respectively.

To identify the major proteins copurified with the 6xMS2-tagged GFP transcript, we performed MALDI-MS on the prominent bands specific to that affinity purification (Figure 3.3c). This identified proteins consistent with expected compositions of a typical RNP population. In addition to multiple ribosomal proteins, most likely associated with RNAs undergoing translation, Pab1p was identified, a general polyA RNA binding protein, as well as Nam7p, characteristic of mRNA surveillance for cytoplasmic nonsense mediated decay (Sheth and Parker, 2006). Thus, the major proteins identified in these mRNP isolations reflect the most highly abundant factors known to associate with cytoplasmic mRNAs undergoing translation. This is consistent with reports that up to 70% of the total mRNA population is cytoplasmic and associated with ribosomes, indicative of rapid mRNA export and a relatively small pool of nuclear mRNA (Arava et al., 2003).

To determine if the yield and purity of this single-step MS2 purification could be improved, we also tested a variety of modifications to the above approach. First, we attempted to avoid co-expression of MS2CP-GFP in cells by using recombinant MS2CP immobilized to beads as a bait

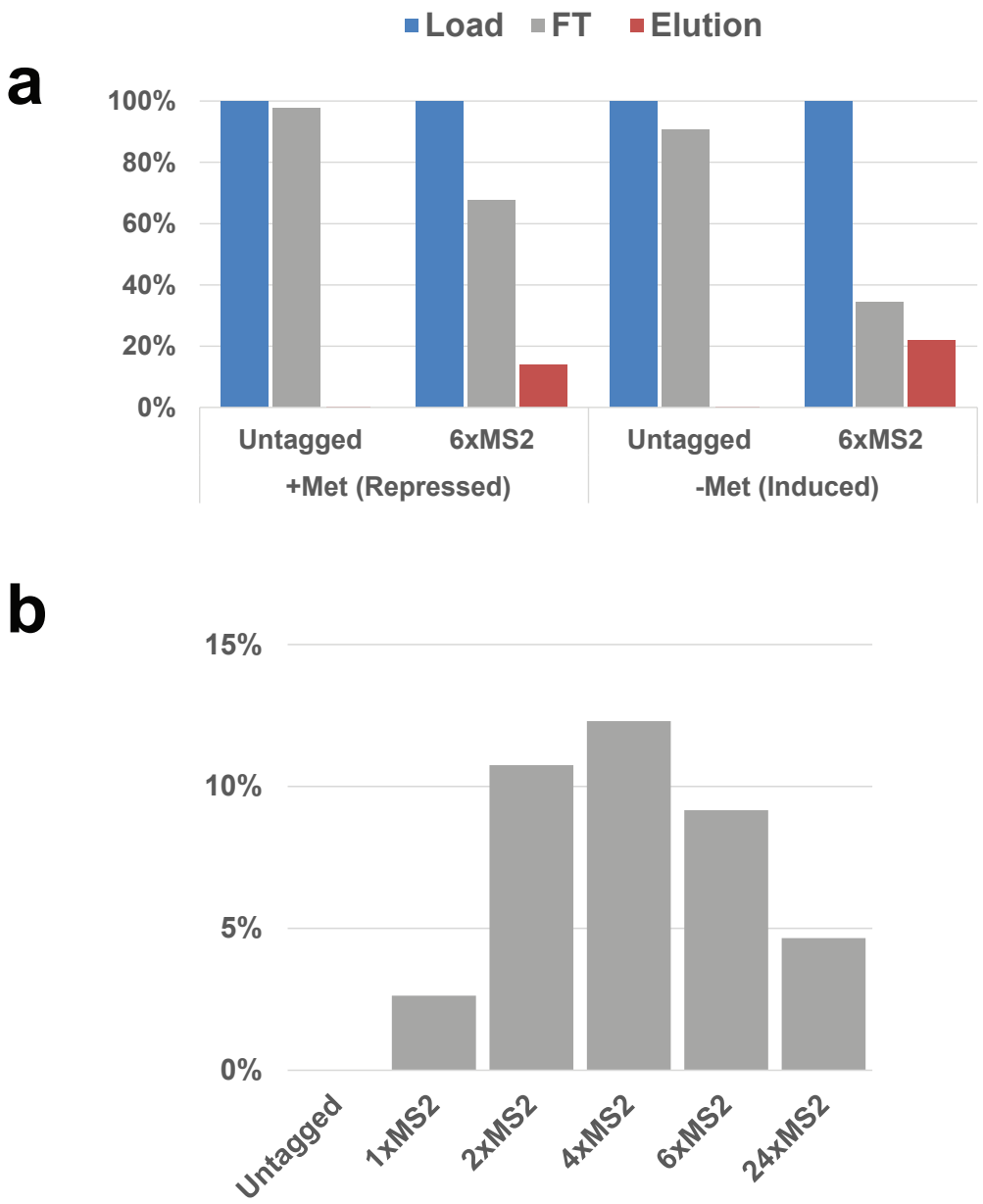



Figure 3.4. qRT-PCR quantification of MS2-tagged transcript recovery. (a) RNA samples from load, flow-through (FT), and elutions of MS2CP-PrA-purified material (Figure 4.3) were used to quantify relative levels of the target GFP transcript using qRT-PCR. Levels of untagged GFP or tagged GFP-6xMS2 transcript relative to the lysate load sample are shown. Experiments were performed with either low level (repressed) or high level (induced) expression of MS2CP-PrA. (b) Relative yields of GFP transcripts tagged with 0-24 repeats of MS2 were determined as in (a), with low level MS2CP-PrA expression.

for MS2-tagged transcripts. Alternatively, soluble MS2CP tagged with PrA was mixed with cell lysate, then purified on IgG Dynabeads. In both cases however, significant nonspecific binding was observed in negative control samples lacking an MS2-tagged transcript. It is likely that the relative concentration of M2SCP protein in these conditions was so high that extensive off-target RNA binding was occurring, expected in extreme cases given MS2CP propensity to bind very weakly to any RNA (Carey et al., 1983). Furthermore, there did not appear to be a point to which recombinant MS2CP could be titrated to allow sufficient recovery of specific MS2-associated material while avoiding excessive background binding of contaminants. Presumably, co-expressing MS2CP in cells allows more time for it to associate with tagged mRNAs as they are transcribed and processed in the nucleus, which makes it possible to use the relatively low levels of the protein necessary to avoid non-specific binding.

In theory, the MS2 tag's hairpins are saturated in the higher expression conditions used. Under a MET3 promoter on a centromeric plasmid (~8 copies per cell), MS2CP could be expected to be present at approximately 3,200 copies per cell (Newman et al., 2006). Given an abundant transcript with four MS2 hairpins, at approximately 400 copies in a typical yeast cell of 41 $\mu m^3$ volume, MS2CP and mRNA concentrations would be 128 nM and 64 nM respectively. Assuming 30% of mRNA and all NLS-containing MS2CP is in the nucleus (~2.5 $\mu m^3$), nuclear concentrations could be estimated at 2,100 nM MS2CP and 315 nM RNA (Jorgensen et al., 2007). With MS2CP binding as a dimer, >99% of MS2 hairpins would be bound, and ~85% of MS2CP-GFP would be free. In practice, this led to efficient isolation of tagged RNA, but significant background binding. Under low MS2CP expression, MET3 repression is predicted to be ~16-fold lower (Jorgensen et al., 2007), giving a theoretical MS2 hairpin saturation of ~20%, and no

unbound MS2CP. This is also consistent with our results, with low background but low yield of tagged mRNA.

As a final comparison, we also attempted equivalent affinity isolations using PP7 coat protein (PP7CP) and its corresponding PP7 hairpin tag in place of MS2. When either coexpressing PP7CP-GFP or using recombinant protein as bait, the results with this alternative bacteriophage system were essentially equivalent to MS2. After assessing the different hairpin types, using different numbers of hairpins in the RNA tag, we ultimately found that MS2 allowed slightly more efficient isolations, with an optimal tag length of four hairpin repeats (Figure 3.4b).

***Sequential protein-RNA affinity purifications***

While the direct purification of 6xMS2-tagged RNA by MS2CP-PrA was successful in the specific isolation of mRNAs at a yield consistent with past studies (Hogg and Goff, 2010), these experiments revealed certain inherent difficulties with this one-step approach. Given the much larger pool of mRNA in the cytoplasm compared to the nucleus, no matter how effective the isolation of total mRNA, it is likely that analysis of lower abundance nuclear proteins would be hampered by the far more abundant cytoplasmic mRNP factors. While more sensitive mass spectrometry is capable of detecting far lower abundance proteins than our initial analysis, the high complexity of these samples and the far higher abundance of cytoplasmic factors would nonetheless make robust analysis and quantitative comparison of different samples extremely difficult.

Additionally, it proved difficult to achieve mRNP purifications at both high yield and high specificity. While MS2CP expression levels could be titered to achieve specific isolations, RNA yield in these conditions remained in the 14-30% range. For robust analysis of different mRNPs,

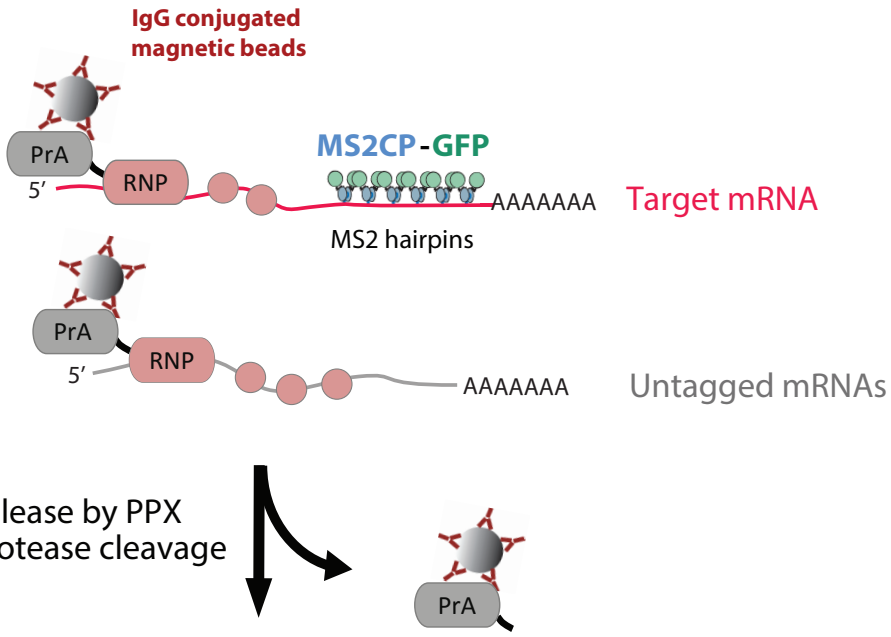
particularly when the associated protein factors of interest are at relatively low abundance, this compromise in yield would be a significant hindrance. We believe this is an inherent limitation of the MS2 system, as MS2CP's nonspecific affinity for RNA, while very weak, will lead to excessive off-target binding if MS2CP's expression level is brought high enough to capture MS2-tagged RNA at an acceptable yield.

To address the dual concerns of non-specific MS2CP binding and unwanted purification of high abundance cytoplasmic mRNPs, we incorporated a second purification step into our protocol. While various orthogonal purifications approaches could potentially help address these issues, we found the most promising option to be isolating a known PrA-tagged mRNP protein in an initial purification step, before using GFP-tagged MS2 to isolate the specific mRNA subpopulation of interest (Figure 3.5). First, this additional purification step provides a significant enrichment in purity simply by providing an independent, highly efficient enrichment of an mRNP population. Our lab has successfully used this approach to routinely isolate dozens of tagged mRNPs at high purity and >90% yield (Oeffinger et al., 2007). Beyond this enhancement of purity, the use of an mRNP protein as a second tagged target has the additional advantage of allowing the isolation of mRNP complexes at defined processing stages; as a great deal is known of the basic functions and associated factors of many nuclear mRNP proteins, it is possible to select a number of such proteins with distinct, well-established roles in mRNP processing as targets. This allows us to isolate a chosen subset of the mRNP processing machinery, before using the MS2 tag to purify the specific mRNA of interest, providing high enrichment at a particular maturation step defined by the tagged mRNA's association with a given mRNA processing factor.

As a starting point, for instance, tagged Nab2p can be used to isolate a large pool of transcripts at steps ranging from transcription to export, but largely excluding the cytoplasmic population

Figure 3.5. Outline of sequential RNP and MS2 RNA isolation method.

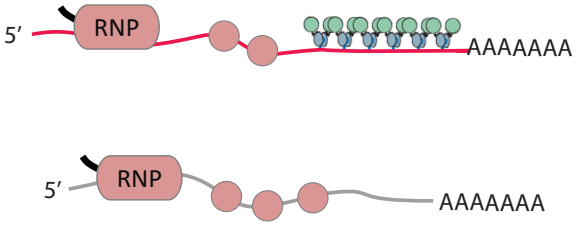
**Step 1:**
Purification of PrA-tagged RNP and all associated mRNAs, untagged or MS2-tagged

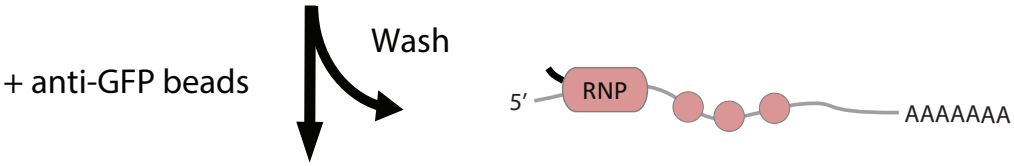




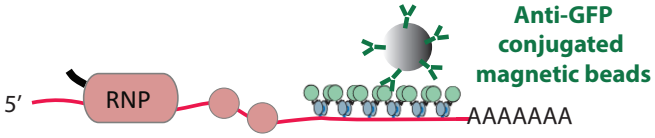


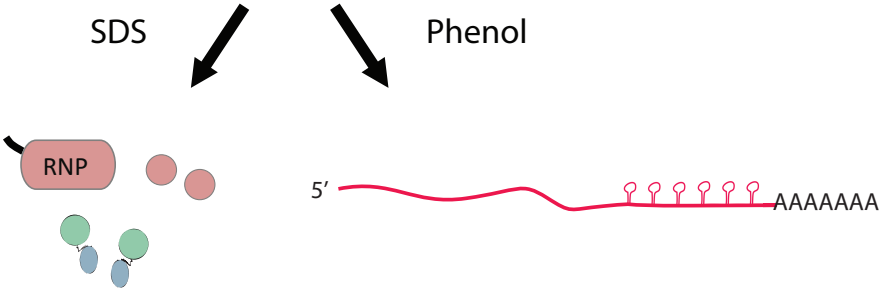
**Step 2:**
MS2CP-GFP-based purification of MS2- tagged mRNA




**Step 3:**
Denaturing elution of tagged transcript's RNPs or RNA



(Batisse et al., 2009; Schmid et al., 2015). Similarly, Cbp80p would be expected to associate with virtually all nascent mRNAs, remaining bound up until the first round of translation (Lewis and Izaurralde, 1997). Conversely, Cdc33/eIF4E, part of the cytoplasmic cap-binding complex, would pre-enrich for cytoplasmic mRNPs (Gingras et al., 1999; Fortes et al., 2000).

Starting with these general mRNA-binding factors, we performed preliminary sequential purifications using MS2-tagged mCherry transcript. Like GFP, this fluorescent protein was chosen for its ease of overexpression and detection. In theory, either the PrA-tagged RNP or MS2CP-GFP could be purified first in our method; for practicality, we chose to purify the RNP first, as IgG is more readily available in large quantities, and the first step must necessarily be done at a larger scale. Additionally, PrA would be likely to bind anti-GFP antibodies given its general IgG affinity, making it better to remove this tag first. To allow a second, independent purification, we also needed a means of releasing bound protein from the first affinity isolation. We chose to do this by inserting a PreScission protease cleavage site upstream of the C-terminal PrA tag. This protease (a derivative of the rhinovirus 3C protease) is highly active even at low temperatures, allowing the RNP protein to be cleaved from the tag and bound IgG resin in under an hour at 4°C (Cordingley et al., 1989). The released material can then be directly incubated with anti-GFP Dynabeads, with PrA no longer present. To start with, we coexpressed MS2CP-GFP under an unrepressed MET3 promoter on a centromeric plasmid, conditions which led to high yield but non-specific single-step isolation of tagged RNA.

In preliminary pullouts using this sequential method, purifications of transcripts tagged with MS2 isolated significant amounts of protein, with a banding pattern similar to an equivalent single-step isolation of the corresponding tagged RNP (Figure 3.6). In contrast, negative controls with untagged transcripts had essentially no associated protein, indicating very little non-specific

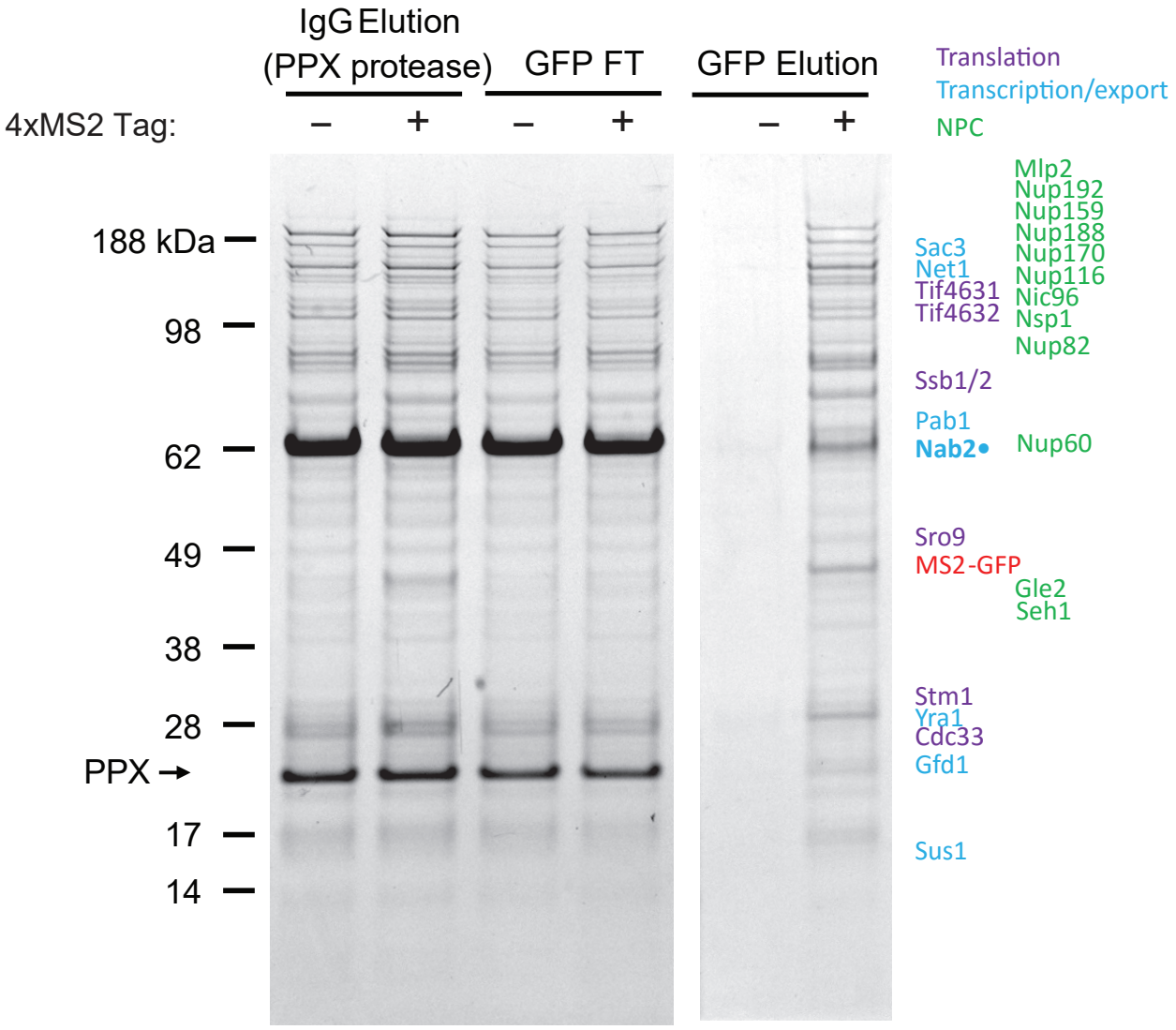


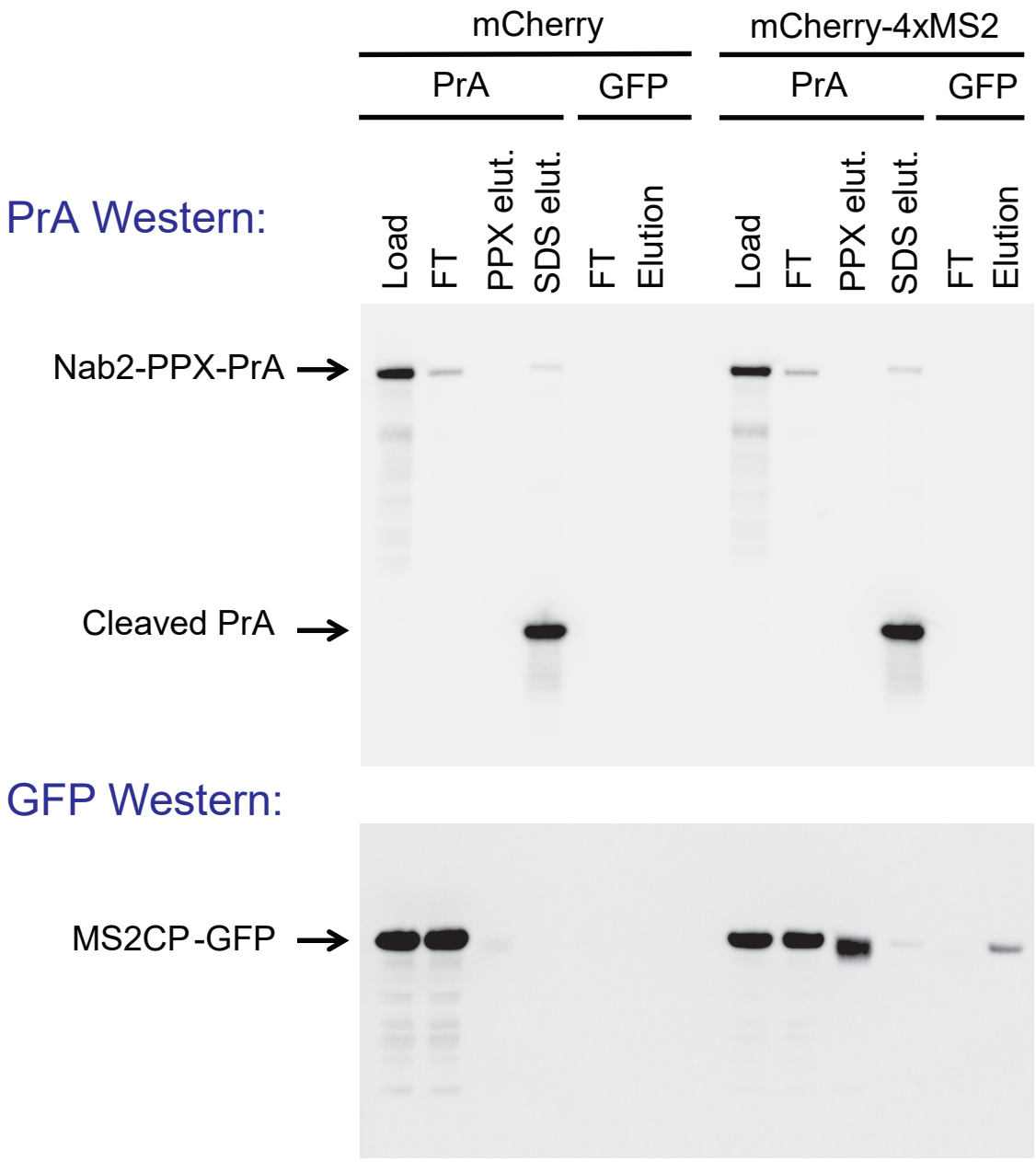

Figure 3.6. Sequential Nab2 RNP and RNA purification of mCherry-4xMS2. mCherry transcripts with or without a 4xMS2 hairpin tag were co-expressed with MS2CP-GFP in yeast and purified in two steps with Nab2-PrA and MS2CP-GFP. Samples of PPX protease elutions from IgG beads (step 1), and flow-throughs (FT) and elutions from anti-GFP beads (step 2) were separated by SDS-PAGE and stained with Coomassie Blue. RNPs are only co-purified in the presence of the 4xMS2 tag. In parallel, equivalent samples were directly eluted from anti-GFP beads with ammonium hydroxide and analyzed by MS. RNP protein IDs from that analysis are listed, categorized according to function.

binding due to MS2CP. This indicates that even in conditions with saturating levels of MS2CP, the addition of the PrA purification step essentially eliminates the non-specific binding previously observed. As we expect the tagged model mCherry transcript to behave as a typical mRNA, it was not surprising that the final protein isolated appeared generally similar to the mixed RNP population observed in a single step RNP purification. However, the relative intensities of a number of bands do in fact appear to differ in the mRNA-specific isolation, and fewer super-stoichiometric bands are present, indicating differences in the purified RNP pools, and in particular less free protein.

***Assessing efficiency of sequential isolations***

To determine the efficiency of each purification step at the protein level, we did Western blotting of each tag (PrA and GFP) across the procedure (Figure 3.7). Depending on the cellular abundance of the PrA-tagged RNP and amount of IgG Dynabeads used, the yield of the PrA isolation was as high as 90%. Judging from the amount of uncleaved PrA-RNP protein remaining on the IgG beads, the efficiency of the cleavage step was highly variable from protein to protein, but generally above 50%. We have observed this type of variability in many protein complexes studied in the lab, and suspect it to be due to protein-specific differences in the accessibility of the cleavage site, or other structural influences. Similar variability is found in the propensity of complexes to stick to the Dynabeads even after tag cleavage, and these issues have shown little dependence on the isolation conditions used. When following GFP signal, MS2CP-GFP was detected in the IgG bead elution only when a tagged transcript was used. This protein was then depleted at high efficiency by anti-GFP Dynabeads. As expected, less than 5% of the total MS2CP-GFP in the lysate was depleted in the initial PrA purification; this is consistent with the small fraction of total mRNA predicted to be associated with any one RNP factor, as well as the excess of MS2CP-GFP present

Figure 3.7. PrA and GFP yield of sequential Nab2-PrA and MS2CP-GFP purification of mCherry-4xMS2. Samples from a sequential purification of untagged or 4xMS2-tagged mCherry were analyzed by Western blot, probing either PrA or GFP. 15 times less material was loaded for PrA Load and flow-through (FT) samples. More than 80% of Nab2-PrA was recovered in the PrA purification step, and more than 90% of this was cleaved by PPX, as judged by cleaved vs uncleaved PrA remaining on beads after protease elution. MS2CP-GFP is recovered after both purification steps only when the mCherry transcript contains a 4xMS2 tag. Due to both free MS2CP-GFP and the mCherry-4xMS2 population not associated with Nab2, only a small fraction of MSCP-GFP is co-purified with Nab2-PrA.


mCherry
mCherry-4xMS2
PrA
GFP
PrA
GFP
Load
FT
PPX elut.
SDS elut.
FT
Elution
Load
FT
PPX elut.
SDS elut.
FT
Elution
PrA Western:
Nab2-PPX-PrA
Cleaved PrA
GFP Western:
MS2CP-GFP

in these expression conditions. The relative amounts of GFP and total protein obtained in these purifications were also dependent on the tagged RNP protein used. Consistent with the higher abundance of cytoplasmic versus nuclear mRNA, Cdc33 isolations had significantly more material than those of Cbp80 or Nab2.

We also performed MS analysis on the material purified from these initial Nab2, Cbp80, and Cdc33-based RNA purifications (Figures 3.6 and 3.8). As expected, the proteins identified were essentially the same as those previously identified from purifications of these RNP factors themselves (Oeffinger et al., 2007). For Cbp80-based isolations, this includes major RNA binding proteins like Pab1p and Yra1p, cap complex components Cbp80p and Cbp20p, members of the THO/TREX complex, and spliceosomal factors. For Cdc33, we identified expected cytoplasmic factors like members of the cytoplasmic cap (eIF4E and eIF4GA), Pab1p, and translation initiation factors. For Nab2, we identified Pab1p, a variety of export factors like Yra1p and members of the TREX-2 complex, as well as a number of nuclear pore complex (NPC) components, consistent with transcripts actively undergoing export. Additional quantitative analysis would be necessary to identify transcript-specific RNP factors from these purifications (Chapter 5).

The purity and yield of MS2-tagged transcript was also assessed from these pilot affinity purifications. Quantitative real-time RT-PCR (qRT-PCR) was also performed on various samples from these isolations (Figure 3.9a). In the case of Cbp80-PrA purifications, this analysis showed a recovery of approximately 20% of mCherry transcript in the first purification step, with or without a tag present. This is broadly consistent with the amount of bulk mRNA expected to be capped and not undergoing translation. In the second MS2CP-GFP purification step, ~75% of this material is recovered, only in the presence of the 4xMS2 hairpin tag. Thus, MS2-tagged transcripts do appear to be efficiently and specifically recovered by this technique. As a control, 18S rRNA

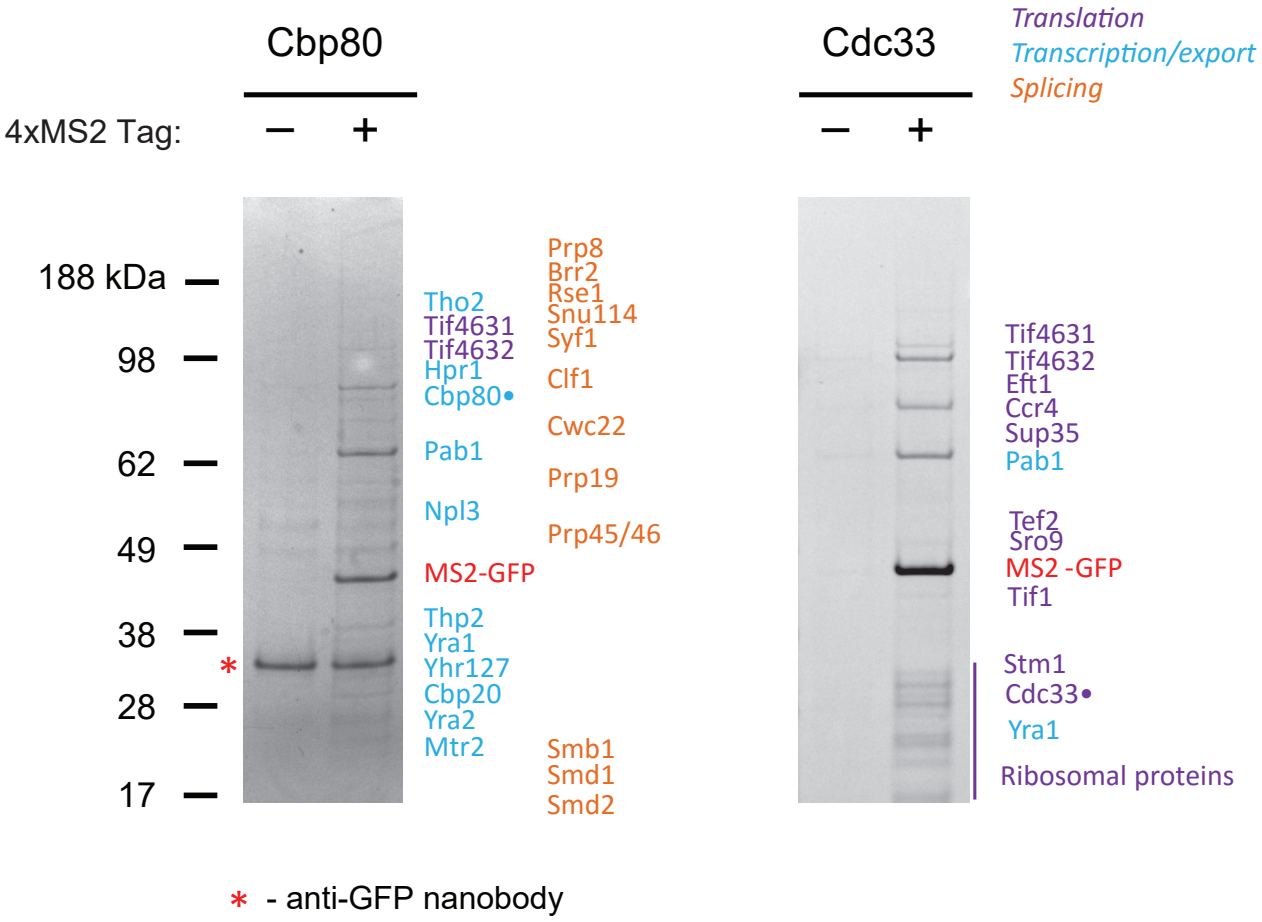



Figure 3.8. Sequential purifications of mCherry-4xMS2 using Cbp80-PrA or Cdc33-PrA. mCherry transcripts with or without a 4xMS2 hairpin tag were co-expressed with MS2CP-GFP in yeast and purified in two steps with Nab2-PrA or Cdc33-PrA and MS2CP-GFP. Final elutions from anti-GFP beads were separated by SDS-PAGE and stained with Coomassie Blue. In parallel, equivalent samples were directly eluted from anti-GFP beads with ammonium hydroxide and analyzed by MS. RNP protein IDs from that analysis are listed, categorized according to function.

Figure 3.9. RNA yield in sequential Cbp80 and MS2CP-GFP purification of mCherry-4xMS2. (a) qRT-PCR was performed on RNA samples taken at each purification step in isolations of untagged or 4xMS2-tagged mCherry transcripts. Primers were used against the mCherry targer sequence, or 18S rRNA to assess non-specific binding of an abundant RNA contaminant. Transcript levels were normalized to total input level (1 and 6).

2) and 7) 0-20% depletion of mCherry is observed in the first step purification of Cbp80-PrA

3) and 8) ~20% of mCherry transcript was recovered in native elutions after the first purification step. This is consistent with the small proportion of mRNA expected to associate with Cbp80, independent of the MS2 tag. No 18S rRNA is detected in this purification step, consistent with Cbp80's specificity for mRNA over rRNA.

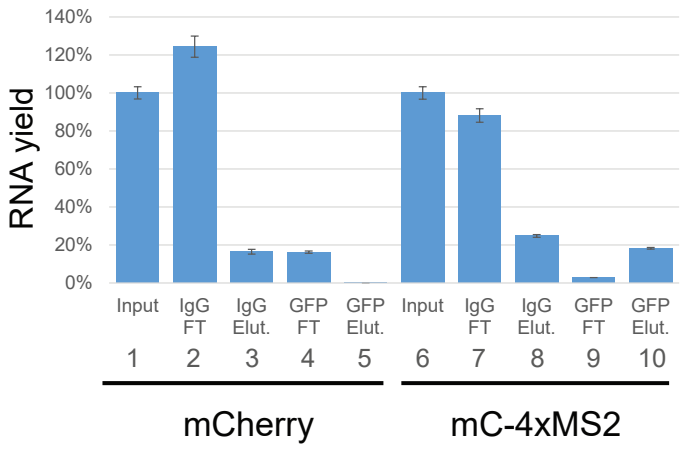
4) and 9) After anti-GFP isolation of MS2CP-GFP, the mCherry transcript is strongly depleted only when 4xMS2-tagged. This reflects MS2CP-GFP's presence only when bound to the tagged transcript.

5) and 10) In elutions from anti-GFP beads, no untagged transcript is detected. Tagged mCherry-4xMS2 is efficiently recovered from the Cbp80-PrA-associated pool (8).

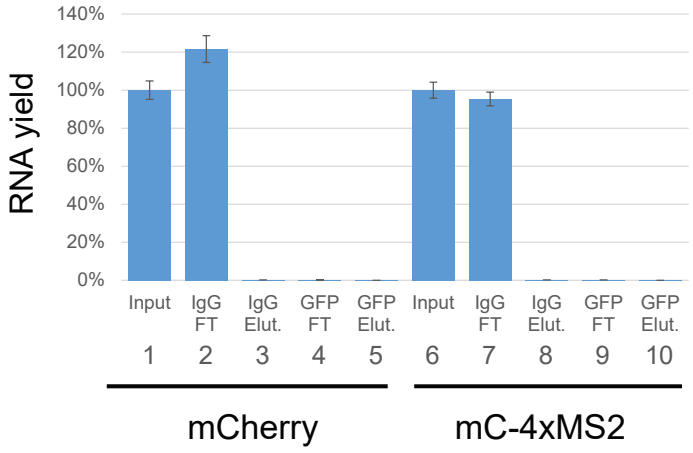
(b) Northern blots were performed against RNA samples equivalent to (a). Radiolabeled probes were complementary to either mCherry or 18S rRNA transcripts. Signals are consistent with the qRT-PCR results.

**a**

**qPCR quantification of mCherry±4xMS2 transcripts throughout MS2 isolation**

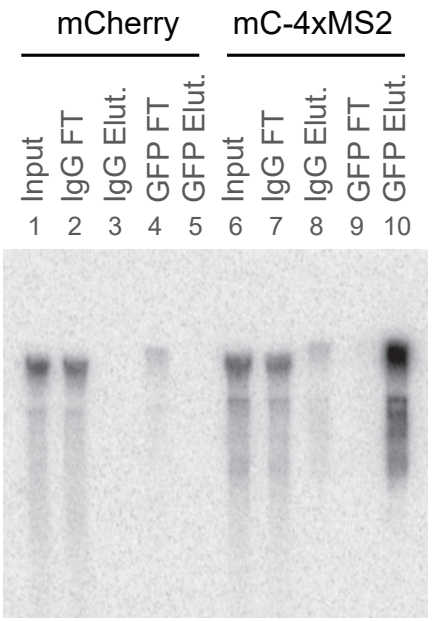



**qPCR quantification of control RNA (18S rRNA) throughout MS2 isolation**

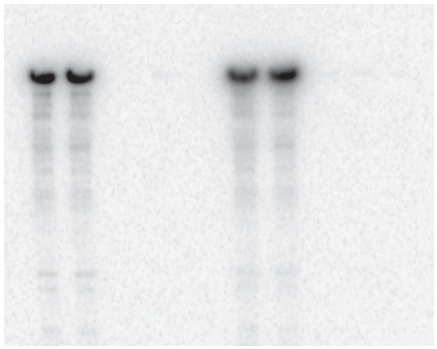



**b**

**mCherry Northern:**




**18S rRNA Northern:**



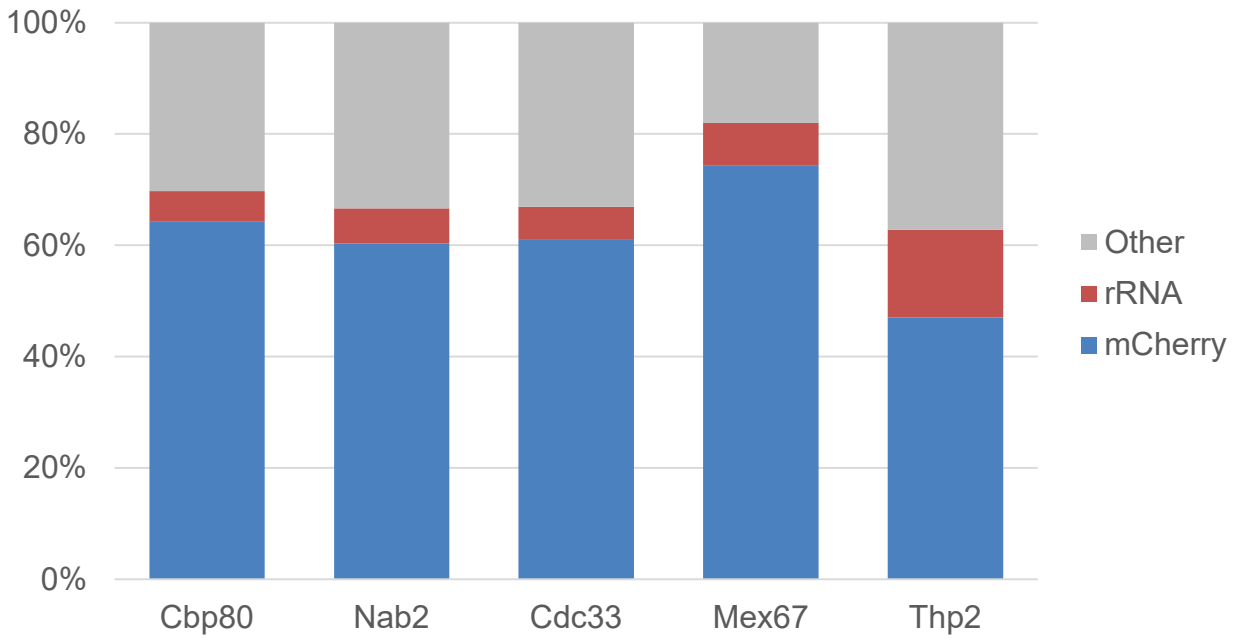



Figure 3.10. RNA-seq analysis of total RNA isolated by sequential purification of mCherry-4xMS2. After sequential purification using the indicated PrA-tagged RNP factor and MS2CP-GFP, RNA was eluted and directly prepared for RNA-seq on an Illumina HiSeq 2000 system. Reads were categorized as either mCherry-4xMS2, rRNA, or other snRNA or mRNA. Depending on the tagged RNP used, 50-70% of all RNA recovered was specific to the tagged mCherry transcript.

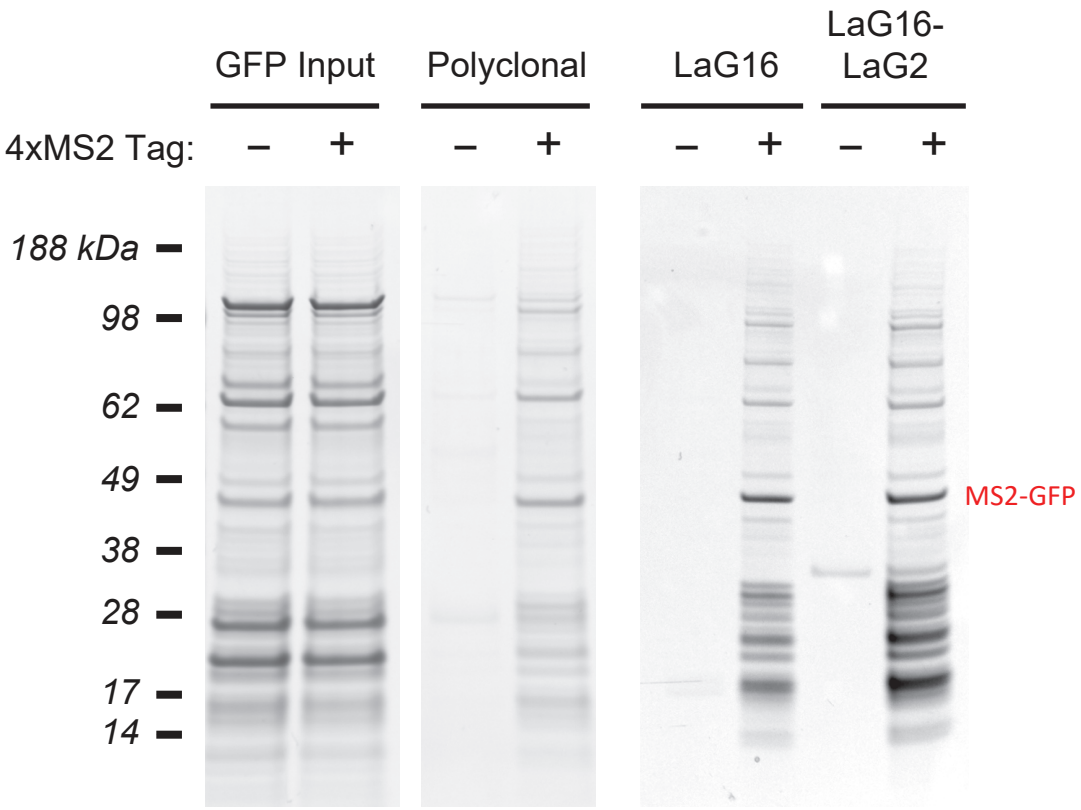



Figure 3.11. Comparing anti-GFP beads in sequential Cdc33-PrA and MS2CP-GFP purifications of mCherry-4xMS2. Tagged or untagged mCherry transcripts were purified first over IgG dynabeads (GFP Input), and then over Dynabeads conjugated to either polyclonal anti-GFP antibodies (Polyclonal), LaG16 nanobody, or the LaG16-LaG2 dimeric nanobody. Protein was eluted with LDS and visualized by Coomassie Blue staining after SDS-PAGE.

was also quantified by qRT-PCR, and ~0.02% was ultimately recovered, with only ~0.2% detected after the first purification step (Figure 3.9a). This is consistent with the divergent pre-rRNA processing pathway expected for this rRNA, and shows that even highly abundant non-specific RNAs are not significantly co-isolated in this method. As a further assessment of RNA recovery, Northern blots were performed probing mCherry and 18S RNA (Figure 3.9b). This recapitulated the qRT-PCR results, and while some degradation is observed, the majority of the RNA purified appears intact.

We also assessed the purity of the final RNA pool recovered in these mRNP isolations by RNA-Seq. This technique allows unbiased quantification of the relative amounts of all transcripts in our samples by sequencing all RNA present (Wang et al., 2009). By this analysis, we determined that the mCherry-4xMS2 transcript comprised approximately 60-70% of purified RNA, depending on the tagged RNP used (Figure 3.10). An additional 5-10% was rRNA, and the remainder was from low levels of other non-specific RNAs. Given the heterogeneity of the non-specific RNA contaminants in these pullouts, it is likely that the final RNP content is largely representative of the highly enriched tagged transcript.

To further optimize the efficiency of the MS2CP-GFP purification step, we transitioned from using polyclonal anti-GFP antibodies from llama serum for the final GFP-MS2 isolations to our own anti-GFP nanobodies (Chapter 2). Both LaG-16 nanobody and the LaG16-G4S-LaG2 dimeric nanobody gave higher yield and lower background in these pullouts, with the dimer performing best in both regards (Figure 3.11). We are thus using this reagent for all subsequent purifications.

# Chapter 4: Pilot survey of RNPs associated with single mRNAs

### *A survey of RNA sequence elements associated with unique processing factors*

As a pilot study to demonstrate the efficacy of the sequential protein-RNA isolation technique, we conducted a survey of several transcripts of interest. As a first step, transcripts were based on a model mCherry sequence, modified in various ways by the introduction of different sequence elements. This allowed us to use the exogenous mCherry sequence as a control, with no expected atypical processing in yeast, and easily expressible at high levels with no biological effects (Shu et al., 2006; Fink et al., 2010; Kafri et al., 2016). This base sequence was then modified with the addition of other promoters, introns, and 3' UTR elements (Figure 4.1). Specifically, we studied introns of different classes, for example comparing those found in highly expressed housekeeping genes like ACT1 to those of ribosomal proteins like RPS30b, whose transcripts are reported to have atypical mRNP compositions (Schmid et al., 2012; Tuck and Tollervey, 2013). An additional interest is comparison of transcripts expressed from promoters with different regulatory mechanisms, such as those from housekeeping or stress-inducible genes. Multiple promoters have also been implicated in post-transcriptional regulation of mRNA stability, and the mRNP complex dynamics involved in these processes are still poorly understood (Bregman et al., 2011; Trcek et al., 2011; Zid and O'Shea, 2014). We began to assess these questions by adding promoters implicated in this downstream transcript regulation to our model mCherry transcript. For these various transcripts of interest, we performed a survey using sequential affinity isolations, beginning with general RNP factors for initial purification.

### *Confirming recovery of Ash1 transcript-specific mRNP factors*

As a first step in determining whether changes in a model transcript affect the co-isolating mRNP composition, we tested transcripts with sequence elements understood to be differentially targeted

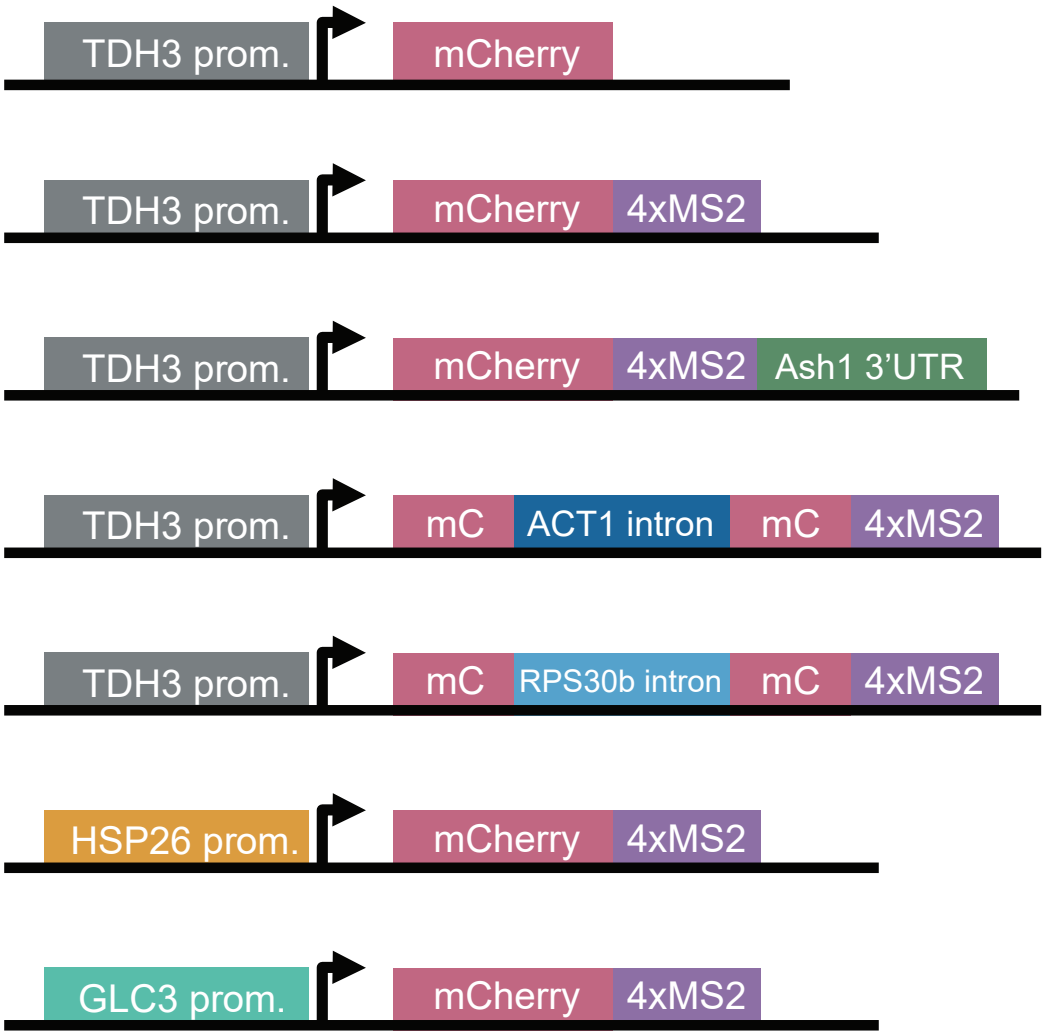



Figure 4.1. MS2-tagged constructs used in survey of mRNPs. An mCherry coding sequence was expressed on a high copy yeast 2μ plasmid under a strong constitutive TDH3 promoter (Mumberg et al., 1995), or promoters from HSP26 or GLC3 genes (Zid and O'Shea, 2014). A 4xMS2 tag was inserted immediately following the mCherry stop codon. As modifications, either the 3'UTR from the ASH1 gene was inserted after the 4xMS2 tag (Bertrand et al., 1998), or intron sequences from ACT1 or RPS30b were inserted 31 bp into the mCherry sequence.

by RNP processing machinery. A particularly well-studied example is 3' UTR region of the Ash1 transcript. Sequence motifs in this UTR recruit She2p and She3p, which target the transcript to the bud tip after export (Long et al., 2000; Olivier et al., 2005). We therefore added the ASH1 3' UTR to our model TDH3-driven mCherry gene, with a 4xMS2 tag between the coding sequence and UTR. The general cap binding protein for nuclear mRNA, Cbp80p, was then used as a PrA-tagged first target for sequential affinity isolation (Chapter 3).

As expected, MS analysis of protein purified from these pullouts identified significant amounts of She2p and She3p, as well as Myo4p, known to be involved in cytoplasmic localization of Ash1 mRNA on actin filaments (Figure 4.2) (Bertrand et al., 1998). Interestingly, we found significant amounts of two other proteins, Htb2p (histone H2B) and Hir3p, part of a nucleosome assembly complex that regulates histone gene expression (Prochasson et al., 2005). Hir3p in particular is a low abundance protein not typically found as a contaminant, but it is unclear what connection might exist between its chromatin remodeling activity and ASH1 transcription. If Hir3p association is confirmed, this connection could be investigated by genetic experiments to assess changes to ASH1 transcription in HIR3 mutants, as well as more targeted proteomics of the HIR complex.

Beyond She2p and She3p, there are other factors reported to be involved in Ash1 localization that were not identified in our purification, notably Loc1p, which stabilizes She2p association, and Puf6p. This may be due to the different dynamics of these proteins' interactions with Ash1 RNA; Loc1p for instance is only bound in the nucleus, and is released upon export through displacement by She3p (Niedner et al., 2013). As we detect She3p but not Loc1p, it may be that Cbp80p-associated Ash1 RNA is highly enriched for the She3p-bound state, presumably due to the rapid kinetics of Loc1p association and dissociation. Puf6p binds Ash1 RNA independently of She2

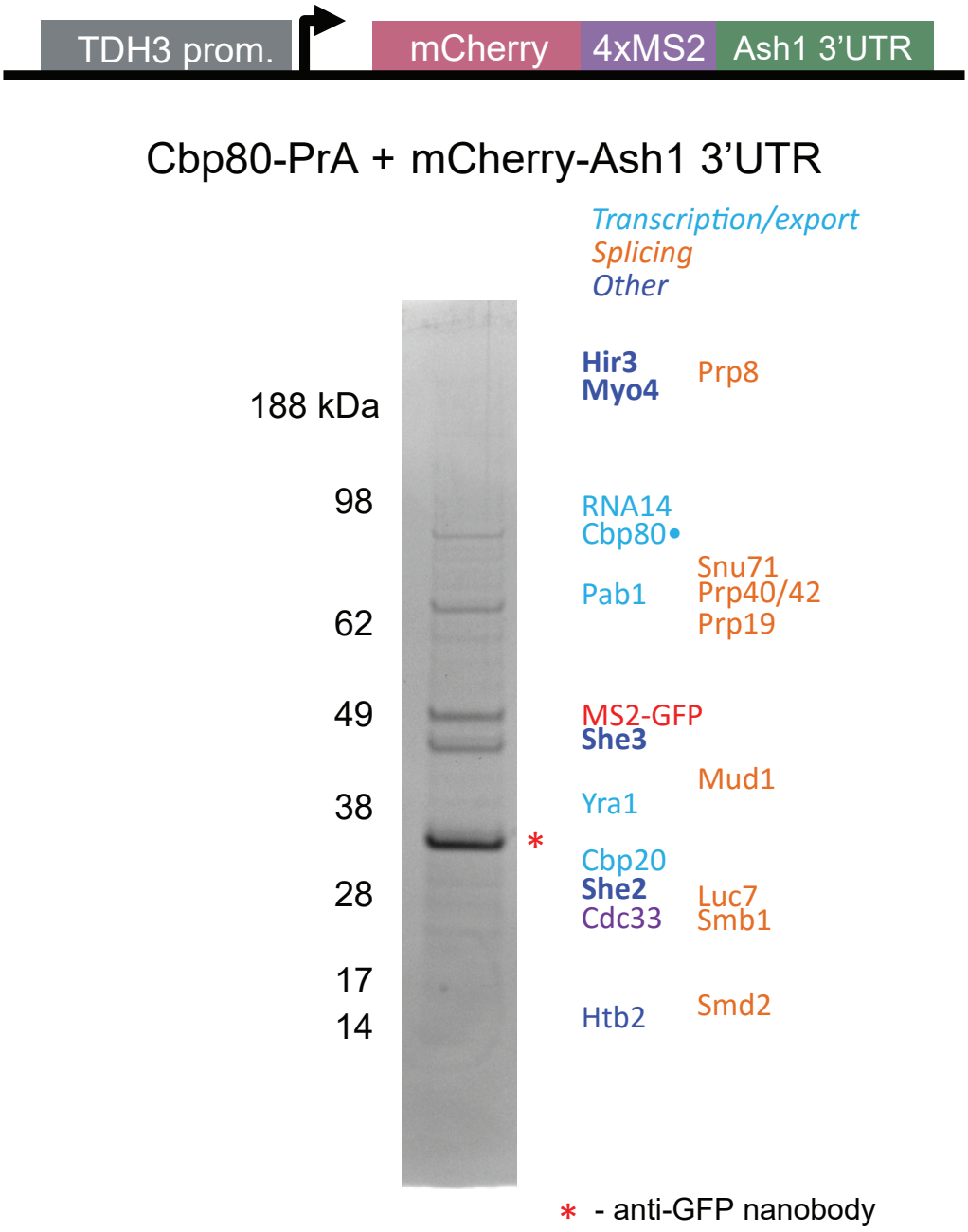



Figure 4.2. Sequential Cbp80-PrA and MS2CP-GFP isolation of mCherry-4xMS2-Ash1 3'UTR mRNPs. A Coomassie-stained gel of the purification was imaged, and protein identifications were made by MS on a whole sample purified in parallel as a gel plug. IDs are shown at predicted molecular weights, and proteins uniquely identifed with the Ash1 3'UTR are shown in bold.

and is thought to remain bound through export. However, it is also highly enriched in the nucleus, and may be less abundant in the purified RNP population due to similar kinetic differences as Loc1p.

Alternatively, it is possible that these proteins are simply less stably associated with the RNA, and are thus not retained throughout the purification process, or else rely on other ASH1 sequence elements not present in the UTR used. In this and other cases though, there is evidence that we are not seeing substantial loss of RNP factors in our isolations. First, previous studies have shown that certain dynamic mRNP proteins, such as Yra1p, do not exchange between mRNAs under our typical purification conditions (Oeffinger et al., 2007). Furthermore, our preliminary mRNP isolations have purified factors suspected to be relatively unstable and dynamic. Intron containing transcripts, for instance, have co-purified with essentially all expected splicing proteins, including relatively transitory proteins like those involved in the second step of catalysis, including Slu7p, Prp16p, and Prp43p (Fabrizio et al., 2009). However, to confirm that dissociation during purification is not a complicating factor, we plan to test this by performing isolations at different time scales, as well as with UV treatment to create RNA-protein crosslinks. By determining if shorter incubations or UV stabilization significantly affects mRNP composition, we can assess the stability of our natively purified complexes.

***mRNP factors specific to introns***

We next performed isolations of a 4xMS2-tagged mCherry transcript containing artificially inserted introns. First, we inserted an intron from the ACT1 gene, a well-studied, highly expressed yeast cytoskeletal gene with a 309 nt intron (Gallwitz and Sures, 1980). In comparison to Nab2p-based pullouts with mCherry transcripts lacking an intron, pullouts of the ACT1 intron-containing mRNA show clear changes in the banding pattern of the co-isolated mRNP complex (Figure 4.3).

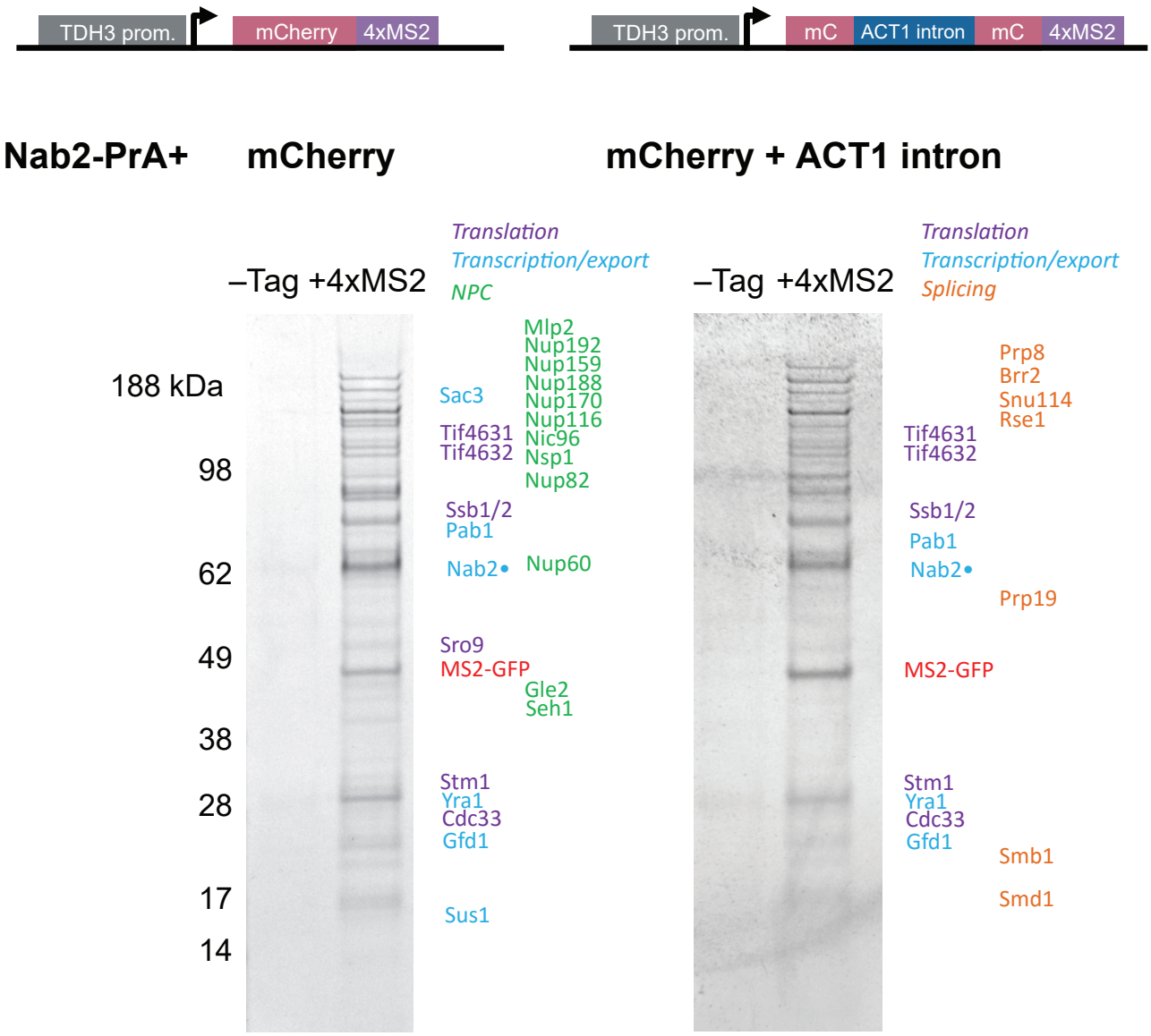



Figure 4.3. Sequential Nab2-PrA and MS2-GFP isolations of mCherry-4xMS2 and mCherry(Act1 intron)-4xMS2 mRNPs, alongside negative controls from transcripts with no MS2 tag. Coomassie-stained gels of the purifications were imaged, and protein identifications were made by MS on directly eluted whole samples prepared in parallel. IDs are shown at predicted molecular weights.

Mass spectrometric analysis of protein elutions from these pullouts provided a finer look at these distinctions. While the intronless mCherry transcript was associated strongly with many nuclear pore complex components, transcripts containing an intron were instead strongly associated with numerous spliceosomal factors (Figure 4.3). This difference is likely explained by a change in the rate-limiting step of the processing of Nab2p-associated mRNA from export to splicing. It is clear that export of intron-containing mRNAs is dependent on completion of splicing, which seems to be required for co-transcriptional recruitment of the Yra1p export adapter (Lei and Silver, 2002). However, the relative kinetics of transcription, splicing, and export in yeast remain somewhat unclear. While a number of studies have shown that splicing occurs co-transcriptionally, it is possible that some transcripts with shorter second exons are spliced post-transcriptionally (Kotovic et al., 2003; Gornemann et al., 2005; Tardiff et al., 2006). There is increasing evidence, though, that transcriptional pausing occurs in terminal exons, leading to primarily co-transcriptional splicing in genes, regardless of exon length (Carrillo Oesterreich et al., 2010; Churchman and Weissman, 2011). Assuming the necessity of splicing for export, this relatively slow process, reported to be in the range of 60 sec (Tardiff et al., 2006; Alexander et al., 2010), would be expected to be rate-limiting in comparison to relatively rapid nuclear export (~180 ms) (Grunwald and Singer, 2010).

As Cbp80p plays a direct role in spliceosome recruitment and is expected to be present in mRNPs undergoing splicing (Lewis et al., 1996), we performed an equivalent pair of affinity isolations with Cbp80p used as the initial mRNP target (Figure 4.4). In this case, all major spliceosomal factors were identified by MS in the presence of an intron, compared to a subset of U1 components with the intronless transcript. Given Cbp80p's involvement in early splicing and the persistence

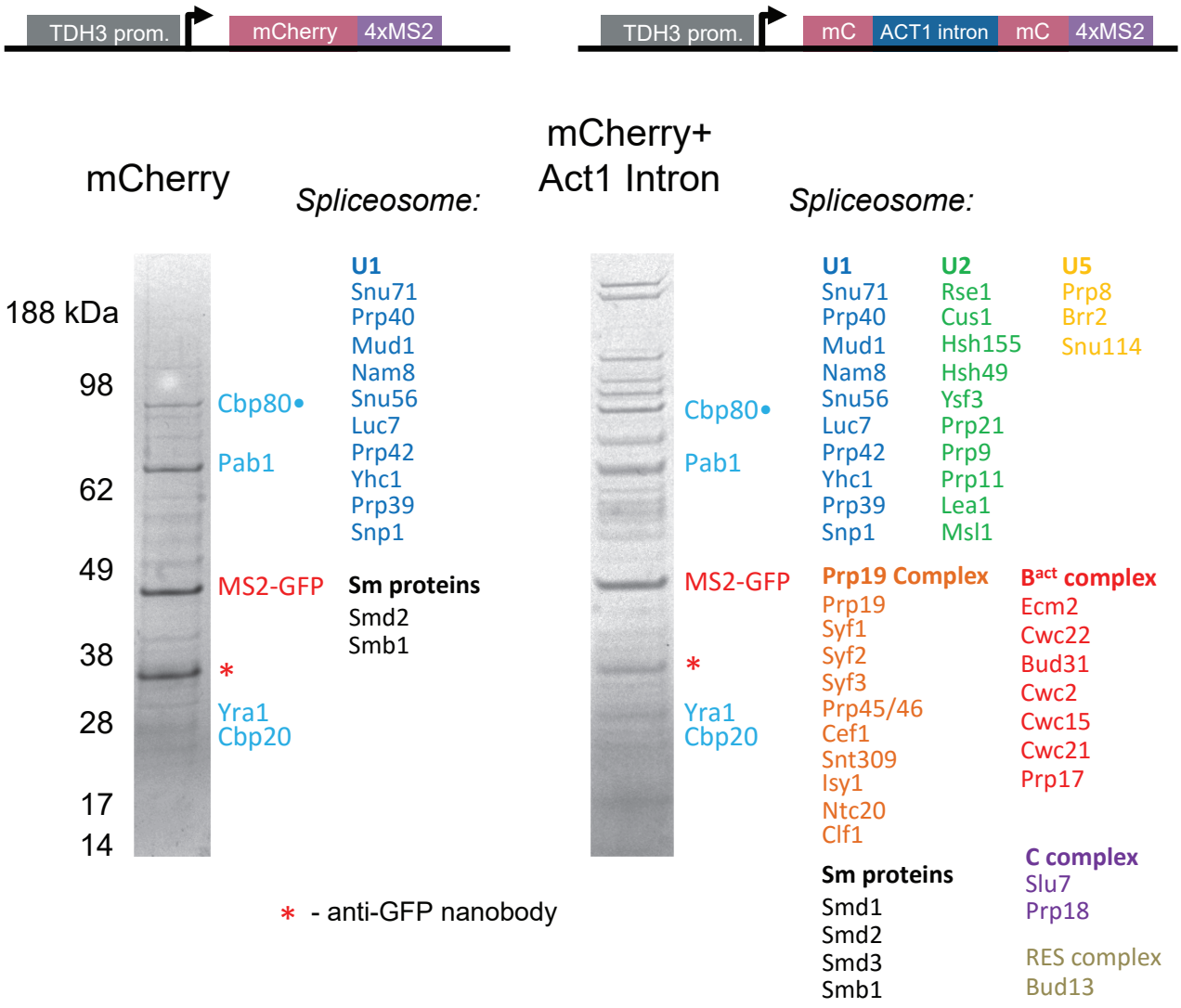



Figure 4.4. Sequential Cbp80-PrA and MS2-GFP isolations of mCherry-4xMS2 and mCherry(Act1 intron)-4xMS2 mRNPs. Coomassie-stained gels of the purifications were imaged, and protein identifications were made by MS on whole samples purified in parallel as gel plugs. IDs corresponding to spliceosome components are listed separately according to subunit classification.

of the cap complex, it is expected to be associated with the active spliceosome, consistent with the presence of the full spliceosome in our isolations of intron-containing RNAs (Lewis et al., 1996).

It has been less clear which spliceosomal factors, if any, associate with intronless transcripts. The U1 snRNP, generally the earliest spliceosome component to bind, is known to recognize the 5' splice site of introns, and is therefore predicted to be highly enriched on intron-containing transcripts (Seraphin and Rosbash, 1989; Will and Luhrmann, 2011). Despite this enrichment, U1 components have been found at low but detectable levels on many intronless genes (Kotovic et al., 2003). While little has been known of U1 snRNP's nonspecific mRNA interactions, our identification of U1 factors on intronless transcripts supports a model in which recognition of true splice site sequences depends on promiscuous scanning, or transitory binding of all transcripts. However, it is also conceivable that the mCherry sequence used in these experiments includes uncharacterized cryptic splice sites, so other transcripts would have to be tested to confirm this model.

***Label free protein quantification by mass spectrometry***

While observing the presence or absence of different RNP factors has proved informative in the case of the clear-cut changes in composition seen with the Ash1 3' UTR or an intron, it would be useful to detect changes in the relative abundance of such factors as well. It is likely that important variations in RNA processing could be associated with several-fold differences in RNP protein abundance, and identifying even the approximate stoichiometry would be useful in these more subtle cases.

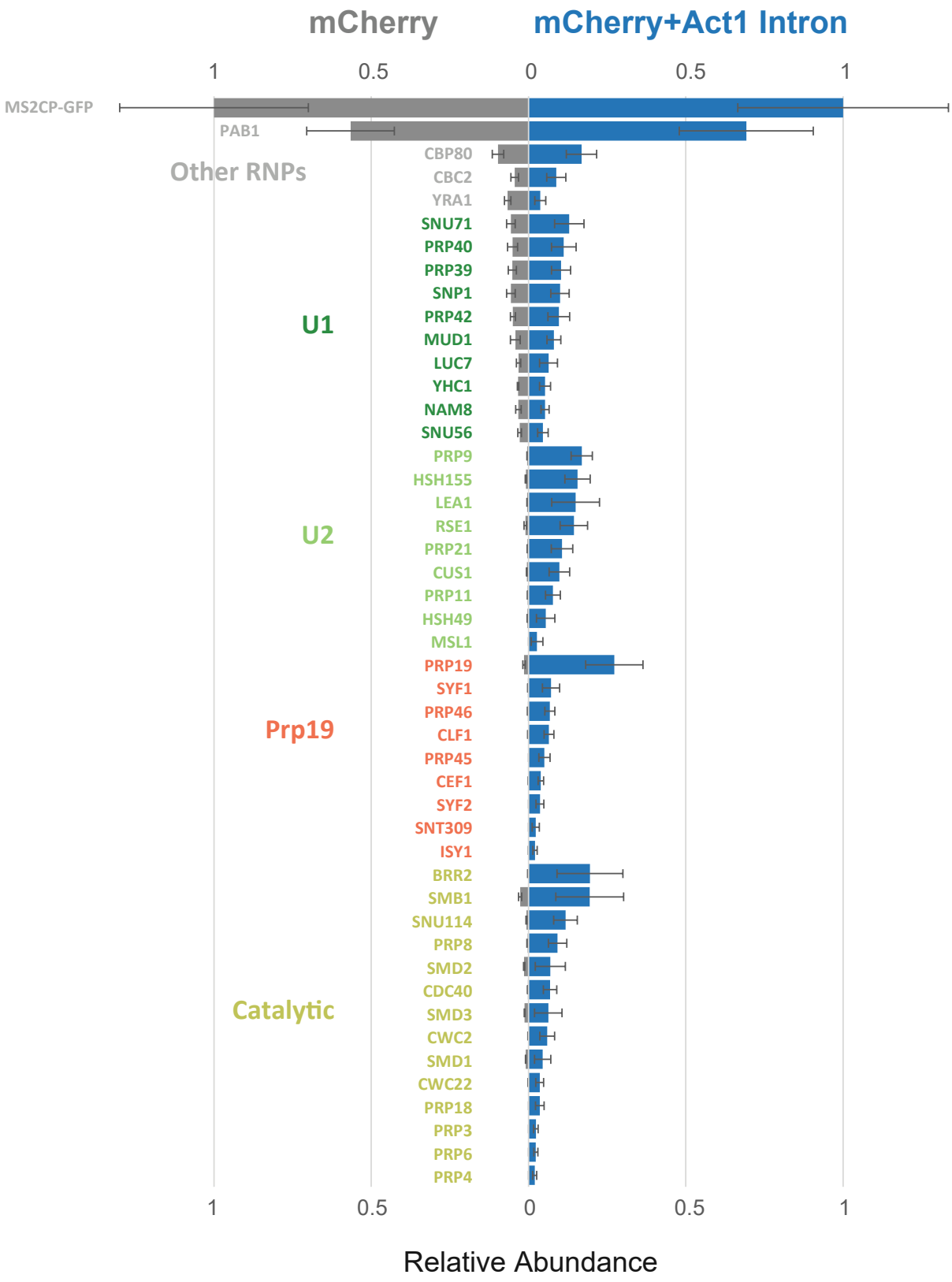
As a straightforward method for quantifying protein levels both between and within samples, we pursued label-free MS quantification techniques (with Erica Jacobs, B. Chait lab). This avoids the

technical difficulties involved in differential stable isotope labeling of all cells to be compared (Oda et al., 1999; Ong et al., 2002), while retaining sufficient accuracy to identify significant changes in protein abundances. While many variations of label-free quantification approaches have been described, it has generally been found that the most reliable methods make use of peak intensities of three or more peptides from proteins to be quantified (Old et al., 2005; Silva et al., 2006; Bantscheff et al., 2007; Schwanhausser et al., 2011). This has led to reported accuracies within 20%, sufficient to detect the multi-fold differences in abundance of most interest to us, and has been found to be somewhat more accurate than similar spectral counting approaches. While most accurate in comparing the relative abundance of the same protein in different samples, this technique has also been successfully used to identify the stoichiometry of different proteins within a single complex (Smits et al., 2013; Taylor et al., 2013). This is a further advantage of label-free quantification, as isotope-based quantification in these cases would require the laborious and difficult production of labeled reference peptides for each protein to be measured (Pratt et al., 2006).

While we attempted a number of different variations of quantification by peptide intensity, including iBAQ, we found that straightforward quantification by proteins' three most intense peptides (TOP3), gave similar, consistent results (Silva et al., 2006; Schwanhausser et al., 2011). Due to the difficulties of detecting multiple peptides from smaller molecular weight proteins, it is also important to note that results from this analysis are most reliable for larger proteins.

We first applied this quantification method to triplicate isolations and MS analyses of mCherry transcripts with or without an Act1 intron, purified with a first step isolation of Cbp80-PrA (Figure 4.5). Not surprisingly, MS2CP-GFP and Pab1p were dominant proteins, with about 5-fold higher intensities than any others. This reflects MS2CP-GFP's ability to bind to four MS2 hairpins as a

Figure 4.5. Label free quantification of sequential Cbp80-PrA and MS2CP-GFP purifications of mCherry with or without an Act1 intron. MS was performed on triplicate isolations of mCherry-4xMS2 or mCherry(Act1 intron)-4xMS2. Proteins were quantified by their top 3 peptide intensities, and normalized to MS2CP-GFP. After filtering out known contaminants, the top 47 protein hits were grouped according to function (U1, U2, or catalytic snRNPs, members of the Prp19 complex, or other RNPs) and visualized. While similar levels of non-splicing RNPs or U1 splicing factors were detected in both samples, other splicing proteins are strongly enriched with the intron-containing transcript.


mCherry
mCherry+Act1 Intron
1
0.5
0
0.5
1
MS2CP-GFP
PAB1
Other RNPs
CBP80
CBC2
YRA1
U1
SNU71
PRP40
PRP39
SNP1
PRP42
MUD1
LUC7
YHC1
NAM8
SNU56
U2
PRP9
HSH155
LEA1
RSE1
PRP21
CUS1
PRP11
HSH49
MSL1
Prp19
PRP19
SYF1
PRP46
CLF1
PRP45
CEF1
SYF2
SNT309
ISY1
Catalytic
BRR2
SMB1
SNU114
PRP8
SMD2
CDC40
SMD3
CWC2
SMD1
CWC22
PRP18
PRP3
PRP6
PRP4
Relative Abundance

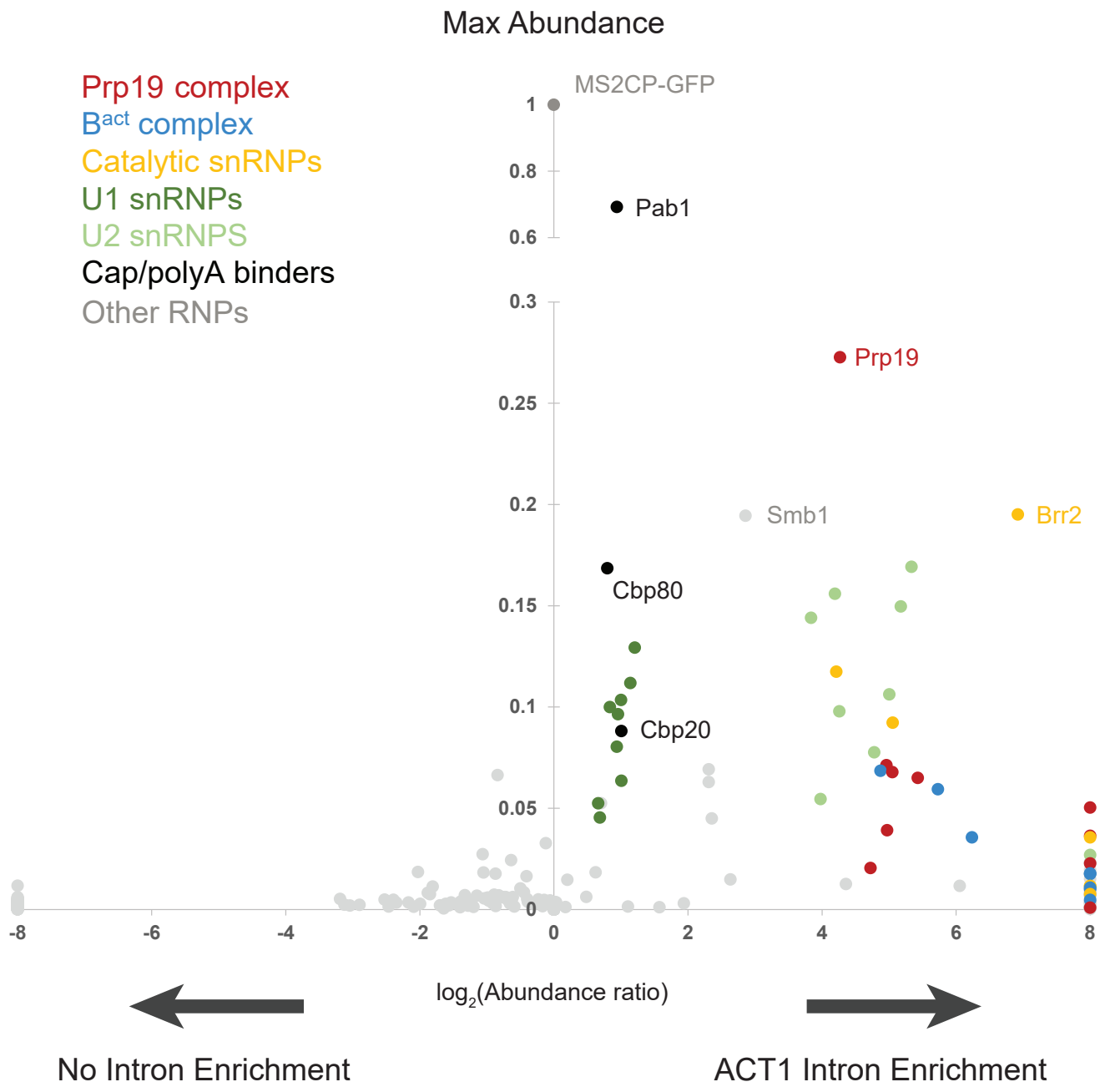
dimer, giving it a predicted stoichiometry of 8 per transcript, and Pab1p's tendency to bind the polyA tail in multiple copies (Sachs et al., 1987). These and other non-splicing related RNPs were found at roughly similar amounts in both intron and non-intron samples. Looking at splicing factors, we also detected similar quantities of U1 snRNPs with both transcripts. However, later spliceosomal factors, such as U2 and catalytic snRNPS or Prp19 complex components, were highly enriched with the ACT1 intron-containing transcript. While most of these factors were not identified in the intronless sample, those that were appeared at 10-100 fold lower levels.

To allow visualization of the relative levels of all identified proteins in two isolations, we also graphed these data in a variation of a volcano plot (Figure 4.6). In this representation, each protein is shown as a point, with the y-axis value indicating its maximum abundance in either sample, and the x-axis value equal to the log value of its enrichment in one condition versus the other. Looking at this plot to compare the presence and absence of the ACT1 intron, it is clear that all U1 snRNPs are detected at approximately the same modest enrichment in the intron sample. A stronger consistent enrichment is also seen for elements of the Prp19 complex, where it can also be noted that Prp19p, which forms a tetrameric core of this complex, is roughly 4-fold more abundant than other members of the complex. Finally, general mRNA-binding proteins like Pab1 and Cbp80 are found along the y-axis, reflecting the expected equivalent abundances between the intron-containing and intronless transcripts.

***Promoter-driven changes in transcript maturation***

Several recent studies have revealed that promoter regions can independently confer downstream, transcript-level regulatory alterations in their gene products (Bregman et al., 2011; Trcek et al., 2011). As representative examples, we pursued classes of promoters responsive to glucose starvation that confer different translational activity to their transcripts (Zid and O'Shea, 2014).

Figure 4.6. Volcano plot of quantifications from sequential Cbp80-PrA and MS2CP-GFP purifications of mCherry with or without an Act1 intron. Proteins quantified as in Figure 4.5 were plotted according to the $\log_2$ ratio of relative abundance in the mCherry(Act1 intron) sample over the mCherry sample (x-axis). Y-axis values reflect protein abundance in the enriched sample, relative to MS2CP-GFP. Proteins from the same complex, like U1 snRNPs, tend to be enriched at similar levels. Proteins found in only one sample are shown at the corresponding edge of the plot.


Max Abundance
Prp19 complex
Bact complex
Catalytic snRNPs
U1 snRNPs
U2 snRNPS
Cap/polyA binders
Other RNPs
MS2CP-GFP
Pab1
Prp19
Smb1
Brr2
Cbp80
Cbp20
1
0.8
0.6
0.3
0.25
0.2
0.15
0.1
0.05
0
-8
-6
-4
-2
0
2
4
6
8
log2(Abundance ratio)
No Intron Enrichment
ACT1 Intron Enrichment

While many genes are up-regulated at the mRNA level in response to glucose stress, only some, such as HSP26, have concomitant increases in protein expression, while others, like GLC3, surprisingly show no protein increase at all. The transcripts of these genes have been found to localize to P bodies and stress granules, and are not actively translated during starvation. The differential regulation of these two types of transcripts was found to be due solely to the corresponding promoter sequence, and was independent of the transcript sequence. We thus sought to identify possible RNP determinants of this regulation by expressing MS2-tagged mCherry with either GLC3 or HSP26 promoters, and performing mRNA isolations under glucose starvation conditions.

As promoter-driven changes in transcript composition would be predicted to occur co-transcriptionally, we isolated mRNA using Thp2-PrA as the tagged mRNP. Thp2p is a member of the THO/TREX complex, recruited during transcription and involved in transcriptional elongation (Jimeno et al., 2002; Strasser et al., 2002). One notable difference in RNP composition was found in mCherry transcripts driven by the HSP26 promoter, rather than GLC3 or the TDH3 control (Figure 4.7 and 4.8). While members of the Prp19 complex are found at significant levels with GLC3 or TDH3, these were not identified with the HSP26 sample. While the Prp19 complex is predominantly recognized as an activator of splicing, it has more recently been shown to recruit the TREX complex to at least some actively transcribed genes (Chanarat et al., 2011). The Prp19 complex appears to associate with RNA polymerase II, and is required for full association of the THO/TREX complex to active genes, even those without introns. It has not been clear whether the Prp19 complex is in fact required for efficient transcription at all genes, and our results suggest that there may be cases, such as HSP26, where it is not necessary for THO/TREX recruitment.

Figure 4.7. Label free quantification of sequential Thp2-PrA and MS2CP-GFP purifications of mCherry expressed from HSP26 or GLC3 promoters. MS was performed on duplicate isolations of HSP26-mCherry-4xMS2 or GLC3-mCherry-4xMS2 from yeast starved of glucose for 30 minutes. Proteins were quantified by their top 3 peptide intensities, and normalized to MS2CP-GFP. After filtering out known contaminants, the top 34 protein hits were grouped according to function (U1, U2, or catalytic snRNPs, members of the Prp19 or THO/TREX complexes, or other RNPs) and visualized. While similar levels of THO/TREX and other non-splicing RNPs were found in each sample, significantly more splicing proteins, particularly factors associating after U1, were detected with the GLC3 promoter-driven transcript.

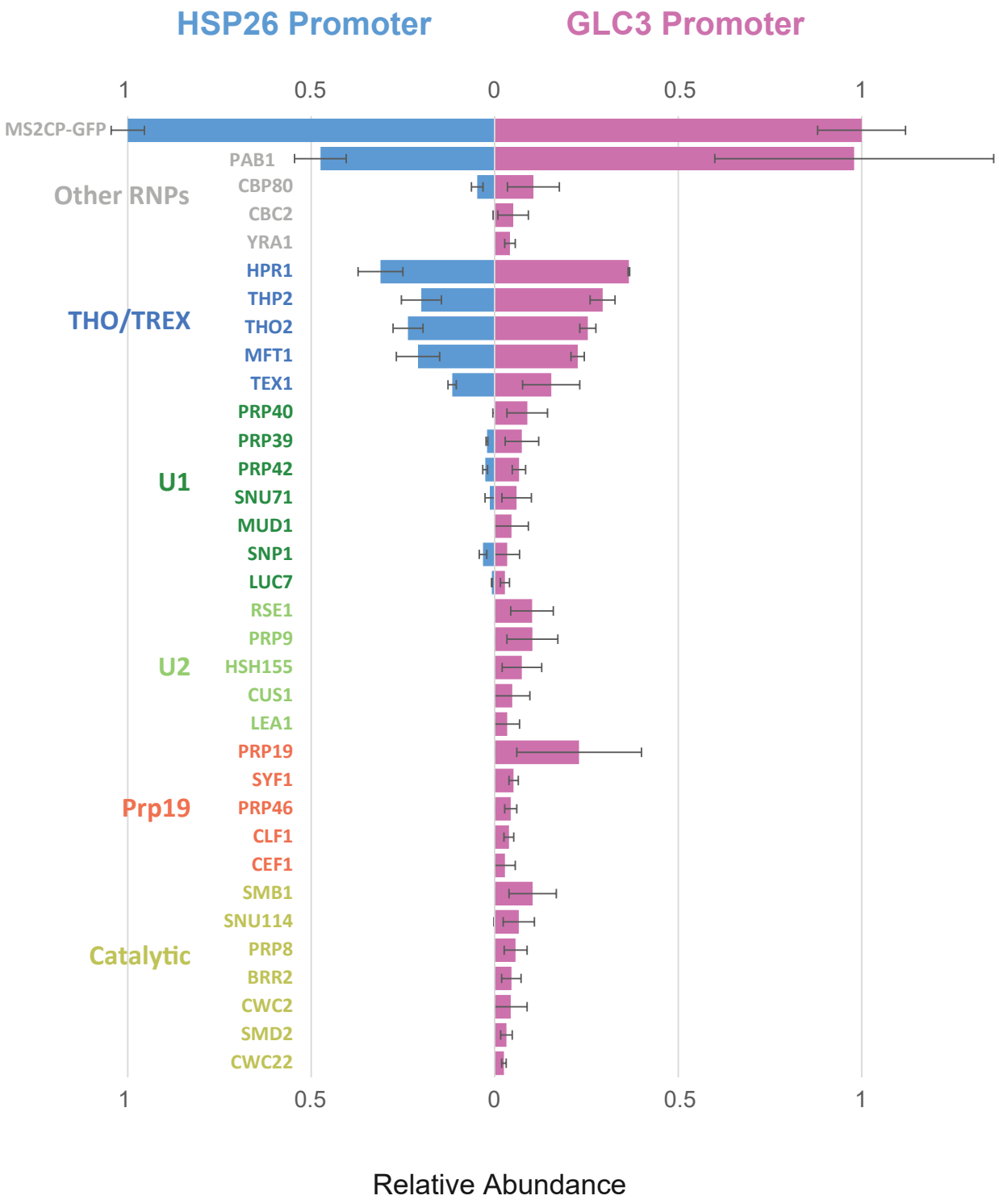
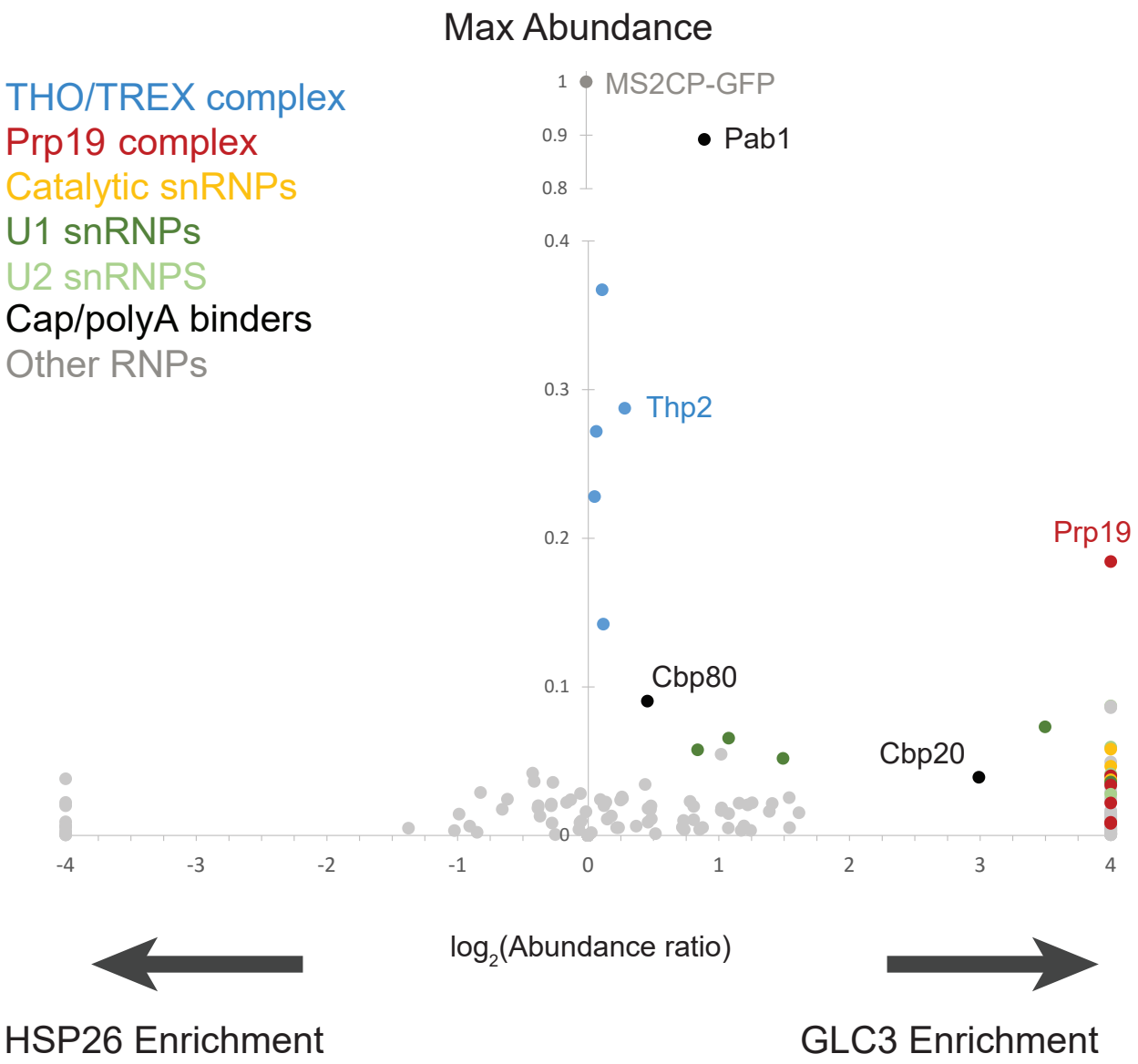

HSP26 Promoter
GLC3 Promoter
1
0.5
0
0.5
1
MS2CP-GFP
PAB1
Other RNPs
CBP80
CBC2
YRA1
HPR1
THP2
THO/TREX
THO2
MFT1
TEX1
PRP40
PRP39
PRP42
U1
SNU71
MUD1
SNP1
LUC7
RSE1
PRP9
U2
HSH155
CUS1
LEA1
PRP19
SYF1
Prp19
PRP46
CLF1
CEF1
SMB1
SNU114
Catalytic
PRP8
BRR2
CWC2
SMD2
CWC22
Relative Abundance

Figure 4.8. Volcano plot of quantifications from sequential Thp2-PrA and MS2CP-GFP purifications of mCherry expressed from an HSP26 or GLC3 promoter. Proteins quantified as in Figure 4.7 were plotted according to the $\log_2$ ratio of relative abundance in the GLC3 promoter sample over the HSP26 promoter sample (x-axis). Y-axis values reflect protein abundance in the enriched sample, relative to MS2CP-GFP. Later splicing factors like U2 or Prp19 complex components are uniquely identified with the GLC3 promoter sample, while general factors like Pab1p or THO/TREX factors are equally abundant in both experiments. Proteins found in only one sample are shown at the corresponding edge of the plot.


Max Abundance
THO/TREX complex
Prp19 complex
Catalytic snRNPs
U1 snRNPs
U2 snRNPS
Cap/polyA binders
Other RNPs
MS2CP-GFP
Pab1
Thp2
Prp19
Cbp80
Cbp20
1
0.9
0.8
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4
log2(Abundance ratio)
HSP26 Enrichment
GLC3 Enrichment

While we do isolate THO/TREX in these pullouts (a TREX component is in fact used to isolate the associated transcript), it remains possible that the lack of Prp19 components found with the HSP26 sample reflect different kinetics of this recruitment.

While more evidence is needed before we can suggest any link between the Prp19 complex and the different fates associated with HSP26 and GLC3 transcripts once exported from the nucleus, it is an intriguing possibility. We can thus perform additional controls to determine whether this complex is indeed required for promoter-dependent regulation of these transcripts. First, we can repeat Thp2-based mRNP isolations of all transcripts with and without glucose starvation. If Prp19 is playing a role in transcriptional response to this stress, it should be detected at different levels when cells are starved. We can also tag a Prp19 factor rather than Thp2 for RNA purifications – this could aid detection of other proteins involved in transcript-specific regulation that interact with the Prp19 complex. Further, to confirm that the conditions used are in fact leading to changes in cytoplasmic processing, we will repeat these purifications using a cytoplasmic mRNP, such as Cdc33p. This would be expected to isolate P body or stress granule components in the case of activated GLC3, but not HSP26.

It is also possible that Prp19 is intrinsically required for full transcription from some promoters and not others, independent of activation of the stress-response genes we investigated. To pursue this, we can investigate the Thp2-bound mRNP composition of a broader range of transcripts, including those with either constitutive or inducible promoters. This could demonstrate differential involvement of Prp19, and potentially reveal other factors able to take the place of Prp19 in THO/TREX recruitment.

***Ribosomal protein introns***

A number of studies have suggested differences in the ways that certain introns are processed. This has been clearest in the case of a number of ribosomal protein genes that appear to be differentially degraded and spliced. For instance, pre-mRNAs of a number of such genes, but not intron-containing genes generally, were found to be targeted for degradation by the nuclear exosome in a Nab2p-dependent manner (Schmid et al., 2012). Moreover, it has been seen that different classes of spliced genes have distinct mechanisms of splicing, as observed by differences in splicing efficiency and sensitivity to various spliceosome mutants (Pleiss et al., 2007). Ribosomal protein genes are notable for their atypically high relative levels of spliced versus unspliced precursor RNA, as well as for their distinct susceptibilities to different spliceosomal defects.

To investigate these questions of intron-dependent processing, we compared mCherry transcripts containing an intron from either ACT1 or RPS30b, a representative ribosomal protein gene, as well as an intronless control. The lengths of these introns (309 bp and 411 bp, respectively) are similar, and thus not expected to contribute to significant differences in splicing activity. We MS2-tagged and isolated these transcripts for MS analysis, using Cbp80-PrA as the RNP purification target to maximize yield of spliceosome-associated RNA. Quantitative comparison of associated RNP factors revealed striking differences between these samples (Figure 4.9 and 4.10). First, U1 and U2 snRNPs, the earliest to bind an intron as part of the prespliceosome complex (complex A), are enriched about 3-fold in the ACT1 vs RPS30b intron. Conversely, elements of the activated and catalytic spliceosome (complex $B_{act}$ or C) are enriched 3-fold in the RPS30b intron. This includes the Prp19 complex, all members of which are enriched to similar extents with the RP30b

Figure 4.9. Label free quantification of sequential Cbp80-PrA and MS2CP-GFP purifications of 4xMS2-tagged mCherry containing an intron from either Act1 or Rps30b. MS was performed after mRNP isolations performed in triplicate. Proteins were quantified by their top 3 peptide intensities, and normalized to MS2CP-GFP. After filtering out known contaminants, the top 47 protein hits were grouped according to function (U1, U2, or catalytic snRNPs, members of the Prp19 complex, or other RNPs) and visualized. Similar levels of non-splicing RNPs are seen in each sample, but while the Act1 intron co-purifies higher levels of U1 and U2 components, factors from later spliceosome subunits are strongly enriched with the Rps30b intron sample.

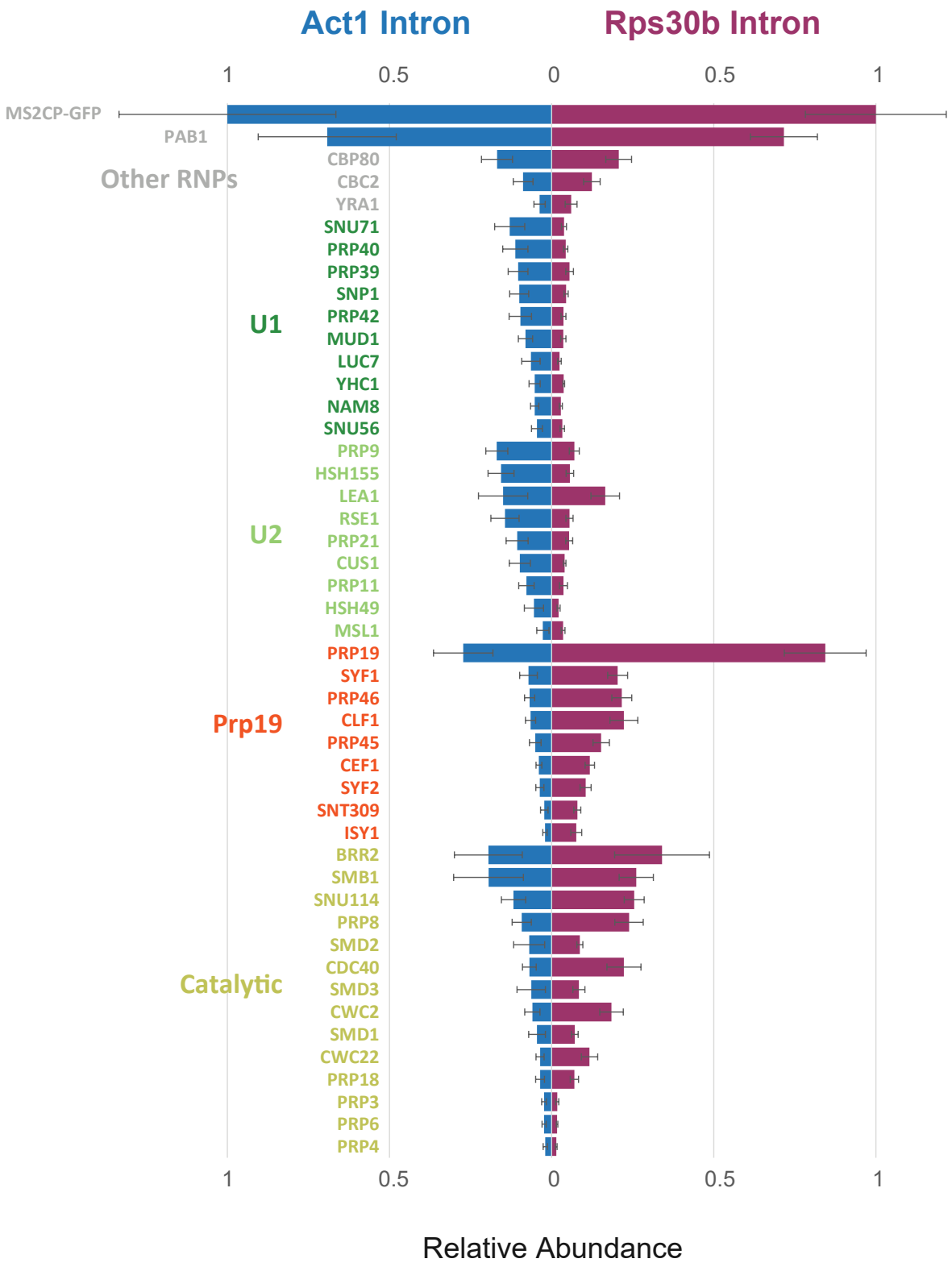
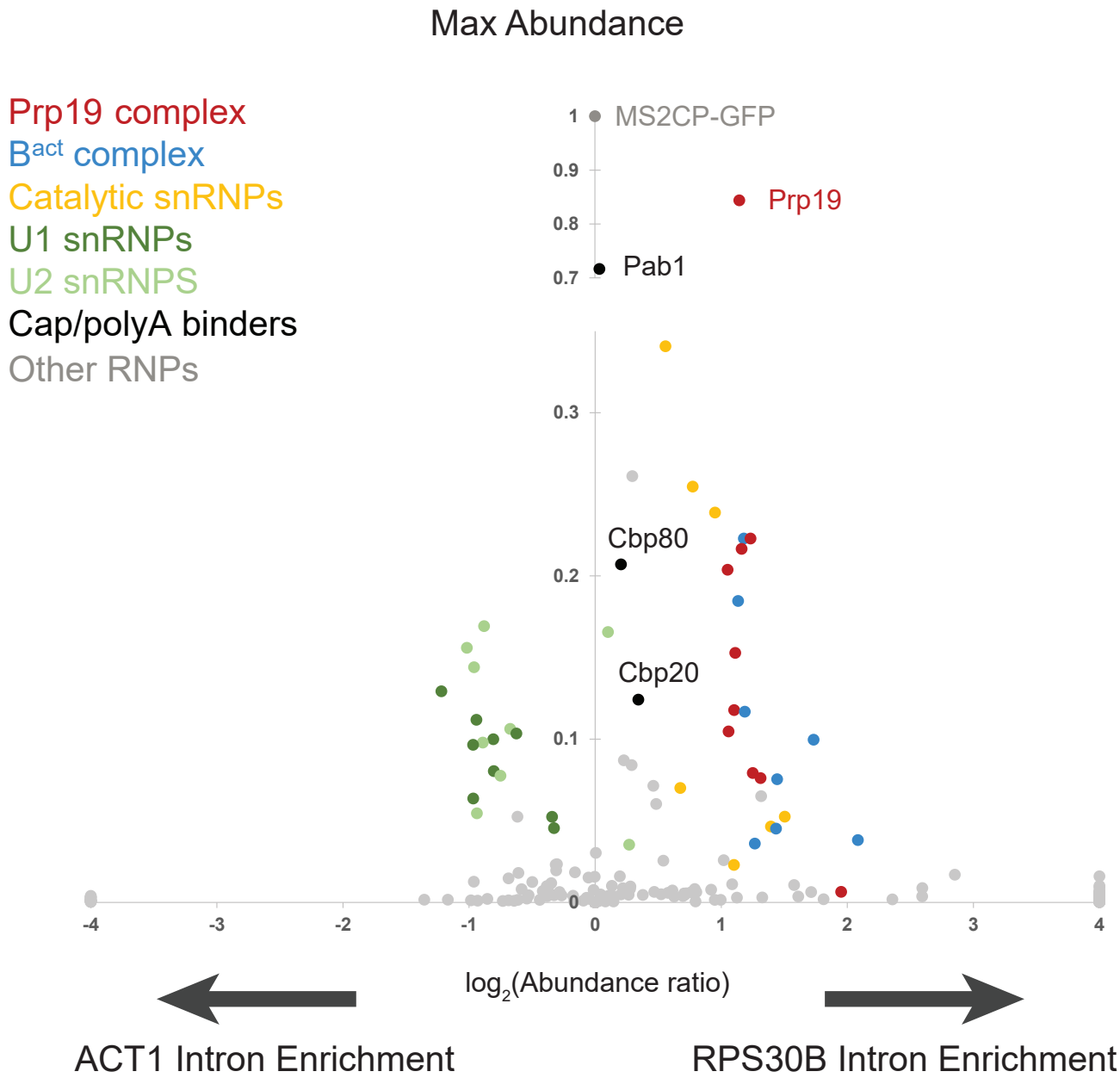

Act1 Intron
Rps30b Intron
1
0.5
0
0.5
1
MS2CP-GFP
PAB1
Other RNPs
CBP80
CBC2
YRA1
U1
SNU71
PRP40
PRP39
SNP1
PRP42
MUD1
LUC7
YHC1
NAM8
SNU56
U2
PRP9
HSH155
LEA1
RSE1
PRP21
CUS1
PRP11
HSH49
MSL1
Prp19
PRP19
SYF1
PRP46
CLF1
PRP45
CEF1
SYF2
SNT309
ISY1
Catalytic
BRR2
SMB1
SNU114
PRP8
SMD2
CDC40
SMD3
CWC2
SMD1
CWC22
PRP18
PRP3
PRP6
PRP4
1
0.5
0
0.5
1
Relative Abundance

Figure 4.10. Volcano plot of quantifications from sequential Cbp80-PrA and MS2CP-GFP purifications of mCherry with an intron from either Act1 or Rps30b. Proteins quantified as in Figure 4.9 were plotted according to the $\log_2$ ratio of relative abundance in the Rps30b intron sample over the Act1 intron sample (x-axis). Y-axis values reflect protein abundance in the enriched sample, relative to MS2CP-GFP. While earlier U1 and U2 splicing subunits are enriched with the Act1 intron, later catalytic subunits are more abundant in the Rps30b intron sample. In all cases, proteins associated with a particular splicing complexe are found at a similar relative enrichment. Proteins found in only one sample are shown at the corresponding edge of the plot.

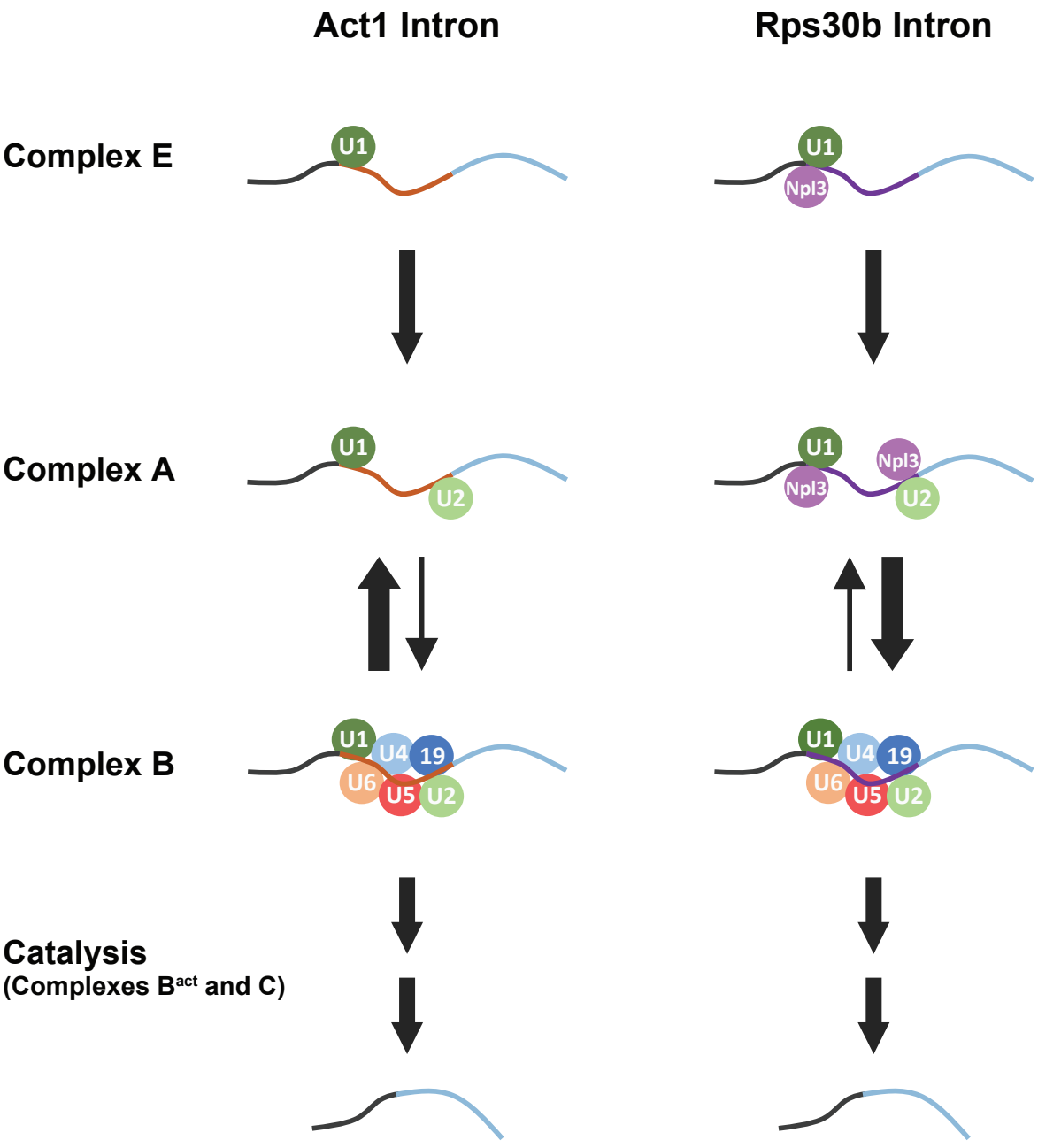

Max Abundance
Prp19 complex
Bact complex
Catalytic snRNPs
U1 snRNPs
U2 snRNPS
Cap/polyA binders
Other RNPs
MS2CP-GFP
Prp19
Pab1
Cbp80
Cbp20
1
0.9
0.8
0.7
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4
log2(Abundance ratio)
ACT1 Intron Enrichment
RPS30B Intron Enrichment

intron. Notably, two U2-associated proteins, Lea1p and Msl1p, are present at equivalent amounts in both samples; these proteins directly associate in a complex with U2 snRNA, and may therefore persist with the core U2 snRNP through the splicing process, in contrast to the other U2 proteins detected.

The differential enrichments of these complexes could be explained by several factors. Most obviously, different kinetics of the splicing process for each intron could lead to varying enrichment of transcripts in earlier versus later steps of splicing. Thus if the ACT1 intron remains associated longer with the prespliceosome before catalytic splicing occurs, we could expect the observed increase in prespliceosomal U1 and U2 snRNPs over catalytic proteins (Figure 4.11). There is in fact reason to believe this is the case – a number of ribosomal protein genes have been seen to have proportionally less unspliced pre-mRNA than ACT1 (Pleiss et al., 2007). While the reason for this has not been clear, our results would suggest that the ACT1 intron is delayed at the presplicesome step, before components of the precatalytic splicesosome (complex B) associate.

Related evidence is also consistent with, and may help explain, our results. Relative to the RPS30b gene, ACT1 shows a significantly different response to a number of spliceosome mutations. Most dramatically, in response to a temperature-sensitive PRP19 mutation, the level of spliced, mature ACT1 mRNA is dramatically reduced (with no change in pre-mRNA level), while RPS30b shows only a strong increase in pre-mRNA (Pleiss et al., 2007). This suggests a difference in the limiting steps of splicing for each of these introns. While a defect in Prp19p (but not other splicing factors, interestingly) impairs the removal of the ACT1 intron, the pool of pre-mRNA is not affected. This may be consistent with a predominance of pre-spliceosome-associated ACT1 mRNA, with a highly active Prp19 complex required for efficient processing of the small subset of mRNA proceeding to splicing. In contrast, the RPS30b intron, which is predominantly associated with

Figure 4.11. Hypothetical model of kinetic differences in splicing of the Act1 and Rps30b introns. Due to its relative enrichment of U1 and U2 snRNPs, the Act1 intron appears to be rate-limited in its transition to the precatalytic form of the spliceosome (Complex B). The Rps30b intron is instead more strongly associated with components of this complex. More efficient procession to Complex B could result from this intron's specific recruitment of Npl3, known to promote recruitment of certain splicing factors.


Act1 Intron
Rps30b Intron
Complex E
Complex A
Complex B
Catalysis
(Complexes B^act and C)
U1
Npl3
U2
U4
19
U6
U5

the catalytic spliceosome, continues to be processed even with a defective Prp19 complex, presumably due to the large pool of mRNA consistently associated with this complex. More thorough investigation of the dynamics of splicing for these and other introns will be necessary to conclusively determine the factors differentiating their processing.

Additional differences of note were observed between the ACT1 and RPS30b introns. Interestingly, only the RPS30b intron was found to co-isolate with Npl3p. This SR protein is primarily known for its role in mRNA export, but has also been shown to promote splicing for a subset of intron-containing genes by aiding recruitment of other splicing factors (Kress et al., 2008). Consistent with our results, efficient splicing of RPS30b, but not ACT1, was found to depend on Npl3p activity. As Npl3p was only detected in the presence of the RPS30b intron, its enhancement of splicing activity appears to depend on intron-specific association. Conceivably, the inability of the ACT1 intron to recruit Npl3p may be associated with its distinctive pattern of early spliceosome association (Figure 4.11). Finally, as expected, Nab2p was found at significantly higher levels in the RPS30b intron sample. This is consistent with studies showing that pre-mRNA of RPS30b and other ribosome protein genes is specifically regulated by the exosome in a Nab2p-dependent manner (Schmid et al., 2012).

The differences in splicing factor occupancy we found in our initial isolations of Cbp80p-associated ACT1 or RPS30b introns suggest a number of avenues for future investigation. First, as Npl3p appears to be a key mRNP factor with a role in splicing specific to the RPS30b intron, this is a promising protein to target in future sequential purifications of intron-containing transcripts. Such mRNA isolations could provide additional evidence relating to the mechanism of Npl3p's apparent activation of splicing. As Npl3p has a number of different roles in multiple steps of mRNA processing, this RNA-specific approach is uniquely able to differentiate

interactions specific to RPS30b intron processing from those involved in generic mRNA maturation. In addition, given the broad differences we observed in introns' enrichment of early versus late splicing factors, a more comprehensive and detailed view of the divergent splicing-related factors could be obtained through sequential isolations using representative spliceosome proteins. Snu71p and Snu114p, for instance, would enrich either the U1-dominated prespliceosome or the catalytic spliceosome, respectively. This more precise targeting would allow us to isolate specific stages of splicing, giving a finer picture of where intron-specific protein factors act during the splicing process. This should indicate, for instance, whether the difference we observe in ACT1's relative enrichment of early splicing factors is only a kinetic phenomenon – in this case a purification with Snu114p should purify the same catalytic spliceosome as one obtained with the RPS30b intron. Differential isolation of intron-specific proteins like Npl3p in these experiments could also suggest where these factors might act to generate differences in splicing activity.

# Chapter 5: Discussion

### *Development and applications of new nanobodies*

The novel mass spectrometry and DNA sequencing-based approach to nanobody discovery that we developed has a range of long-term applications. Our method is well-suited to the development of nanobody reagents against various types of protein targets, including proteins that are difficult to tag. The versatility and potential of nanobodies is huge, as reflected by the interest of the research community (Huang et al., 2010; Romer et al., 2011; Vanlandschoot et al., 2011; Muyldermans, 2013). Nanobodies have great potential in drug development, as they are much smaller than antibodies, resistant to aggregation, and can be readily humanized (Revets et al., 2005; Vincke et al., 2009; Muyldermans, 2013); they can bind with great specificity and efficacy to disease targets such as tumor cells, either independently (as a monomer or an ultra-high affinity nanobody dimer), or as a fusion with other protein domains, molecules, or drugs (Els Conrath et al., 2001; Jahnichen et al., 2010; Roovers et al., 2011; Ulrichts et al., 2011). As demonstrated here, the ability of our method to quickly and easily identify large repertoires of high affinity bacterially-expressed nanobodies against a chosen target antigen has the potential to significantly advance a field that otherwise can take months or years to generate such reagents (Muyldermans, 2013).

Nanobodies have a number of unique properties that make them well-suited for certain other applications. In microscopy, their small size allows for improved permeability into cell and tissue samples, improving performance, for instance, in super-resolution studies (Ries et al., 2012). These reagents have further potential to improve other techniques reliant on efficient fluorescent labeling. Volume imaging of tissue samples is a particularly demanding technique, as immunolabeling and clearing of these large structures can take weeks with standard antibodies (Renier et al., 2014). Fluorescently labeled nanobodies, being an order of magnitude smaller, are

a promising alternative for these studies, and preliminary results have indeed indicated their suitability for more effective large tissue labeling. An additional advantage of nanobodies' small size and monovalency is their ability to bind their target at an equimolar ratio. When covalently conjugated with dye, this means that their fluorescent signal will increase linearly with the amount of target protein. This allows for quantitative labeling of samples, as we have observed in preliminary experiments (with the M. Tessier-Lavigne lab). Finally, in all microscopy applications, the ability of multiple nanobodies to simultaneously bind separate epitopes on their antigen means that total signal intensity can be correspondingly increased. With our anti-GFP nanobody repertoire, we have found that up to three nanobodies with distinct epitopes can be used to label GFP samples without effect on each other's binding, leading to three-fold higher signal.

More broadly, our pipeline for nanobody identification has allowed us to simultaneously immunize llamas with many antigens at once, and successfully generate several repertoires concurrently. While we have only begun taking advantage of this ability, we are already significantly increasing the output of nanobodies against many new targets of interest. As a start, we have generated several nanobodies against three proteins of recent biomedical interest: the cancer-associated CTLA4 and EPHA2 receptors, and dystrophy-linked dysferlin protein. CTLA4 and EPHA2 nanobodies are being actively assessed for *in vivo* effects on their targets' signaling pathways, and dysferlin nanobodies are undergoing validation. Given the medical importance of these three proteins, biologically active nanobodies can potentially be of immediate diagnostic or therapeutic use. Moving beyond these three targets, with which we have found early success, we are actively pursuing nanobodies against novel targets of interest. This includes BTLA and HVEM, two receptors expressed by a variety of lymphocytes, normally involved in inhibiting T cell activation when they interact with each other (Gavrieli et al., 2006). Similar to CTLA4 however, changes in

expression of these two receptors are common features of certain types of cancer, resulting in suppression of the immune response to these tumors (Pasero and Olive, 2013). Antibodies to these two proteins blocking their interactions could thus be valuable therapeutic tools, and we are currently working to generate nanobodies for this purpose.

Alongside our efforts on immediately relevant biomedical targets, we are also developing nanobodies against a number of other biologically significant proteins. Given our lab's interest in the structure and function of the nuclear pore complex, we have begun generating nanobodies against a variety of nucleoporins. This has started with immunizations of 11 different mammalian proteins, with a particular focus on those factors implicated in cancer biology (Simon and Rout, 2014). Lastly, we are aiming to develop nanobody reagents against several mammalian RNA processing factors, as well as key retrotransposition factors (Taylor et al., 2013).

In sum, the ability to rapidly generate sets of nanobodies has allowed us to produce reagents enabling diverse avenues of research. We can generate nanobodies against more than ten antigens in parallel with limited labor, making this an efficient and widely available pipeline. The nanobodies identified can be readily produced in unlimited quantities, and are immediately useful for affinity capture or related techniques. Moreover, the recombinant proteins are easily modified to allow for myriad other uses, from straightforward dye labeling for fluorescent imaging, to protein engineering of customized nanobodies designed for specialized applications.

***Tools for the study of single mRNP complexes***

For our investigations into yeast mRNP processing pathways, high efficiency anti-GFP nanobodies were one element enabling the development of a new approach to the isolation of RNA-specific pools of RNP complexes. With the use of highly efficient isolations of affinity tags, we are able

to perform robust purifications of MS2-tagged RNAs associated with any RNP factor of interest. This has allowed us to begin investigating the composition of mRNP complexes associated with the many different alternative processing pathways suspected to regulate mRNA transcripts. Until now, only the highest abundance RNP factors have been successfully identified through isolations of tagged RNAs (Said et al., 2009; Hogg and Goff, 2010; Tsai et al., 2011), in contrast to the many specialized proteins we have identified in our two-step approach. Aside from greater overall sensitivity, the use of a tagged RNP factor as a first purification target allows analysis of a specific step or mechanism of mRNA processing.

We have started to take advantage of this novel RNP purification approach in a preliminary survey of RNA sequence elements of interest. In addition to confirming the efficacy of our approach by identifying known protein interactors of well-studied mRNA elements like the ASH1 transcript's 3' UTR, our survey has already begun to provide insight into potentially novel mRNP interactions. For example, isolations of transcripts containing either ACT1 or RPS30b introns has revealed significant differential enrichment in various subcomplexes of the spliceosome. While the ACT1 intron predominantly isolates early prespliceosome factors, the RPS30b intron shows significant enrichment of catalytic spliceosomal proteins, along with specialized activators of splicing like Npl3p (Kress et al., 2008).

While previous studies in yeast have provided strong evidence that certain introns, particularly those of ribosomal protein genes, require different factors for efficient splicing, or are spliced at different rates, these conclusions have largely relied on indirect observations (Pleiss et al., 2007; Kress et al., 2008; Schmid et al., 2012). This has typically involved observing changes in the relative amounts of spliced and unspliced transcript in various mutant strains. While instructive, this experimental approach provides only indirect genetic insight into the mechanisms behind these

intron-specific differences. Through proteomic analysis of isolated intron-containing transcripts, we have now directly identified changes in the composition of mRNP components specific to different types of introns. While still preliminary, this data has been consistent with the previous genetic work on these introns, and has provided insight into potential mechanisms involved in transcript-specific differences in splicing.

***Potential applications of sequential RNP purifications***

Beyond the set of mCherry derivatives that we used for our preliminary survey of mRNA sequence elements, there are countless other questions of RNA biology that could be studied with variations of our sequential purification methodology. An obvious next step is to tag and isolate wholly endogenous transcripts. Given a moderate level of transcriptional activity, our approach is likely to allow equally efficient purifications of any transcript – there is nothing in our system uniquely tailored to the mCherry sequence so far used. There are many questions that could be addressed by isolations of such RNAs. As it is unclear how a coding sequence itself does or does not affect mRNP composition, this could be examined by comparing the RNPs co-purified with different types of mRNA sequence. Transcripts of different lengths, for instance, could be examined to identify factors whose stoichiometry are or are not dependent on transcript size. Evidence from genome-wide studies suggests that different RNPs may in fact have different length-dependence (Baejen et al., 2014). Alternatively, transcripts expressed at different levels could be similarly investigated, with a potential focus on codon adaption index, which is closely associated with a gene's protein expression level (Sharp and Li, 1987; Jansen et al., 2003). Beyond these questions of RNPs preferentially associated with certain broad transcript characteristics, any specific mRNA believed to have unique RNP components of interest can be easily tagged and purified with our sequential method. More and more such transcripts continue to be identified by various groups,

and there remain many such outstanding candidates: heat shock proteins with unique RNA-level regulation, or transcripts regulated during the cell cycle are two examples of many (Saavedra et al., 1997; Trcek et al., 2011).

Importantly, our system of RNA isolation is not specific to mRNA at all. As long as an MS2 hairpin tag can be inserted into a sequence of interest without undue effect on RNA structure, a fundamentally identical approach can be taken with any class of RNA, such as rRNAs or non-coding RNAs (ncRNAs). ncRNAs in particular have been of increasing interest to the scientific community, and comparatively little is known about processing specific to such sequences, particularly long ncRNAs (lncRNAs) (Mattick and Makunin, 2006; Gupta et al., 2010; Tuck and Tollervey, 2013). While often degraded in the nucleus by surveillance machinery, studies have suggested that certain ncRNA sequences are exported to the cytoplasm (Xu et al., 2009). Hrp1p and Nab2p have been implicated in these different outcomes through preliminary genome-wide studies (Tuck and Tollervey, 2013), and are thus prime candidates for targets in RNP-specific RNA isolations. Cytoplasmic factors can similarly be used to identify possible roles for these transcripts after export.

Finally, while our work in MS2-based RNA isolations has so far been limited to yeast, there is nothing preventing these techniques from being directly applied to other systems, such as mammalian cells. The MS2/MS2CP system itself has been extensively used in mammalian cell culture (Rodriguez et al., 2007; Querido and Chartrand, 2008), and our lab has optimized our protein tagging and affinity isolation approach to achieve results comparable to those in yeast in a multitude of model organisms, including mammalian tissue culture and even tissue (Domanski et al., 2012; Shi et al., 2015; Obado et al., 2016). As the sequential RNA purification method

fundamentally relies only on the combination of these two technologies, there is little reason to believe it cannot be easily adapted to any common laboratory model system.

***Conclusions***

The work presented here has aimed to provide a methodological framework first and foremost for the isolation of RNA-specific mRNPs. This involved identifying highly effective reagents for affinity isolation of protein tags, primarily PrA and GFP. In the process of generating such reagents in the form of high affinity nanobodies, we have also developed a pipeline for rapid identification of nanobody repertoires against any antigen of interest. This has allowed us to efficiently pursue sets of novel affinity reagents against a variety of targets of biomedical interest.

With improved nanobodies available, we have been able to develop a specific and efficient sequential affinity technique for isolating a single species of mRNA transcript associated with a given RNP factor. After confirming and optimizing the effectiveness of this approach, we began applying it to a small number of transcript sequence elements suspected to undergo unusual mRNA processing. While this survey is still in progress, initial results have revealed a number of transcript-specific changes in mRNP composition. Specifically, we have isolated confirmed factors associated with the well-studied Ash1 3'UTR, and identified potential new interacting factors. We have also found differences in patterns of spliceosome association with different introns, offering potential mechanistic insight into differences in splicing activity previously associated with these introns. Moreover, early experiments have suggested possible RNP factors responsible for mediating promoter regions' determination of their transcripts' downstream cellular fate. While more work remains to confirm and explore our preliminary results relating to

the biology of the mRNA transcripts we surveyed, the methods we have developed are already allowing routine investigation of RNA-specific RNP complexes. This will enable expansion of our survey to additional mRNAs or ncRNAs of biological interest.

# Chapter 6: Materials and Methods

## Nanobody generation and characterization

### *Isolation of $V_HH$ antibodies*

A 5 year old female llama, Barbie, was immunized with recombinant GFP-$His_6$, and a 4 year old male llama, Marley, with recombinant mCherry-$His_6$ through subcutaneous injections of 5 mg of protein with CFA. Three additional injections of 5 mg protein, with IFA, were performed at three week intervals. Serum bleeds were obtained 10 days after the final injection. 2.5 ml of serum was diluted ten-fold in 20 mM sodium phosphate, pH 7.0, and incubated with Protein G-agarose resin for 30 min. The flow-through was then incubated for 30 min with Protein A-agarose resin. Both resins were washed with 20 mM sodium phosphate, pH 7.0, and bound VHH IgG was eluted with 100 mM acetic acid, pH 4.0 and 500 mM NaCl (Protein G resin) or 100 mM acetic acid, pH 3.5 and 150 mM NaCl (Protein A resin). These elutions were pooled and dialyzed into PBS. 3 mg of this VHH fraction was then incubated with Sepharose-conjugated GFP. This resin was washed with 10 mM sodium phosphate, pH 7.4 and 500 mM NaCl, followed by 1-4.5 M $MgCl_2$ in 20 mM Tris, pH 7.5, and then equilibrated in PBS. The resin was then digested with 0.3 mg/ml papain in PBS plus 10 mM cysteine, for 4 hours at 37°C. The resin was then washed with 1) 10 mM sodium phosphate, pH 7.4 and 500 mM NaCl 2) PBS plus 0.1% Tween-20 3) PBS 4) 0.1 M $NH_4OAc$, 0.1 mM $MgCl_2$, 0.02% Tween-20. Bound protein was then eluted for 20 min with 0.1 M $NH_4OH$ and 0.5mM EDTA, pH 8.0. These elutions were dried down in a SpeedVac and resuspended in LDS plus 25 mM DTT. The samples were alkylated with iodoacetamide and run on a 4-12% Bis-Tris gel. The ~15 kDa band corresponding to the digested VHH region was then cut out and prepared for MS.

***RT-PCR and DNA sequencing***

Bone marrow aspirates were obtained from immunized llamas concurrent with serum bleeds. Bone marrow plasma cells were isolated on a Ficoll gradient using Ficoll-Paque (GE Healthcare). RNA was isolated from approximately 1-6 x $10^7$ cells using Trizol LS reagent (Thermo Fisher), according to the manufacturer's instructions. cDNA was reverse-transcribed using Ambion RETROscript (Thermo Fisher). A nested PCR was then performed with IgG specific primers. In the first step, CALL001 (5'-GTCCTGGCTGCTCTTCTACAAGG-3') and CALL002 (5'-GGTACGTGCTGTTGA ACTGTTCC-3') primers were used to amplify the IgG variable domain into the CH2 domain (Conrath et al., 2001). The approximately 600-750 bp band from VHH variants lacking a CH1 domain was purified on an agarose gel. Next, for 454 sequencing, VHH regions were specifically reamplified using framework 1- and 4-specific primers with 5' 454 adaptor sequences: 454-VHH-forward (5'-<u>CGTATCGCCTCCCTCGCGCCATCAG</u> ATGGCT[C/G]A[G/T]GTGCAGCTGGTGGAGTCTGG-3') and 454-VHH-reverse (5'-<u>CTATGCGCCTTGCCAGCCCGCTCAG</u> GGAGACGGTGACCTGGGT-3') (adaptor sequences are underlined) (Conrath et al., 2001). The approximately 400 bp product of this reaction was gel purified, then sequenced on a 454 GS FLX system after emPCR amplification, on one Pico Titer Plate. For Illumina MiSeq sequencing, the second PCR was instead performed with random 12-mers replacing adaptor sequences, to aid in cluster identification: MiSeq-VHH-forward (5'-NNNNNNNNNNNN ATGGCT[C/G]A[G/T]GTGCAGCTGGTGGAGTCTGG-3') and MiSeq-VHH-reverse (5'-NNNNNNNNNNNN GGAGACGGTGACCTGGGT-3'). The product of this PCR was gel purified, ligated to MiSeq adaptors before library preparation using Illumina kits, and run on a MiSeq sequencer with 2 x 300 bp paired end reads.

***Database preparation***

The protein sequence databases used for identification were prepared by translating sequencing reads in all 6 reading frames, and for each read the longest Open Reading Frame (ORF) was selected. The selected ORF was digested with trypsin *in silico* and a list of unique tryptic peptides of 7 amino acids or longer was constructed and saved in a FASTA file. It is important to construct a FASTA file only containing unique peptides because even though most search engines can handle some sequence redundancy, they are not well equipped to handle the extreme redundancy that is provided by next generation sequencing of the single chain antibody locus and search engines either become very slow or crash if presented with such an extreme redundancy.

***Mass spectrometry***

Gel sections containing $V_HH$ domains were excised, destained, and dehydrated. MS analysis was performed by Yinyin Li as follows. The dehydrated gel slices were subjected to in-gel digestion with proteomic-grade trypsin (80 µL; 25 ng trypsin, 25 mM ammonium bicarbonate) (Promega) at 37 °C overnight. The gel was extracted once with extraction solution (140 µL; 67% acetonitrile, 1.7 % formic acid). The resulting proteolytic digest was cleaned with a STAGE tip(Rappsilber et al., 2003) and loaded onto a home-packed reverse phase C18 column (75 µm I.D., 15 µm tip) (New Objective) with a pressurized bomb. The loaded peptides were subsequently separated with a linear gradient (0 % to 42 % acetonitrile, 0.5 % acetic acid, 120 min, 150 nL/min after flow splitting) generated by an Agilent 1260 HPLC and directly sprayed into an LTQ-Velos-Orbitrap mass spectrometer (Thermo Scientific) for analysis. In the mass spectrometer, a survey scan was carried out in the orbitrap (resolution = 30,000, AGC target = 1E6) followed by tandem MS in the ion trap (AGC target = 5E3) of the top twenty most intense peaks. Tandem MS was carried out with collision induced dissociation (isolation width = 2 Th, CE = 35 %, activation time = 5 ms). Internal

calibration was used for improved mass accuracy (lock mass m/z = 371.1012). In order to scan more peptides, both predictive AGC and dynamic exclusion were enabled (Repeat counts: 2, repeat duration: 12 s, exclusion duration: 60 s). Single and unassigned charge species were excluded from tandem MS scans. The raw files were converted into mzXML format with ReAdW (version 4.3.1).

***MS-based identification of VHH sequences***

Searching and ranking of candidate sequences was carried out by Yinyin Li and Sarah Keegan. The MS search was performed on the custom database of tryptic peptides using the X! Tandem search engine. Then, the identified peptides filtered by expectation value were mapped to the sequences translated from 454 reads (longest ORF only, as described above). The CDR regions were located within the sequence based on approximate position in the sequence and the presence of specific leading and trailing amino acids. For example, to locate the CDR3 region, the algorithm searched for the left anchor YXC (X representing any amino acid) between position 93 and 103 of the sequence, and the right anchor WG between position n-14 and n-4 of the sequence, where n is the length of the sequence. Once the peptides were mapped to the sequences and their CDR regions, a metric was calculated to rank each sequence as a potential candidate based on the bioinformatics evidence available. The factors included in the metric were: MS coverage and length of individual CDR regions with CDR3 carrying highest weight, overall coverage including framework region, and a count of the 454 reads producing the sequence. Finally, sequences with similar CDR3 regions were grouped together, allowing for the identification of the highest confidence sequence corresponding to a particular CDR3. A sequence was assigned to a group where its hamming distance to an existing member was 1, i.e. there was one amino acid difference in the sequence, and different groups that have one shared sequence were further combined. By choosing sequence hits from different groups for production, we maximized the overall sequence

diversity of the candidate pool. The candidate list was displayed for manual inspection as an interactive HTML page with CDR regions annotated, peptide mapping information and the ranking metrics shown for each sequence. All algorithms described above were implemented in Perl.

***Web-based application for nanobody sequence identification: "Llama-Magic"***

The pipeline that was used for identification of the Nanobody sequences has been automated and can be accessed through a web-based interface at http://www.llamamagic.org. *Llama-Magic* allows upload of FASTA files containing reads from High-throughput DNA sequencing. Once uploaded, the reads will be automatically translated and digested to create an MS searchable database of tryptic peptides, as described above. Next, the MS (mgf) files can be uploaded for a selected tryptic peptide sequence database, and the parent and fragment error can be chosen for the X! Tandem search. Once the mgf files are uploaded, the X! Tandem search will be executed and the matching peptides saved. Then (1) annotation of CDR regions, (2) mapping of the identified peptides and (3) ranking and grouping of candidates are performed automatically, producing an interactive display of the candidate list showing detailed information regarding each sequence and its corresponding rank. *Llama-Magic* is implemented in Perl, HTML and JavaScript. Manual inspection was performed to make sure a) long CDR3 peptides, which embrace both variable regions and framework regions, have fragmentation pattern within the variable regions; b) CDR3 peptides are unique enough (uniqueness score < 100);

***Cloning***

Nanobody sequences were codon-optimized for expression in *E. coli* and cloned into pCR2.1 after gene synthesis (Eurofins MWG Operon), incorporating BamHI and XhoI restriction sites at 5' and 3' ends, respectively. A pelB leader sequence was cloned into pET21b at NdeI and BamHI restriction sites using complementary primers: 5'-tatgaaatacttattgcctacggcagccgctggattgttattact

cgcggcccagccggccatggctg-3’ and 5’-gatccagccatggccggctgggccgcgagtaataacaatccagcggctgccgtag gcaataagtatttca-3'. Nanobody sequences were then subcloned into pET21b-pelB using BamHI and XhoI restriction sites, with primers also encoding a PreScission Protease cleavage site just before the C-terminal 6xHis tag.

***Purification of nanobodies***

pelB-fused nanobodies were expressed under a T7 promoter in Arctic Express (DE3) cells (Agilent), induced with IPTG at a final concentration of 0.1 mM. Cells were induced for 18-20 hours at 12°C, then pelleted by a 10 min spin at 5000 x g. The periplasmic fraction was then isolated by osmotic shock(Skerra and Pluckthun, 1988). This fraction was bound to His-Select nickel affinity resin (Sigma), washed with His wash buffer (20 mM sodium phosphate pH 8.0, 1 M NaCl, 20 mM imidazole), and eluted with His elution buffer (20 mM sodium phosphate pH 8.0, 0.5 M NaCl, 0.3 M imidazole). The elution was then dialyzed into PBS.

***Fluorescent protein binding assays***

2 µg of fluorescent protein was added to 50 µl of 2 mg/ml *E. coli* lysate diluted in binding buffer (20mM HEPES, pH 7.4, 350 mM NaCl, 0.01% Tween-20, 0.1 M PMSF, 3 µg/ml pepstatin A). This was incubated with 25 µl of nanobody-Dynabead slurry. After a 30 minute incubation at 4°C, beads were washed with binding buffer and bound protein was eluted with 15 µl LDS. Elutions were run on a 4-12% Bis-Tris gel.

**$K_d$** ***determinations***

SPR measurements were obtained on a Proteon XPR36 Protein Interaction Array System (Bio-Rad). Recombinant 1xproA, 2xproA, and 4xproA was immobilized on a ProteOn GLC sensor chip: the chip surface was first activated with 50 mM sulfo-NHS and 50 mM EDC, run at a flow-

rate of 30 µl/min for 300 sec. The ligand was then diluted to 2-6 µg/ml in 10 mM sodium acetate, pH 5.0, and injected at 25 µl/min for 75 sec. Finally, the surface was deactivated by running 1 M ethanolamine-HCl (pH 8.5) at 30 µl/min for 300 sec. This led to immobilization of approximately 200-600 response units (RU) of ligand.

$K_d$s of recombinant nanobodies were determined by injecting 4 or 5 concentrations of each protein, in triplicate, with a running buffer of 20 mM HEPES, pH 8.0 / 150 mM NaCl / 0.01% Tween. Proteins were injected at 50 µl/min for 120 sec, or 100 µl/min for 90 sec, followed by a dissociation time of 600 sec. Between injections, residual bound protein was eliminated by regeneration with 4.5 M $MgCl_2$ in 10 mM Tris, pH 7.5, run at 100 µl/min for 36 sec. Binding sensorgrams from these injections were processed and analyzed using the ProteOn Manager software. Binding curves were fit to the data with a Langmuir model, using grouped $k_a$, $k_d$, and $R_{max}$ values.

***Immunofluorescence microscopy***

HeLa cells were cultured on coverslips in DMEM media with 10% FBS and penicillin/streptomycin at 37°C with 8% $CO_2$ in a humidified environment. Cells tested negative for mycoplasma. Cells were transfected with CellLight Tubulin-GFP or Mitochondria-GFP BacMam 2.0 reagents (Thermo Fisher) using 4 µL of reagent per 5,000 cells, and processed after 18-20 hrs. Cells were fixed in ice-cold methanol for 10 minutes (for Tubulin-GFP) or in 2% paraformaldehyde for 10 minutes (for Mitochondria-GFP). Cells were permeabilized with 0.5% Triton for 10 min. and blocked for 1 hr with 10% goat serum / 1% BSA in PBS. They were then incubated for 1 hr at room temperature with recombinant nanobody conjugated to Alexa Fluor® 568 succinimidyl ester (Thermo Fisher), diluted to 100 ng/ml in 1% BSA in PBS. Cells were washed four times with PBS / 0.01% BSA, with 300 nM DAPI included in the final wash, then mounted with ProLong Diamond (Thermo Fisher).

Wild-type and Sec13-GFP tagged *T. brucei* strains were cultured to a cell density of $1x10^7$ as previously described (DeGrasse et al., 2009). Cells from each strain were mixed 1:1, and fixed for 10 minutes with cold 4% formaldehyde. Approximately $1x10^6$ cells were spotted onto coverslips, allowed to settle for 30 min., permeabilized with 0.1% Triton for 5 min., and blocked with 10% goat serum / 1% BSA in PBS for 30 minutes. Cells were then stained, washed, and mounted identically to HeLa cells.

A *S. cerevisiae* W303 strain with Htb2 genomically tagged at the C-terminus with mCherry was grown to mid-log phase, and allowed to settle on Concanavalin A-coated coverslips. Yeast were fixed in 4% paraformaldehyde / 2% sucrose / PBS, and blocked and permeabilized for 30 minutes in 0.25% Triton / 2% milk / PBS(Ries et al., 2012). Cells were stained overnight at 4°C with nanobody diluted to 3.3 µg/ml in 0.25% Triton / 1% BSA / PBS. They were then washed 5 times with 0.01% BSA in PBS, the final two washes for 5 min. Cells were mounted in 70% glycerol / PBS.

All images were obtained on a Deltavision Image Restoration Microscrope (Applied Precision/Olympus), with an Olympus 100x/1.40 numerical aperture objective, or 60x/1.42 objective in the case of HeLa cells. Raw images were processed by a deconvolution algorithm using softWorX software (Applied Precision/GE Healthcare).

***Affinity isolations of tagged protein complexes***

Recombinant nanobodies were conjugated to epoxy-activated magnetic Dynabeads (Thermo Fisher), with minor modifications to published IgG coupling conditions (Alber et al., 2007). 10 µg recombinant protein was used per 1 mg of Dynabeads, with conjugations carried out in 0.1 M sodium phosphate, pH 8.0 and 1 M ammonium sulfate, with an 18-20 hour incubation at 30°C.

Affinity isolations of yeast Nup84-GFP were carried out as previously described, using binding buffer consisting of 20 mM HEPES, pH 7.4, 500 mM NaCl, 2 mM $MgCl_2$, 0.1% CHAPS, 0.1M PMSF, and 3 µg/ml pepstatin A(Alber et al., 2007). For each experiment, 50 µl of bead slurry was used with 0.5 g of yeast cells. Similar conditions were used for HTB2-mCherry isolations (from yeast with HTB2 genomically tagged at the C-terminus with mCherry(Rout et al., 2000)), except lysate was sonicated 4 times for 10 s before centrifugation, and the binding buffer consisted of 20 mM HEPES, pH 8.0, 300 mM NaCl, 110 mM KOAc, 0.1% Tween-20, 0.1% Triton X-100, 0.1M PMSF, and 3 µg/ml pepstatin A. Isolations of RBM7-LAP from HeLa cells were performed as previously described(Domanski et al., 2012). 10 µl of bead slurry was used with 100 mg of cells, using a binding buffer of 20 mM HEPES, pH 7.4, 300 mM NaCl, 0.5% Triton X-100, with cOmplete Protease Inhibitor, EDTA-free (Roche).

To determine affinity isolation yields, samples of resuspended lysate were taken before and after Dynabead binding. These were run on a 4-12% Novex Bis-Tris gel in MES running buffer (Thermo Fisher), and probed by Western blotting using mouse anti-GFP antibody (Roche, cat. no. 11 814 460 001) diluted 1:1,000 in TBST / 2% dry milk and an anti-mouse, HRP-conjugated secondary (GE Healthcare, cat. no. NA931V) diluted 1:3,000 in TBST / 2% dry milk. Signals were quantified using ImageJ software.

***Fluorescence spectra***

Samples of recombinant GFP at 0.5 µM in PBS were mixed with either buffer or 10 µM of a LaG protein. Fluorescence spectra were obtained on a Synergy Neo microplate reader (BioTek). Excitation spectra from 300 nm to 530 nm were taken at an emission wavelength of 560 nm, and emission spectra were measured from 450 nm to 600 nm at an excitation wavelength of 425 nm.

***Phylogenetic analysis***

Phylogenetic trees and alignments were generated from LaG amino acid sequences using the Phylogeny.fr web service (Dereeper et al., 2008; Dereeper et al., 2010).

***Mapping of nanobody binding epitopes on GFP by NMR***

Three variants of GFP were used in the preparation of NMR samples. GFP-$His_6$ (eGFP), the variant used for immunization; GFPuv, the variant for which backbone $^{15}N$-$^{1}H$ chemical shift assignments were available from BMRB file 5666 (Khan et al., 2003) and a crystal structure was available from PBD ID 1B9C (Battistutta et al., 2000); GFPuv_A206K (GFPuv_M), a monomeric version of GFPuv (Zacharias et al., 2002).

All NMR samples contained between 500 and 20 µM $^{15}N$-GFP either alone or in the presence of a 1-1.2 molar excess of LaG, 10mM sodium phosphate buffer, pH 7.4, 150mM NaCl and 90% $H_2O$/10% $D_2O$. All NMR spectra (2D $^{1}H$-$^{15}N$ HSQC) were measured at 310K on a Bruker Avance DPX-600 MHz spectrometer equipped with a TCI cryoprobe. NMR experiments and analysis were performed by Ilona Nudelman.

Backbone $^{1}H$-$^{15}N$ assignments of GFPuv were obtained from a comparison between a $^{1}H$-$^{15}N$ HSQC spectrum of GFPuv alone and a simulated $^{1}H$-$^{15}N$ HSQC based on BMRB 5666 (Khan et al., 2003). Due to a very high similarity between the two, $^{1}H$-$^{15}N$ backbone assignment of GFPuv was obtained for 97% of $^{1}H$-$^{15}N$ backbone resonances for which assignment was available in BMRB5666. The accuracy of the GFPuv assignment was verified by mapping the binding site of a previously identified nanobody, GFP-Trap, on GFPuv. The crystal structure of the GFP/GFP-Trap complex is available in the PDB (PDB ID 3K1K)(Kirchhofer et al., 2010) and a comparison between the X-ray crystallography-derived binding site (obtained by analysis of 3K1K by PISA -

'Protein interfaces, surfaces and assemblies' service at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel and Henrick, 2007)) and the one determined by the chemical shift perturbation method, reveals they overlap, thereby confirming our assignment of GFPuv residues.

Backbone $^{1}H$-$^{15}N$ assignments of GFPuv_M were obtained from a comparison between a $^{1}H$-$^{15}N$ HSQC spectrum of GFPuv and that of GFPuv_M. Assignment was verified by mapping the dimerization site of GFPuv and comparing it to the crystal structure of PDB ID 1B9C(Battistutta et al., 2000) (analyzed for interacting residues using PISA(Krissinel and Henrick, 2007)).

All chemical shift differences were calculated using equation (1) where CSD is the total

$$CSD = \sqrt{\frac{\left(\frac{\Delta\delta N}{5}\right)^2 + \Delta\delta H^2}{2}} \quad (1)$$

chemical shift difference and ΔδN and ΔδH are the chemical shift differences in the free and bound states between the amide nitrogens and protons, respectively. The CSD cutoff for binding site residues was 0.05ppm for GFP-Trap binding site and for GFPuv dimerization site and 0.03ppm for all LaG binding sites.

## mRNP purifications

### *Cloning*

The MET3 promoter was subcloned from p404MET3 into pRS414 (from N. Buchler and F. Cross) using SacI and SpeI restriction sites to create pRS414-MET3. HA-MS2(ΔFG)-eGFP-SV40NLS was PCR-amplified from pMS2-GFP (Fusco et al., 2003) and cloned into pRS414-MET3 to create pRS414-MET3-MS2CP-GFP. Wild-type eGFP and mCherry sequences were PCR-amplified and

cloned into pRS426-TDH3 (Mumberg et al., 1995) using SpeI and XmaI restriction sites. Four repeats of the MS2 hairpin were inserted after these sequences between XmaI and XhoI sites, using complementary primers: 5'- ccgggCCTAATTGCCTAGAAAACATGAGGATCACCCATGT CTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGA AAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCC ATGTCTGCAGTATTCc-3' and 5'-tcgagGAATACTGCAGACATGGGTGATCCTCATGTTT TCTAGAGTCGACCTGCAGACATGGGTGATCCTCATGTTTTCTAGAGTCGACCTGCAG ACATGGGTGATCCTCATGTTTTCTAGAGTCGACCTGCAGACATGGGTGATCCTCATG TTTTCTAGGCAATTAGGc-3'.

ACT1 and RPS30b intron sequences were amplified from yeast w303 genomic DNA, and then blunt-end ligated into pRS426-mCherry-4xMS2 and pRS426-mCherry plasmids digested with MscI. HSP26 and GLC3 promoters were PCR amplified from yeast genomic. A 2x repeat sequence of Rap1p binding sites (5'GTGGTGCACAGATGTAACGTTCCAAAATGTATG GATGGTA-3') was synthesized and inserted immediately upstream of the TDH3 promoter. These promoters were subcloned into the pRS426-mCherry and pRS426-mCherry-4xMS2 plasmids using SacI and SpeI restriction sites. Finally, ASH1 and CTH2 3' UTR sequences were amplified from yeast genomic DNA and subcloned into pRS426-mCherry and pRS426-mCherry-4xMS2 plasmids after the 4xMS2 tag using XhoI and EagI restriction sites.

***Strains***

Strains used in this work were derived from wild-type W303 (*MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100*). NAB2, CBP80, CDC33, and THP2 were genomically tagged at the C-terminus using a PPX-Protein A::HIS5 cassette (Rout et al., 2000). The PPX sequence (GLEVLFQGPS) is the target of human rhinovirus 3C protease. For sequential RNA purifications,

pRS414-MS2CP-GFP was transformed into the appropriate strain, followed by transformation with the transcript-encoding pRS426 plasmid. Transformations were performed using standard lithium acetate techniques. Cells were grown at 30°C in complete synthetic defined medium lacking L-tryptophan, uracil, and L-methionine.

***Sequential RNA affinity purifications***

Yeast cells were grown, harvested at mid-log phase and cryogenically milled as previously described (Alber et al., 2007; Oeffinger et al., 2007). 2 g of frozen cell powder was resuspended in 8 ml of RNP buffer (20 mM HEPES, pH 7.4; 150 mM NaCl; 110 mM K-acetate; 1 mM DTT; 0.1% Tween 20; 0.5% Triton X-100; 1:5000 RNasin (Promega); 1:5000 Antifoam B; 1 mM PMSF; 3 µg/ml pepstatin A). Lysate was spun at 10,000 x g for 10 min., and the supernatant incubated with a 200 µl suspension (30 mg beads) of IgG-conjugated Dynabeads equilibrated in RNP buffer (Thermo Fisher). After a 30 min. incubation at 4°C, beads were washed 3 times with 1 ml RNP buffer, and resuspended in 150 µl RNP with 80 ng/µl PreScission Plus protease. This was incubated for 1 hr at 4°C, and the supernatant was collected. This was pooled with an additional 75 µl RNP wash of the beads, and the combined solution was incubated for 30 min with a 10 µl suspension (1.5 mg beads) of Dynabeads conjugated to LaG16-$G_4$S-LaG2 nanobody. After 3 washes with RNP buffer, protein was eluted at 72°C for 10 min. with 25 µl NuPAGE LDS buffer. When prepared for MS analysis, this elution was reduced for 10 min. at 72°C with 25 mM DTT, then alkylated with 0.1 M iodoacetamide for 30 min. at room temperature. For RNA isolation, beads were instead resuspended in 600 µl RLT buffer for RNeasy purification (Qiagen).

To determine yields, samples of resuspended lysate were taken before and after each binding step. These were run on a 4-12% Novex Bis-Tris gel in MES running buffer (Thermo Fisher), and probed by Western blotting using, for GFP, mouse anti-GFP antibody (Roche, cat. no. 11 814 460

001) diluted 1:1,000 in TBST / 2% dry milk and an anti-mouse, HRP-conjugated secondary (GE Healthcare, cat. no. NA931V) diluted 1:3,000 in TBST / 2% dry milk. For PrA, membranes were instead probed with peroxidase anti-peroxidase (Sigma, cat. No. P1291) diluted 1:5,000 in TBST. Signals were quantified using ImageJ software.

***Mass spectrometry-based proteomic analysis***

Alkylated affinity isolation elutions were purified as single gel plugs by SDS-PAGE: samples were run at 125 V for 7 min. on a 10% Bis-Tris gel (Thermo Fisher), which was then stained with Coomassie Blue dye. Bands were then excised as a single piece, destained, and dehydrated. In-gel digestion with trypsin was performed on destained dehydrated gel pieces overnight with 0.5-1 µg trypsin in 50 µl 50 mM ammonium bicarbonate (ABC) buffer with 10% acetonitrile (ACN). Tryptic peptides were extracted with POROS20 C18 beads (Krutchinsky et al., 2001), desalted on Zip Tips (0.6 µl C18 resin, Millipore), sequentially eluted with 40% and then 90% acetonitrile / 0.1% acetic acid, and the eluates pooled, concentrated by vacuum centrifugation, and suspended in 10 µl of 0.1% formic acid (FA). Desalted peptides (2 µl) were analyzed by direct injection onto a $C_{18}$ analytical column (EASY-Spray 3 µm, 150 mm x 0.075 mm) using a Thermo EASY-nLC 100 nanoLC coupled online to an Orbitrap Q Exactive Plus (QE+) or Fusion mass spectrometer (Thermo Fisher). Peptide separation and MS detection was performed over 60 min at ~300 nl/min using a discontinuous ACN gradient from 0 to 30 % mobile phase B over 60 min, followed by 30 to 100% B over 15 min (mobile phase A, 0.1% FA; mobile phase B, 97% ACN in 0.1% FA). The spray voltage was set at 1.9–2.3 kV. The instruments were operated in the data-dependent mode, in which the most abundant ions were fragmented by higher-energy collisional dissociation (HCD) (HCD collision energy 30 (QE+) or 35 (Fusion)), and analyzed in the Orbitrap mass analyzer. The target resolution for MS1 was 70,000 (QE+) and 120,000 (Fusion), and for MS2 was 17,500 (QE+)

and 30,000 (Fusion). Ions (300–1,700 *m/z* QE+, 400-1500 *m/z* Fusion) with a charge state of >1 were selected for fragmentation. A dynamic exclusion of 10 s (QE+) and 30 s (Fusion) and a 'lock mass' at 371.1012 Da were used.

***Mass spectrometry data processing and quantitation***

Combined MS/MS spectra were extracted from Thermo RAW files derived from three biological replicate samples, and searched by Proteome Discoverer/SEQUEST (version 2.0/2.1, Thermo Fisher) against a *Saccharomyces cerevisiae* protein sequence database and a common contaminants database. SEQUEST search parameters were as follows: full trypsin specificity with 2 missed cleavages, precursor and fragment tolerances of 20 ppm and 0.6 Da, fixed modification of carbamidomethylation of cysteine, and variable modification for methionine oxidation. Putative SEQUEST peptide-spectrum matches (PSMs) were scored by the Percolator algorithm (Spivak et al., 2009) to assign individual expectation scores (PEP). PSMs were filtered to achieve an estimated 1% peptide false discovery rate (q-value < 0.01) and then PSMs were assembled into protein groups using the "strict maximum parsimony principle" option. Relative label-free quantitation was performed using the Proteome Discoverer precursor ions area quantification workflow on the top 3 most intense peptides per protein. Results were exported to Excel for downstream analysis. Abundances for the three biological replicates were averaged. Proteins with fewer than three peptides were not analyzed.

***RNA-Seq and real-time PCR***

RNA from sequential affinity purifications was purified using the RNeasy kit (Qiagen). RNA was fragmented and ligated to adaptors before library preparation using Illumina kits, and run on a HiSeq 2500 sequencer with 50 bp single reads.

For qRT-PCR, RNA was reverse transcribed using SuperScript VILO (Thermo Fisher). Real-time PCR was then with using PowerUp SYBR Green mix (Thermo Fisher) on a QuantStudio 12K Flex system using standard conditions (Thermo Fisher). The following primer pairs were used to quantify each gene: mCherry, AGATCAAGCAGAGGCTGAAGCTGA and TGTGGGAGG TGATGTCCAACTTGA; GFP, TGGTGTTCAATGTTTTGCGAG and GCTCTGGTCTTGT AGTTACCG; 18S rRNA, GAGTCCTTGTGGCTCTTGGC and AATACTGATGCCCCCGACC (Merz et al., 2008).

***Northern hybridization***

RNA samples were separated on a 6% TBE-Urea acrylamide gel, and Northern blotting was performed as previously described (Tollervey, 1987). Briefly, gels were electrotransferred to Hybond N+ (GE Healthcare), and hybridized at 37°C overnight with probe diluted in a buffer of 50% formamide, 5x SSPE, 5x Denhardt's buffer, 0.2 mg/ml fish sperm DNA, and 0.2% SDS. Blots were then washed twice for 15 min. at 42°C with 6xSSPE before being exposed to a Phosphorimager screen. The following DNA oligomers were end-labeled with [$\gamma$-$^{32}$P]-ATP and used as probes: mCherry, TGTGGGAGGTGATGTCCAACTTGA; 18S rRNA, GACATGCATGGCTTAATCTTTGAGAC.

# Appendix: Improving and extending the nanobody pipeline

### *Improvements to the nanobody discovery pipeline*

Our initial efforts to development a new method for identifying nanobodies were based on GFP and mCherry antigens (Fridy et al., 2014). As the GFP tag in particular is a key part of our mRNP isolation protocol, and mCherry nanobodies could be used as an important control (Chapter 4), nanobodies against these targets were of the greatest practical interest to our lab. Given the success of this first application of our new approach, we have since sought to apply it to other antigens of biomedical interest, taking advantage of the relative ease and flexibility of the approach. In the process of producing nanobodies against new antigens, we have also made a variety of improvements in the methodology, allowing more reliable identification of nanobody repertoires against many antigens at once.

After GFP and mCherry, we first aimed to target three human proteins of biomedical interest: CTLA4, EPHA2, and dysferlin. CTLA4 is a T-cell receptor that inhibits the activation of cytotoxic T-cells; inhibition of CTLA4 thus enhances T-cell activation, and an inhibitory antibody, ipilimumab, has in fact been shown to promote anti-tumor immune responses in clinical trials of cancer patients (Leach et al., 1996; Hodi et al., 2010). Similarly, ephrin receptor A2 (EPHA2) has been implicated as a key oncogene in certain cancers, and inhibition of this receptor has been shown to suppress tumor growth (Oricchio et al., 2011; Udayakumar et al., 2011). The current use of monoclonal antibodies in these cancer therapeutic approaches is hindered by both high cost (e.g. $120,000 per course of ipilimumab treatment), and the difficulties of delivery into dense tumors (Couzin-Frankel, 2013; Maleki et al., 2013). Smaller and less complex nanobodies are expected to avoid some of these issues, and in fact a number of nanobodies are showing success as therapeutic agents (Van Bockstaele et al., 2009; Roovers et al., 2011).

Beyond these two cancer targets, we are also attempting to generate nanobody reagents against dysferlin. Mutations in this gene are associated with multiple forms of muscular dystrophy, such as Miyoshi myopathy and limb girdle muscular dystrophy (Barthelemy et al., 2011). The dysferlin protein is a very large (~240 kDa) transmembrane protein highly enriched in muscle cells; its functions are not fully understood, but it appears to play an important role in membrane maintenance and muscle repair. Due to its involvement in disease and the many questions related to its function, reagents able to target dysferlin's many domains have long been sought. To date, very few antibodies have been generated against dysferlin, possibly due to the difficulties in producing stable protein and subsequently raising a response against this structurally and proteolytically unstable target. We therefore aimed to generate a large repertoire of nanobodies against dysferlin, with the goal of improving the reagents available for research and diagnostic purposes. While some efforts have been made to produce nanobodies against dysferlin, so far only two clones have been generated using traditional phage display approaches (Huang et al., 2005).

We immunized animals with these three antigens, as part of a mixture with four other proteins of interest to collaborators. It has been reported that up to five antigens can be simultaneously injected into an animal with no adverse effect on the immune response, and informal accounts suggest that responses can even be generated against dozens of proteins at once (Pardon et al., 2014). This is perhaps not surprising, as animals in the wild are expected to be exposed to many natural invaders at once, each presenting multiple antigens, and must be able to mount simultaneous immune responses. As with GFP and mCherry, we immunized three llamas, but also injected two alpacas with dysferlin-GFP. It was reasoned that the smaller size of alpacas to llamas could result in a relatively stronger immune response to the same antigen dose, and other groups

have successfully generated nanobodies from alpacas with no changes in protocol (Rothbauer et al., 2006; Maass et al., 2007).

***Overcoming the limitations of papain digestion***

Sera from immunized animals showed strong activity against EPHA2, CTLA4, and dysferlin, and $V_HH$ fractions and $V_HH$ sequencing data was obtained as before from sera and bone marrow samples. However, in contrast to GFP and mCherry, we found that after affinity purification of antigen-specific $V_HH$ IgG, the papain digest used to release Fc fragments unavoidably led to complete proteolysis of immobilized EPHA2 and dysferlin. It is well established that papain cleaves polypeptides with very broad specificity, with only moderate preference for bulky hydrophobic residues at one amino acid position in its active site (Kimmel and Smith, 1954; Schechter and Berger, 1967). With careful titration however, papain preferentially targets the IgG hinge region, and has long been the standard protease used to cleave IgG into Fab and Fc fragments (Porter, 1959; Nisonoff et al., 1960; Parham, 1983). With GFP and mCherry, two highly stable and compact beta-barrel proteins (Ormo et al., 1996; Yarbrough et al., 2001), bound $V_HH$ IgG could be readily digested with papain to release Fc, in easily optimized conditions in which the immobilized antigen was not degraded. With less stable proteins however, it became clear that no digest conditions are available that allowed sufficient cleavage of IgG while avoiding excessive proteolysis of the antigen. With the antigen degraded, no specific $V_HH$ binders can remain on the resin to be analyzed.

To address this papain digestion problem, we first attempted to avoid the papain digestion step entirely. In theory, this should leave intact IgG on the antigen resin, complete with both Fab and Fc regions. When performing MS on these samples to identify nanobody candidates, the remaining highly conserved Fc region was expected to give abundant extraneous peptides that

could overwhelm detection of more diverse peptides from the $V_HH$ region, but with sufficient sensitivity, we anticipated adequate identification of the informative peptides. However, performing these non-digest purifications with anti-GFP $V_HH$ as a control, we consistently found that while positive candidates could be detected, the candidate list had far more non-specific clones throughout, making it difficult to identify positives without impractically extensive screening (Figure A.1). The reason for this increase in non-specific hits is unclear, but is presumably due to complications from additional peptides from either Fc domains, or from non-specific $V_HH$ regions that are no longer washed away as efficiently from the resin; as papain is no longer separating IgG molecules into monovalent variable domains, it is possible that the higher avidity of bivalent antibodies allows very weak binders to be retained. It is also possible that sets of "sticky" non-specific $V_HH$ binders always tend to associate with the antigen resin, but that papain could destabilize these by partial proteolysis, causing them to dissociate.

Due to the disadvantages of both papain and complete omission of digestion, we sought an alternative protease to achieve more specific IgG cleavage. A variety of papain alternatives have been suggested to offer more specific digestion of IgG hinge regions, including lysyl endopeptidase, V8 protease, or ficin (Mariani et al., 1991; Yamaguchi et al., 1995). More recently though, a protease with strong site specificity for a sequence found in certain IgG hinges was identified from *S. pyogenes*, known as IdeS (von Pawel-Rammingen et al., 2002). While this enzyme does not recognize every IgG subtype, it has no known non-IgG substrates, avoiding the risk of off target proteolysis in our method. Upon testing commercially available IdeS on llama $V_HH$ fractions, we found that more than 90% of IgG was specifically cleaved, with no over-digestion observed (with Junjie Wang, B. Chait lab) (Figure A.2a). Additionally, no other proteins or antigens we tested were cleaved by IdeS, in stark contrast to papain (Figure A.2b). We next

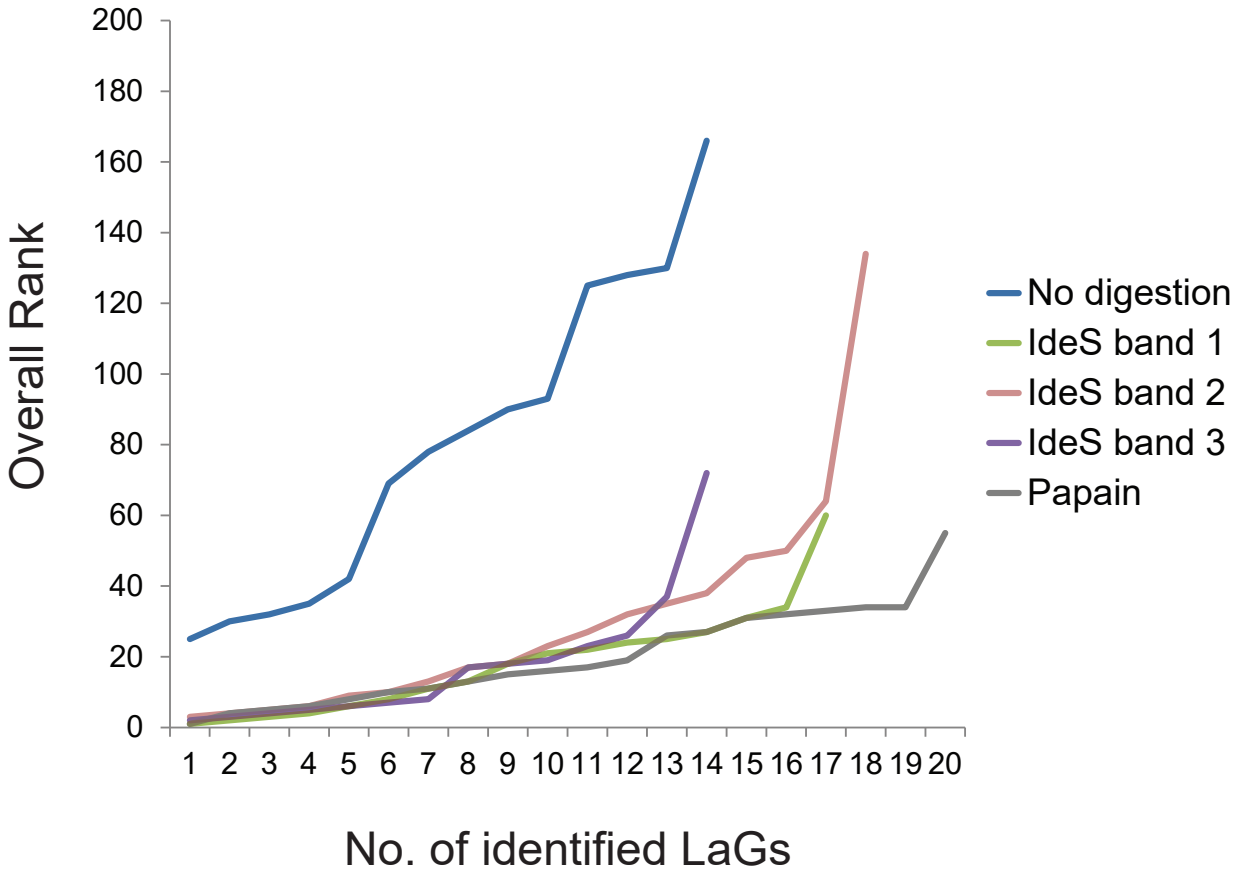
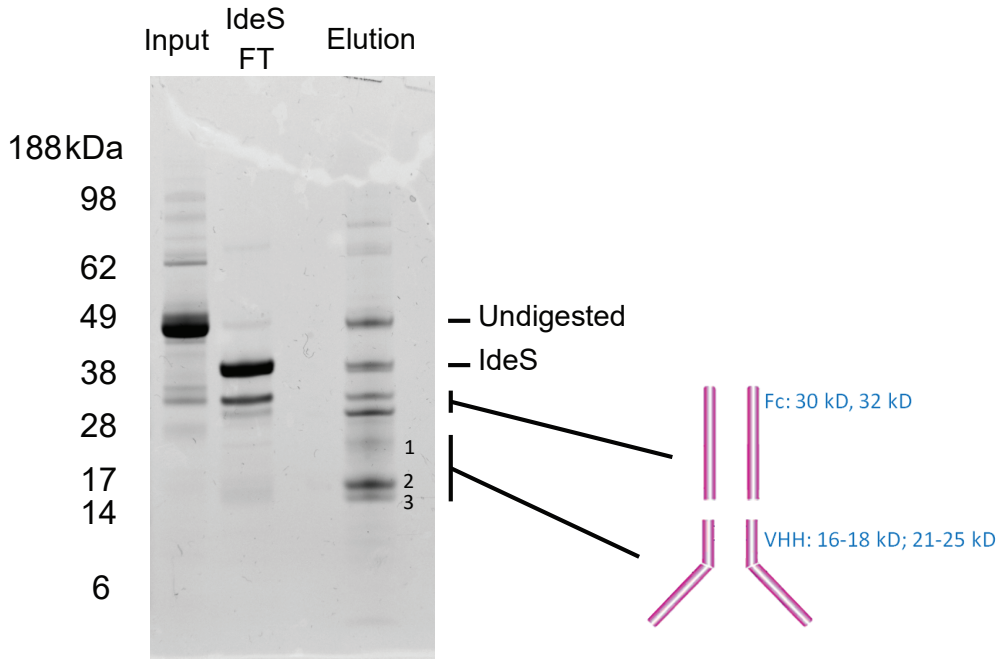


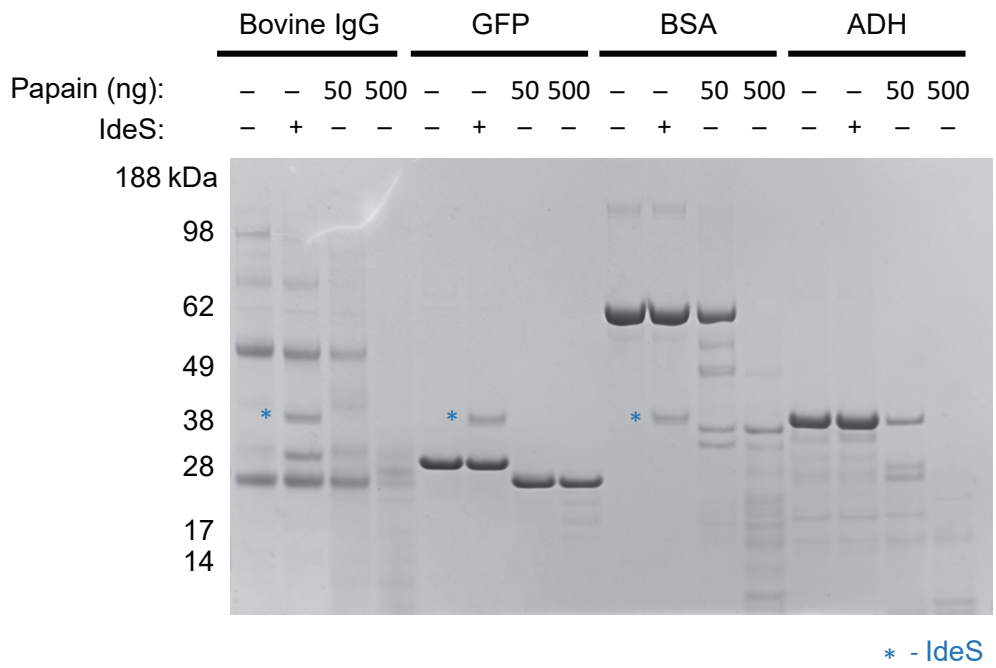

Figure A.1. Identification of verified anti-GFP nanobodies with papain, IdeS, or no digestion. $V_HH$ affinity-purified over GFP resin was digested in different ways before MS analysis to identify nanobody candidates. The resulting candidate lists were searched for the 26 verified LaG nanobody sequences. The overall number of candidates searched (y-axis) was then plotted against the corresponding number of LaGs identified (x-axis).

Figure A.2. Efficient and specific IgG digestion by IdeS. (a) A $V_HH$ IgG fraction was purified over GFP sepharose, then digested on resin with IdeS. Input, flow-through (FT) and ammonium hydroxide elutions were analyzed by SDS-PAGE. Fc fragments were found at the predicted MW of 30-32 kDa, and VHH fragments at 15-25 kDa. Variation in Fc and $V_HH$ sizes are consistent with short (IgG3) versus long (IgG2) hinge variants (Hamers-Casterman et al., 1993). Bands corresponding to long hinge (1) and short hinge (2 and 3) $V_HH$ were analyzed by MS (Figure 3.1). (b) IgG, GFP, bovine serum albumin (BSA), and yeast alcohol dehydrogenase (ADH) were digested with 0, 50, or 500 ng of papain, or 67 units of IdeS. All proteins were sensitive to papain, while only IgG was cleaved by IdeS.

**a**


Input
IdeS FT
Elution
188 kDa
98
62
49
38
28
17
14
6
1
2
3
Undigested
IdeS
Fc: 30 kD, 32 kD
VHH: 16-18 kD; 21-25 kD

**b**


Bovine IgG
GFP
BSA
ADH
Papain (ng): – – 50 500 – – 50 500 – – 50 500 – – 50 500
IdeS: – + – – – + – – – + – – – + – –
188 kDa
98
62
49
38
28
17
14
* - IdeS

tested IdeS-digested antigen-specific $V_HH$ by MS. Using GFP as a control antigen, we saw far fewer non-specific sequences in the resulting candidate lists when compared to non-digested samples. Confirmed GFP-binding sequences thus tended to rank much more highly, similar to our original papain results (Figure A.1).

As we were able to recapitulate our original papain results for GFP using IdeS, we have now reanalyzed IdeS-digested samples for the previously problematic CTLA4, EPHA2, and dysferlin antigens. With this new approach, we have successful identified many nanobody candidates, with positive hit rates averaging approximately 50% (Figure A.3). We are now determining the binding characteristics of these nanobodies, and in the case of CTLA4 and EPHA2 nanobodies, providing them to collaborators to assess *in vivo* inhibitory function.

***Nanobodies targeting Protein A***

In the course of generating nanobody repertoires against various antigens, we observed that approximately half of identified nanobody clones bound robustly to Protein A (PrA), consistent with previous studies, but somewhat counterintuitive given PrA's more commonly known affinity for IgG Fc domains (Frenken et al., 2000; Fridy et al., 2014). PrA and its derivatives continue to be widely used biological tools, due to their high affinity for the Fc domain of many antibodies. It is a commonly used protein tag, with multiple copies of its five homologous antibody-binding domains often used to increase overall avidity to IgG binding partners (Uhlen et al., 1983; Rigaut et al., 1999). Curiously, in addition to its well-known affinity for the Fc region, PrA has been found to bind certain immunoglobulins through their Fab domains using a distinct binding mechanism (Sasso et al., 1991; Jansson et al., 1998). It has similarly been observed that recombinant nanobodies, which are derived from camelid heavy chain-only $V_HH$ antibody

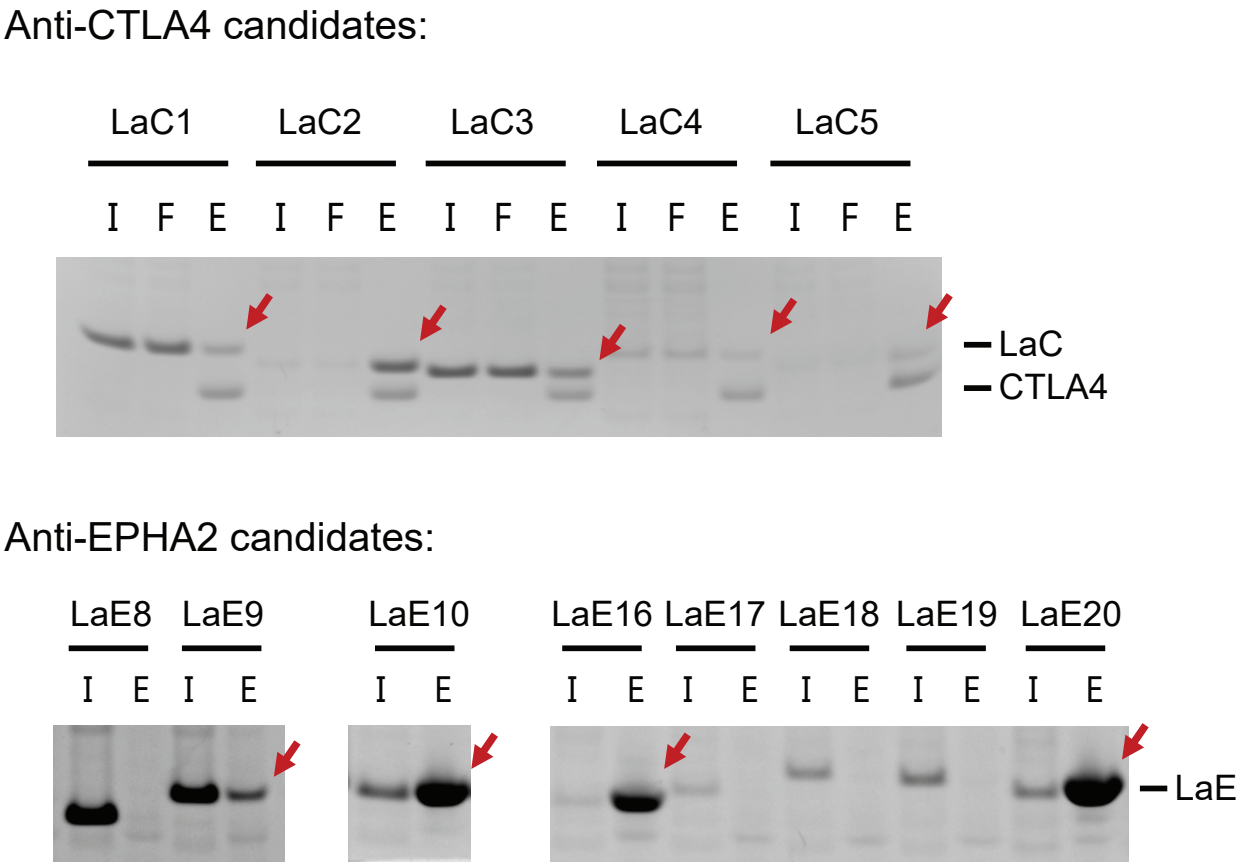


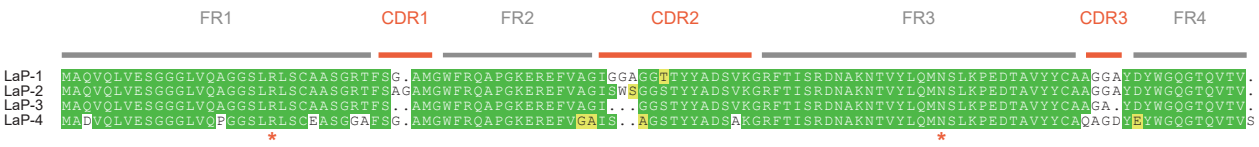

Figure A.3. Verification of nanobody candidates against CTLA4 and EPHA2. Llama antibodies against CTLA4 (LaCs) or EPHA2 (LaEs) were expressed in *E. coli*, and lysates were incubated with the corresponding antigen immobilized to sepharose resin, washed and eluted. Input (I), flow-through (F), and SDS elutions (E) were analyzed by SDS-PAGE. Five LaCs and four LaEs were found in elutions (red arrows), indicating positive antigen binding.

variants, can sometimes bind PrA through their Fab-like single variable domain (Frenken et al., 2000).

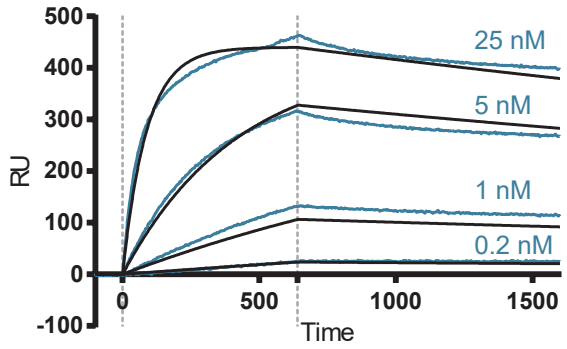
To date, standard purified IgG protein has been the reagent of choice for isolating PrA-tagged complexes (Lindmark et al., 1983; Moks et al., 1986), recently supplemented by an engineered high affinity affibody (Lindborg et al., 2013). However, the heterogeneous nature of IgG preparations, high molecular weight, lot-to-lot variation, and cost make these less than ideal for many applications, particularly in proteomics. Additionally, we have found that any particular recombinant protein used for affinity binding of a single target can display unpredictable off-target binding in certain systems, making it valuable to have more than one available option (Fridy et al., 2014). Given the well-established ease of use and versatility of nanobodies, we sought to generate such a reagent with high affinity for PrA.

When assessing the repertoire of nanobodies that we had previously identified, sequence alignments of PrA binders revealed highly conserved framework regions. These sequences all contained CDR loops specific for their original antigen, however, and to generate a nanobody with no extraneous binding activity, we synthesized four sequence variants derived from these consensus sequences with designed minimal or absent CDR loops (Figure A.4a). Once synthesized and expressed in *E. coli* with a periplasmic leader sequence, all four proteins (termed llama antibody against Protein A, or LaP 1-4) showed high levels of expression, and high solubility after periplasmic purification. These crude periplasm preparations were incubated with PrA-Sepharose, and all constructs displayed high affinity, particular LaP-1, which was selected for follow-up studies. In all cases, no detectable binding was observed for the original antigen (data not shown).

**a**




**b**




**c**

| | | 1xPrA | 2xPrA | 4xPrA |
|---|---|---|---|---|
| LaP-1 | $k_a$ ($M^{-1}s^{-1}$) | $9.9 \times 10^4$ | $7.7 \times 10^4$ | $1.6 \times 10^5$ |
| | $k_d$ ($s^{-1}$) | $1.0 \times 10^{-2}$ | $8.2 \times 10^{-3}$ | $1.3 \times 10^{-2}$ |
| | $K_D$ (nM) | 105±10 | 112±32 | 72±12 |
| 2xLaP-1 | $k_a$ ($M^{-1}s^{-1}$) | $2.9 \times 10^5$ | $4.5 \times 10^5$ | $4.0 \times 10^5$ |
| | $k_d$ ($s^{-1}$) | $3.6 \times 10^{-4}$ | $1.7 \times 10^{-4}$ | $1.4 \times 10^{-4}$ |
| | $K_D$ (nM) | 1.3±0.1 | 0.37±0.02 | 0.36±0.03 |

Figure A.4. Design and characterization of PrA nanobodies. (a) Highly conserved regions from multiple identified nanobody sequences were used as a framework for four engineered nanobodies against PrA (LaP-1-4), with varying minimal linkers used in place of existing CDR regions. Asterisks: residues whose mutation eliminates PrA binding. (b) SPR sensorgrams are shown for injections of multiple concentrations of 2xLaP-1 nanobody over immobilized 4xPrA. Curves fit from a Langmuir model are shown in black. (c) Binding constants determined by SPR for either LaP-1 nanobody or 2xLaP-1 dimerized nanobody against immobilized 1xPrA, 2xPrA, or 4xPrA. Corresponding association rates ($k_a$), dissociation rates ($k_d$), and dissociation constants ($K_d$) are shown.

Surface plasmon resonance (SPR) analysis was performed to determine the binding kinetics of LaP-1's interactions with multiple recombinant PrA constructs, containing 1, 2, or 4 repeats of the IgG-binding domain (Figure A.4c). Regardless of the number of PrA repeats, LaP-1 bound these proteins with a $K_D$ of 70-120 nM. While suitable for many applications, it was reasoned that this affinity could be increased by generating a homodimeric form of the nanobody. Two copies of the LaP-1 sequence were therefore fused using a glycine-rich peptide linker (3 repeats of GGGGS). As the LaP-1 monomer is only 13 kDa, even this dimerized form remains a relatively small 27 kDa. After purification, the PrA affinities of this 2xLaP-1 fusion protein were similarly assessed by SPR (Figure A.4b,c). When binding to 2xPrA or 4xPrA constructs, the affinity of the dimer was more than 300-fold stronger than the LaP-1 monomer, with a $K_D$ of 360-370 pM.

Given the structural similarity of the nanobody variable region to other mammalian Fab fragments, it was hypothesized that the LaP-1 nanobodies interacted with PrA through an analogous binding surface, rather than an Fc-like binding mechanism (Desmyter et al., 1996; Jansson et al., 1998). To test this, mutagenesis was done across the homologous sequences corresponding to this binding region (Figure A.5a) (Graille et al., 2000). Mutations in two residues, R21 and N85, eliminated PrA binding. These are both present in the homologous binding region, and expected to be necessary for PrA binding via an Fab-like interaction. A model of the proposed PrA:LaP-1 interaction was also generated via homology to a PrA:Fab crystal structure (PDB ID: 1DEE) (Graille et al., 2000) using the program I-TASSER (Zhang, 2008; Roy et al., 2010; Roy et al., 2012), and is consistent with the mutagenesis results (Figure A.5b).

To investigate the specificity and versatility of these anti-PrA nanobodies, we assessed their effectiveness in *ex vivo* affinity isolations of PrA-tagged protein complexes from yeast and bacteria. LaP-1 and 2xLaP-1 proteins were conjugated to magnetic beads and used to isolate tagged

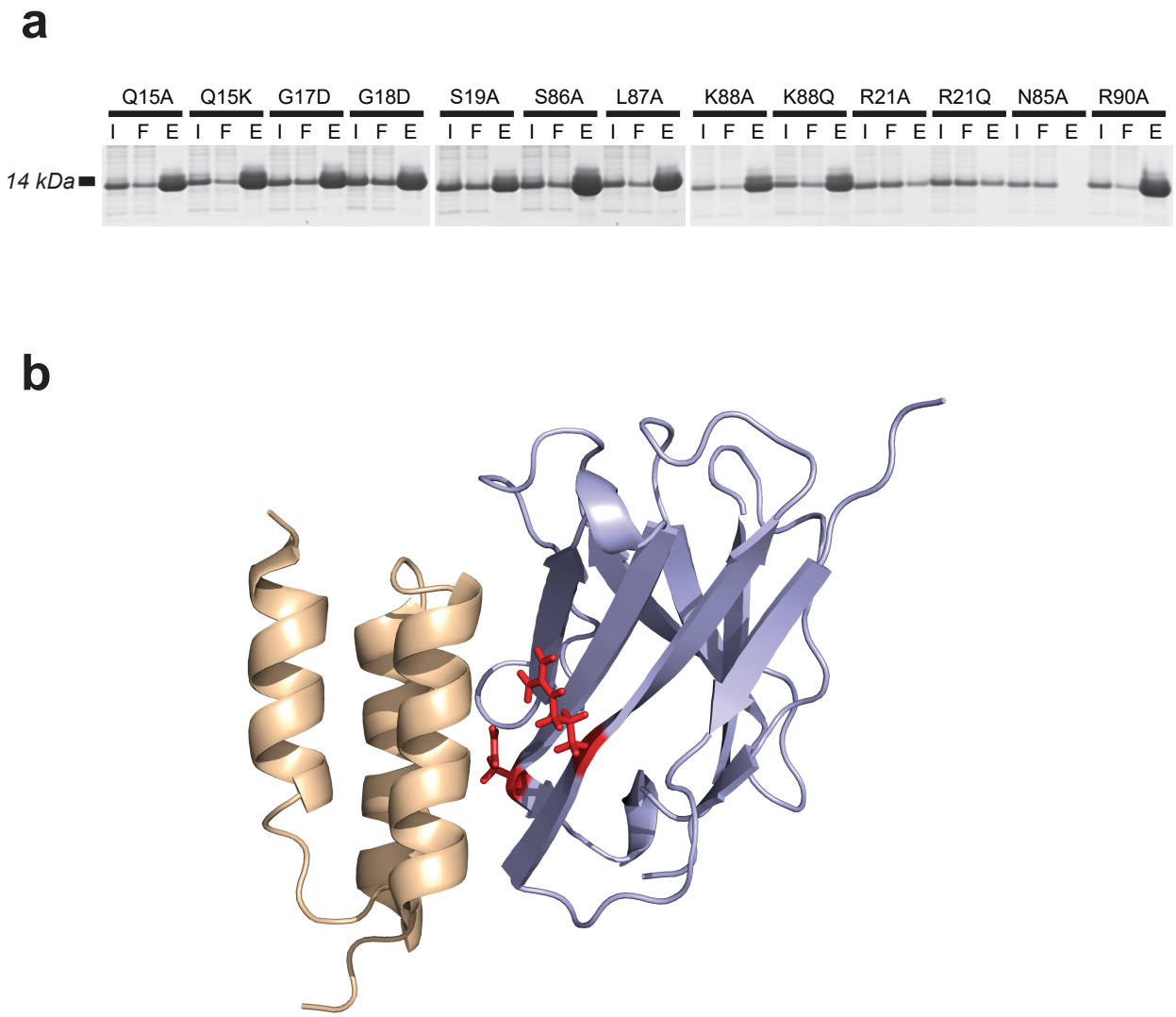
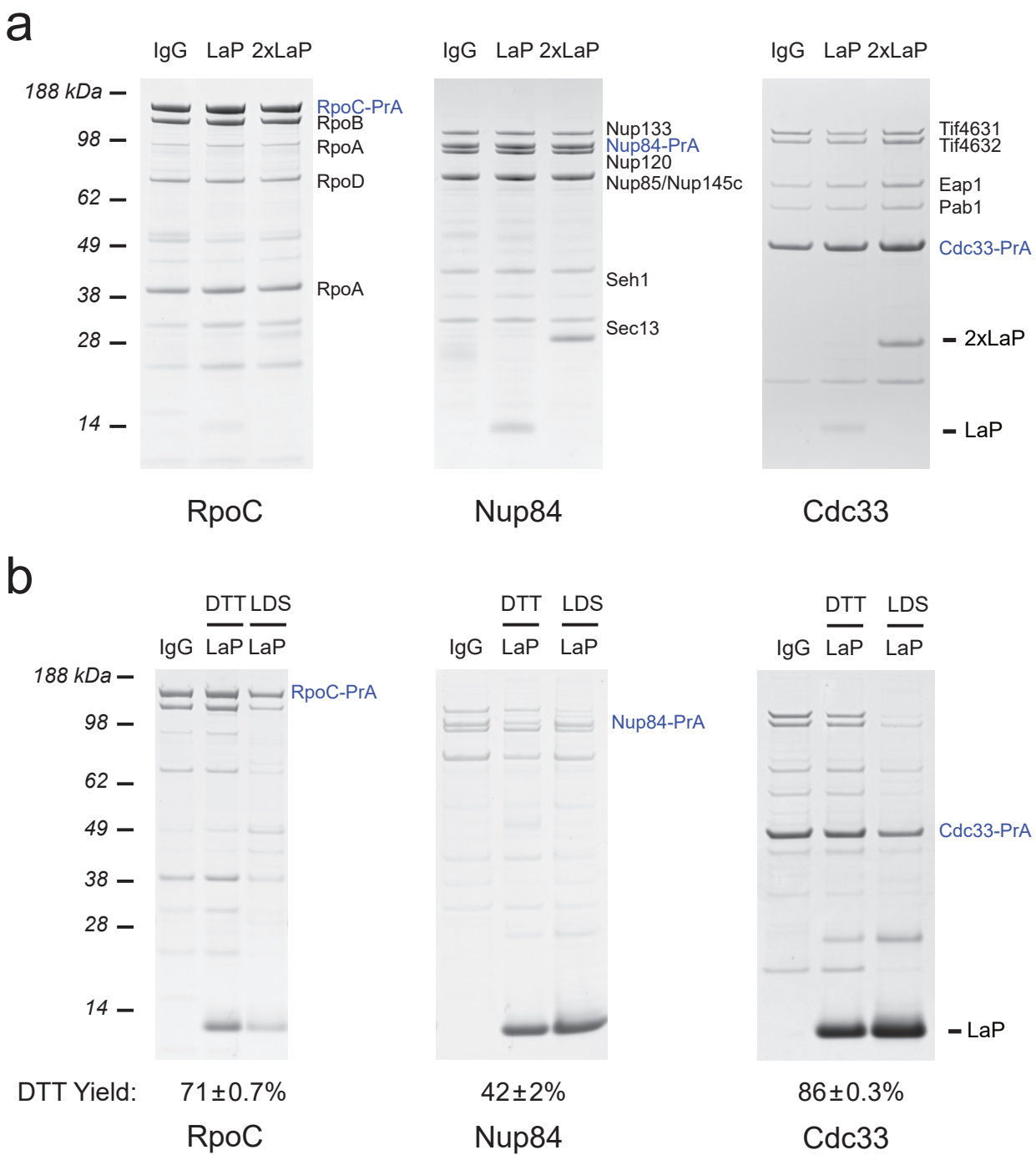



Figure A.5. Mapping of PrA:LaP-1 binding interface. (a) LaP-1 nanobodies with the designated point mutant were expressed in bacteria, and periplasmic extracts were incubated with PrA-Sepharose. For each mutant, input (I), flow-through (F), and elution (E) samples from the Sepharose binding were run by SDS-PAGE and Coomassie-stained. (b) The PrA:LaP-1 interaction was modeled using I-TASSER, via LaP-1's homology to Fab in a PrA:Fab crystal structure (PDB ID: 1DEE). The structure was visualized in PyMOL. Grey: LaP-1; Beige: PrA; Red: R21 and N85 residues, which are required for PrA binding.

RNA polymerase from *E. coli* (Westblade et al., 2008), the *S. cerevisiae* Nup84 subcomplex of the nuclear pore complex (NPC) (Brohawn et al., 2009; Fernandez-Martinez et al., 2012) and mRNA cytoplasmic cap binding complex (Goyer et al., 1989) (Figure A.6a). In all cases, both the monomeric and dimeric LaP-1 proteins were able to efficiently isolate the targeted complex with yield and purity comparable to control affinity isolations with polyclonal IgG, and negligible non-specific binding or contamination. The tandem affinity purification (TAP) tag is a frequently used PrA alternative, containing two artificial PrA Z domains (Rigaut et al., 1999). However, consistent with other studies showing weak binding between this Z domain and Fab fragments (Ljungberg et al., 1993; Jansson et al., 1998), our LaP-1 proteins are only able to recover very limited amounts of TAP-tagged protein in affinity isolations. The PrA Z domains in the TAP tag have only a single glycine to alanine point mutation however, so binding should be restored by reverting this mutation.

In addition to these original LaP-1 proteins, we generated constructs containing an additional free cysteine at the C-terminus. Thiol chemistry can thus be targeted to the C-terminus, which allowed us to reversibly immobilize LaP-1 to magnetic amine-coated Dynabeads using an N-Succinimidyl 3-(2-pyridyldithio)-propionate (SPDP) crosslinker with a 12 unit PEG spacer. This produces a disulfide-containing crosslink to the bead, which is easily cleavable in mild reducing conditions (Carlsson et al., 1975; Kiefer, 1975). Beads with LaP-1 immobilized in this manner were tested in isolations of RNA polymerase, Nup84, and Cdc33 as before. Again, yield and purity comparable to IgG was observed, and by a short incubation with 25 mM DTT, 40-90% of the isolated protein could be eluted (Figure A.6b). This level of elution efficiency is comparable to that seen in other native methods of PrA release (Rigaut et al., 1999; LaCava et al., 2013). The complex-to-complex difference in yield is also a consistent phenomenon, likely due to differences in interactions with

Figure A.6. Affinity isolations performed with LaP-1 nanobodies. (A) LaP-1, 2xLaP-1, or rabbit IgG were conjugated to epoxy-activated magnetic beads and used to isolate *E. coli* RpoC-PrA (an RNA polymerase subunit), as well as *S. cerevisiae* Nup84-PrA (an NPC subcomplex component) and Cdc33-PrA (a protein in the mRNA cytoplasmic cap binding complex). Bands previously determined by MS are labeled. (B) LaP-1 was reversibly immobilized to magnetic beads using SPDP crosslinker, and used for affinity isolations as in (A). Bound protein was first eluted in 25 mM DTT (DTT), before release of remaining bound protein in LDS (LDS). The yield of protein recovered in the DTT elution step, as compared to total eluted protein, is listed with the s.e.m. All experiments were performed in duplicate.


a
IgG LaP 2xLaP
188 kDa
98
62
49
38
28
14
RpoC-PrA
RpoB
RpoA
RpoD
RpoA
Nup133
Nup84-PrA
Nup120
Nup85/Nup145c
Seh1
Sec13
Tif4631
Tif4632
Eap1
Pab1
Cdc33-PrA
2xLaP
LaP
RpoC
Nup84
Cdc33
b
DTT LDS
IgG LaP LaP
188 kDa
98
62
49
38
28
14
RpoC-PrA
Nup84-PrA
Cdc33-PrA
LaP
DTT Yield:
71±0.7%
42±2%
86±0.3%
RpoC
Nup84
Cdc33

the bead surface. This approach does necessarily lead to release of both PrA-associated and unassociated LaP-1 nanobody in the elution. Due to the small size (13 kDa) of LaP-1 however, an additional purification step, such as gradient centrifugation, which is often necessary for sensitive downstream applications like electron microscopy, can easily remove unbound nanobody.

In either monomeric or dimeric form, the LaP-1 nanobody is suitable for highly efficient isolation of PrA-tagged protein targets. As a small recombinant protein, it is an especially flexible reagent, as shown in its use in native elutions of protein complexes, making it a convenient option for any application making use of a PrA-based tag.

# References


Abruzzi, K.C., Lacadie, S. and Rosbash, M., 2004, Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J 23, 2620-31.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Rout, M.P. and Sali, A., 2007, Determining the architectures of macromolecular assemblies. Nature 450, 683-94.

Alcazar-Roman, A.R., Tran, E.J., Guo, S. and Wente, S.R., 2006, Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. Nat Cell Biol 8, 711-6.

Alexander, R.D., Barrass, J.D., Dichtl, B., Kos, M., Obtulowicz, T., Robert, M.C., Koper, M., Karkusiewicz, I., Mariconti, L., Tollervey, D., Kufel, J., Bertrand, E. and Beggs, J.D., 2010, RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in Saccharomyces cerevisiae. RNA 16, 2570-80.

Alvarez-Rueda, N., Behar, G., Ferre, V., Pugniere, M., Roquet, F., Gastinel, L., Jacquot, C., Aubry, J., Baty, D., Barbet, J. and Birkle, S., 2007, Generation of llama single-domain antibodies against methotrexate, a prototypical hapten. Mol Immunol 44, 1680-90.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O. and Herschlag, D., 2003, Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 100, 3889-94.

Arbabi-Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. and Muyldermans, S., 1997, Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. FEBS letters 414, 521-6.

Arbabi-Ghahroudi, M., Tanha, J. and MacKenzie, R., 2005, Prokaryotic expression of antibodies. Cancer Metastasis Rev 24, 501-19.

Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. and Roberts, R.W., 2002, Designed arginine-rich RNA-binding peptides with picomolar affinity. J Am Chem Soc 124, 10966-7.

Bachler, M., Schroeder, R. and von Ahsen, U., 1999, StreptoTag: a novel method for the isolation of RNA-binding proteins. RNA 5, 1509-16.

Baejen, C., Torkler, P., Gressel, S., Essig, K., Soding, J. and Cramer, P., 2014, Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Mol Cell 55, 745-57.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B., 2007, Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 389, 1017-31.

Bardwell, V.J. and Wickens, M., 1990, Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. Nucleic Acids Res 18, 6587-94.

Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E. and Hentze, M.W., 2004, Using the lambdaN peptide to tether proteins to RNAs. Methods Mol Biol 257, 135-54.

Barthelemy, F., Wein, N., Krahn, M., Levy, N. and Bartoli, M., 2011, Translational research and therapeutic perspectives in dysferlinopathies. Mol Med 17, 875-82.

Batisse, J., Batisse, C., Budd, A., Bottcher, B. and Hurt, E., 2009, Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. J Biol Chem 284, 34911-7.

Battistutta, R., Negro, A. and Zanotti, G., 2000, Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein. Proteins 41, 429-37.

Becker, R.S. and Knight, K.L., 1990, Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. Cell 63, 987-97.

Benner, R., Hijmans, W. and Haaijman, J.J., 1981, The bone marrow: the major source of serum immunoglobulins, but still a neglected site of antibody formation. Clin Exp Immunol 46, 1-8.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H. and Long, R.M., 1998, Localization of ASH1 mRNA particles in living yeast. Mol Cell 2, 437-45.

Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S.M., Lee, T., Pope, S.H., Riordan, G.S. and Whitlow, M., 1988, Single-chain antigen-binding proteins. Science 242, 423-6.

Bobola, N., Jansen, R.P., Shin, T.H. and Nasmyth, K., 1996, Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. Cell 84, 699-709.

Bousquet-Antonelli, C., Presutti, C. and Tollervey, D., 2000, Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell 102, 765-75.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A. and Choder, M., 2011, Promoter elements regulate cytoplasmic mRNA decay. Cell 147, 1473-83.

Brohawn, S.G., Partridge, J.R., Whittle, J.R. and Schwartz, T.U., 2009, The nuclear pore complex has entered the atomic age. Structure 17, 1156-68.

Butter, F., Scheibe, M., Morl, M. and Mann, M., 2009, Unbiased RNA-protein interaction screen by quantitative proteomics. Proc Natl Acad Sci U S A 106, 10626-31.

Carey, J., Cameron, V., de Haseth, P.L. and Uhlenbeck, O.C., 1983, Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. Biochemistry 22, 2601-10.

Carlsson, J., Axen, R. and Unge, T., 1975, Reversible, covalent immobilization of enzymes by thiol-disulphide interchange. European journal of biochemistry / FEBS 59, 567-72.

Carrillo Oesterreich, F., Preibisch, S. and Neugebauer, K.M., 2010, Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Molecular cell 40, 571-81.

Casolari, J.M., Brown, C.R., Komili, S., West, J., Hieronymus, H. and Silver, P.A., 2004, Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. Cell 117, 427-39.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., Krijgsveld, J. and Hentze, M.W., 2012, Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149, 1393-406.

Chanarat, S., Seizl, M. and Strasser, K., 2011, The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. Genes Dev 25, 1147-58.

Chao, J.A., Patskovsky, Y., Almo, S.C. and Singer, R.H., 2008, Structural basis for the coevolution of a viral RNA-protein complex. Nat Struct Mol Biol 15, 103-5.

Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A. and Reed, R., 2006, Human mRNA export machinery recruited to the 5' end of mRNA. Cell 127, 1389-400.

Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S., 1997, mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev 11, 3319-26.

Churchman, L.S. and Weissman, J.S., 2011, Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368-73.

Conrath, K.E., Lauwereys, M., Galleni, M., Matagne, A., Frere, J.M., Kinne, J., Wyns, L. and Muyldermans, S., 2001, Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae. Antimicrob Agents Chemother 45, 2807-12.

Cordingley, M.G., Register, R.B., Callahan, P.L., Garsky, V.M. and Colonno, R.J., 1989, Cleavage of small peptides in vitro by human rhinovirus 14 3C protease expressed in Escherichia coli. J Virol 63, 5037-45.

Cortez-Retamozo, V., Backmann, N., Senter, P.D., Wernery, U., De Baetselier, P., Muyldermans, S. and Revets, H., 2004, Efficient cancer therapy with a nanobody-based conjugate. Cancer Res 64, 2853-7.

Cotten, M., Oberhauser, B., Brunar, H., Holzner, A., Issakides, G., Noe, C.R., Schaffner, G., Wagner, E. and Birnstiel, M.L., 1991, 2'-O-methyl, 2'-O-ethyl oligoribonucleotides and phosphorothioate oligodeoxyribonucleotides as inhibitors of the in vitro U7 snRNP-dependent mRNA processing event. Nucleic Acids Res 19, 2629-35.

Couzin-Frankel, J., 2013, Breakthrough of the year 2013. Cancer immunotherapy. Science 342, 1432-3.

Cristea, I.M., Williams, R., Chait, B.T. and Rout, M.P., 2005, Fluorescent proteins as proteomic probes. Molecular & cellular proteomics : MCP 4, 1933-41.

Daigle, N. and Ellenberg, J., 2007, LambdaN-GFP: an RNA reporter system for live-cell imaging. Nat Methods 4, 633-6.

DeGrasse, J.A., DuBois, K.N., Devos, D., Siegel, T.N., Sali, A., Field, M.C., Rout, M.P. and Chait, B.T., 2009, Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. Molecular & cellular proteomics : MCP 8, 2119-30.

Dereeper, A., Audic, S., Claverie, J.M. and Blanc, G., 2010, BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol 10, 8.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M. and Gascuel, O., 2008, Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic acids research 36, W465-9.

Desmyter, A., Transue, T.R., Ghahroudi, M.A., Thi, M.H., Poortmans, F., Hamers, R., Muyldermans, S. and Wyns, L., 1996, Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Biol 3, 803-11.

Dieppois, G., Iglesias, N. and Stutz, F., 2006, Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes. Mol Cell Biol 26, 7858-70.

Dieppois, G. and Stutz, F., 2010, Connecting the transcription site to the nuclear pore: a multi-tether process that regulates gene expression. J Cell Sci 123, 1989-99.

Dolman, N.J., Kilgore, J.A. and Davidson, M.W., 2013, A review of reagents for fluorescence microscopy of cellular compartments and structures, part I: BacMam labeling and reagents for vesicular structures. Curr Protoc Cytom Chapter 12, Unit 12 30.

Domanski, M., Molloy, K., Jiang, H., Chait, B.T., Rout, M.P., Jensen, T.H. and LaCava, J., 2012, Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. Biotechniques 0, 1-6.

Dong, S., Li, C., Zenklusen, D., Singer, R.H., Jacobson, A. and He, F., 2007, YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol Cell 25, 559-73.

Dorner, T. and Radbruch, A., 2007, Antibodies and B cell memory in viral immunity. Immunity 27, 384-92.

Dumoulin, M., Conrath, K., Van Meirhaeghe, A., Meersman, F., Heremans, K., Frenken, L.G., Muyldermans, S., Wyns, L. and Matagne, A., 2002, Single-domain antibody fragments with high conformational stability. Protein Sci 11, 500-15.

Els Conrath, K., Lauwereys, M., Wyns, L. and Muyldermans, S., 2001, Camel single-domain antibodies as modular building units in bispecific and bivalent antibody constructs. J Biol Chem 276, 7346-50.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H. and Luhrmann, R., 2009, The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell 36, 593-608.

Fernandez-Martinez, J., Phillips, J., Sekedat, M.D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J.D., Williams, R., Stokes, D.L., Chait, B.T., Sali, A. and Rout, M.P., 2012, Structure-function mapping of a heptameric module in the nuclear pore complex. The Journal of Cell Biology 196, 419-34.

Fink, D., Wohrer, S., Pfeffer, M., Tombe, T., Ong, C.J. and Sorensen, P.H., 2010, Ubiquitous expression of the monomeric red fluorescent protein mCherry in transgenic mice. Genesis 48, 723-9.

Fortes, P., Inada, T., Preiss, T., Hentze, M.W., Mattaj, I.W. and Sachs, A.B., 2000, The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. Molecular cell 6, 191-6.

Franklin, N.C., 1985, Conservation of Genome Form but Not Sequence in the Transcription Antitermination Determinants of Bacteriophage-Lambda, Bacteriophage-Phi-21 and Bacteriophage-P22. Journal of Molecular Biology 181, 75-84.

Frenken, L.G., van der Linden, R.H., Hermans, P.W., Bos, J.W., Ruuls, R.C., de Geus, B. and Verrips, C.T., 2000, Isolation of antigen specific llama VHH antibody fragments and their high level secretion by Saccharomyces cerevisiae. J Biotechnol 78, 11-21.

Fridy, P.C., Li, Y., Keegan, S., Thompson, M.K., Nudelman, I., Scheid, J.F., Oeffinger, M., Nussenzweig, M.C., Fenyo, D., Chait, B.T. and Rout, M.P., 2014, A robust pipeline for rapid production of versatile nanobody repertoires. Nat Methods 11, 1253-60.

Fridy, P.C., Thompson, M.K., Ketaren, N.E. and Rout, M.P., 2015, Engineered high-affinity nanobodies recognizing staphylococcal Protein A and suitable for native isolation of protein complexes. Anal Biochem 477, 92-4.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S.M., Blanchard, J.M., Singer, R.H. and Bertrand, E., 2003, Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol 13, 161-7.

Futcher, A.B. and Cox, B.S., 1984, Copy number and the stability of 2-micron circle-based artificial plasmids of Saccharomyces cerevisiae. J Bacteriol 157, 283-90.

Gallouzi, I.E. and Steitz, J.A., 2001, Delineation of mRNA export pathways by the use of cell-permeable peptides. Science 294, 1895-901.

Gallwitz, D. and Sures, I., 1980, Structure of a split yeast gene: complete nucleotide sequence of the actin gene in Saccharomyces cerevisiae. P Natl Acad Sci USA 77, 2546-50.

Galy, V., Gadal, O., Fromont-Racine, M., Romano, A., Jacquier, A. and Nehrbass, U., 2004, Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. Cell 116, 63-73.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B. and Superti-Furga, G., 2006, Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631-6.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. and Superti-Furga, G., 2002, Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141-7.

Gavrieli, M., Sedy, J., Nelson, C.A. and Murphy, K.M., 2006, BTLA and HVEM cross talk regulates inhibition and costimulation. Adv Immunol 92, 157-85.

Geiger, T., Cox, J. and Mann, M., 2010, Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. Mol Cell Proteomics 9, 2252-61.

Georgescu, J., Rehm, T., Wiehler, J., Steipe, B. and Holak, T.A., 2003, Backbone H(N), N, C(alpha) and C(beta) assignment of the GFPuv mutant. Journal of biomolecular NMR 25, 161-2.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S., 2003, Global analysis of protein expression in yeast. Nature 425, 737-41.

Gilbert, W. and Guthrie, C., 2004, The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. Mol Cell 13, 201-12.

Gilbert, W., Siebel, C.W. and Guthrie, C., 2001, Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. RNA 7, 302-13.

Gingras, A.C., Aebersold, R. and Raught, B., 2005, Advances in protein complex analysis using mass spectrometry. J Physiol 563, 11-21.

Gingras, A.C., Raught, B. and Sonenberg, N., 1999, eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem 68, 913-63.

Goldflam, M., Tarrago, T., Gairi, M. and Giralt, E., 2012, NMR studies of protein-ligand interactions. Methods in molecular biology 831, 233-59.

Gornemann, J., Kotovic, K.M., Hujer, K. and Neugebauer, K.M., 2005, Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell 19, 53-63.

Goyer, C., Altmann, M., Trachsel, H. and Sonenberg, N., 1989, Identification and characterization of cap-binding proteins from yeast. The Journal of biological chemistry 264, 7603-10.

Graille, M., Stura, E.A., Corper, A.L., Sutton, B.J., Taussig, M.J., Charbonnier, J.B. and Silverman, G.J., 2000, Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. P Natl Acad Sci USA 97, 5399-404.

Greenberg, A.S., Avila, D., Hughes, M., Hughes, A., McKinney, E.C. and Flajnik, M.F., 1995, A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374, 168-73.

Grunwald, D. and Singer, R.H., 2010, In vivo imaging of labelled endogenous beta-actin mRNA during nucleocytoplasmic transport. Nature 467, 604-7.

Grunweller, A., Wyszko, E., Bieber, B., Jahnel, R., Erdmann, V.A. and Kurreck, J., 2003, Comparison of different antisense strategies in mammalian cells using locked nucleic acids, 2'-O-methyl RNA, phosphorothioates and small interfering RNA. Nucleic Acids Res 31, 3185-93.

Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K. and Izaurralde, E., 1998, TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Mol Cell 1, 649-59.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R.B., van de Vijver, M.J., Sukumar, S. and Chang, H.Y., 2010, Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071-6.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E.B., Bendahman, N. and Hamers, R., 1993, Naturally occurring antibodies devoid of light chains. Nature 363, 446-8.

Harmsen, M.M. and De Haard, H.J., 2007, Properties, production, and applications of camelid single-domain antibody fragments. Appl Microbiol Biotechnol 77, 13-22.

Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M. and Luhrmann, R., 2002, Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A 99, 16719-24.

Heim, R., Prasher, D.C. and Tsien, R.Y., 1994, Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A 91, 12501-4.

Hereford, L.M. and Rosbash, M., 1977, Number and distribution of polyadenylated RNA sequences in yeast. Cell 10, 453-62.

Hieronymus, H. and Silver, P.A., 2003, Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat Genet 33, 155-61.

Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. and Jensen, T.H., 2001, Quality control of mRNA 3'-end processing is linked to the nuclear exosome. Nature 413, 538-42.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D. and Tyers, M., 2002, Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415, 180-3.

Hocine, S., Raymond, P., Zenklusen, D., Chao, J.A. and Singer, R.H., 2013, Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nature methods 10, 119-21.

Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., Akerley, W., van den Eertwegh, A.J., Lutzky, J., Lorigan, P., Vaubel, J.M., Linette, G.P., Hogg, D., Ottensmeier, C.H., Lebbe, C., Peschel, C., Quirt, I., Clark, J.I., Wolchok, J.D., Weber, J.S., Tian, J., Yellin, M.J., Nichol, G.M., Hoos, A. and Urba, W.J., 2010, Improved survival with ipilimumab in patients with metastatic melanoma. N Engl J Med 363, 711-23.

Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. and Brown, P.O., 2008, Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 6, e255.

Hogg, J.R. and Collins, K., 2007, RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. RNA 13, 868-80.

Hogg, J.R. and Goff, S.P., 2010, Upf1 senses 3'UTR length to potentiate mRNA decay. Cell 143, 379-89.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S. and Young, R.A., 1998, Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95, 717-28.

Huang, L., Muyldermans, S. and Saerens, D., 2010, Nanobodies(R): proficient tools in diagnostics. Expert Rev Mol Diagn 10, 777-85.

Huang, Y., Verheesen, P., Roussis, A., Frankhuizen, W., Ginjaar, I., Haldane, F., Laval, S., Anderson, L.V., Verrips, T., Frants, R.R., de Haard, H., Bushby, K., den Dunnen, J. and van der Maarel, S.M., 2005, Protein studies in dysferlinopathy patients using llama-derived antibody fragments selected by phage display. Eur J Hum Genet 13, 721-30.

Hurt, E., Luo, M.J., Rother, S., Reed, R. and Strasser, K., 2004, Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. Proc Natl Acad Sci U S A 101, 1858-62.

Hurt, E., Strasser, K., Segref, A., Bailer, S., Schlaich, N., Presutti, C., Tollervey, D. and Jansen, R., 2000, Mex67p mediates nuclear export of a variety of RNA polymerase II transcripts. J Biol Chem 275, 8361-8.

Huston, J.S., Levinson, D., Mudgetthunter, M., Tai, M.S., Novotny, J., Margolies, M.N., Ridge, R.J., Bruccoleri, R.E., Haber, E., Crea, R. and Oppermann, H., 1988, Protein Engineering of Antibody-Binding Sites - Recovery of Specific Activity in an Anti-Digoxin Single-Chain Fv Analog Produced in Escherichia-Coli. P Natl Acad Sci USA 85, 5879-5883.

Hutchins, J.R., Toyoda, Y., Hegemann, B., Poser, I., Heriche, J.K., Sykora, M.M., Augsburg, M., Hudecz, O., Buschhorn, B.A., Bulkescher, J., Conrad, C., Comartin, D., Schleiffer, A., Sarov, M., Pozniakovsky, A., Slabicki, M.M., Schloissnig, S., Steinmacher, I., Leuschner, M., Ssykor, A., Lawo, S., Pelletier, L., Stark, H., Nasmyth, K., Ellenberg, J., Durbin, R., Buchholz, F., Mechtler, K., Hyman, A.A. and Peters, J.M., 2010, Systematic analysis of human protein complexes identifies chromosome segregation proteins. Science 328, 593-9.

Iglesias, N. and Stutz, F., 2008, Regulation of mRNP dynamics along the export pathway. FEBS Lett 582, 1987-96.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C. and Stutz, F., 2010, Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev 24, 1927-38.

Inoue, H., Hayase, Y., Imura, A., Iwai, S., Miura, K. and Ohtsuka, E., 1987, Synthesis and hybridization studies on two complementary nona(2'-O-methyl)ribonucleotides. Nucleic acids research 15, 6131-48.

Jahnichen, S., Blanchetot, C., Maussang, D., Gonzalez-Pajuelo, M., Chow, K.Y., Bosch, L., De Vrieze, S., Serruys, B., Ulrichts, H., Vandevelde, W., Saunders, M., De Haard, H.J., Schols, D., Leurs, R., Vanlandschoot, P., Verrips, T. and Smit, M.J., 2010, CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells. Proc Natl Acad Sci U S A 107, 20565-70.

Jansen, R., Bussemaker, H.J. and Gerstein, M., 2003, Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. Nucleic Acids Res 31, 2242-51.

Jansen, R.P., Dowzer, C., Michaelis, C., Galova, M. and Nasmyth, K., 1996, Mother cell-specific HO expression in budding yeast depends on the unconventional myosin myo4p and other cytoplasmic proteins. Cell 84, 687-97.

Jansson, B., Uhlen, M. and Nygren, P.A., 1998, All individual domains of staphylococcal protein A show Fab binding. FEMS Immunol Med Microbiol 20, 69-78.

Jimeno, S., Rondon, A.G., Luna, R. and Aguilera, A., 2002, The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. The EMBO journal 21, 3526-35.

Jorgensen, P., Edgington, N.P., Schneider, B.L., Rupes, I., Tyers, M. and Futcher, B., 2007, The size of the nucleus increases as yeast cells grow. Mol Biol Cell 18, 3523-32.

Kafri, M., Metzl-Raz, E., Jona, G. and Barkai, N., 2016, The Cost of Protein Production. Cell Rep 14, 22-31.

Khan, F., Stott, K. and Jackson, S., 2003, 1H, 15N and 13C backbone assignment of the green fluorescent protein (GFP). Journal of biomolecular NMR 26, 281-2.

Kiefer, H., 1975, Separation of antigen-specific lymphocytes. A new general method of releasing cells bound to nylon mesh. Eur J Immunol 5, 624-8.

Kim Guisbert, K., Duncan, K., Li, H. and Guthrie, C., 2005, Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. RNA 11, 383-93.

Kimmel, J.R. and Smith, E.L., 1954, Crystalline papain. I. Preparation, specificity, and activation. The Journal of biological chemistry 207, 515-31.

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C., Leonhardt, H., Hopfner, K.P. and Rothbauer, U., 2010, Modulation of protein properties in living cells using nanobodies. Nature structural & molecular biology 17, 133-8.

Knight, K.L., 1992, Restricted VH gene usage and generation of antibody diversity in rabbit. Annu Rev Immunol 10, 593-616.

Kohler, A. and Hurt, E., 2007, Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol 8, 761-73.

Kotovic, K.M., Lockshon, D., Boric, L. and Neugebauer, K.M., 2003, Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Molecular and cellular biology 23, 5768-79.

Kozak, M. and Nathans, D., 1972, Translation of the genome of a ribonucleic acid bacteriophage. Bacteriol Rev 36, 109-34.

Kress, T.L., Krogan, N.J. and Guthrie, C., 2008, A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol Cell 32, 727-34.

Krissinel, E. and Henrick, K., 2007, Inference of macromolecular assemblies from crystalline state. Journal of molecular biology 372, 774-97.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A. and Greenblatt, J.F., 2006, Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440, 637-43.

Krutchinsky, A.N., Kalkum, M. and Chait, B.T., 2001, Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. Anal Chem 73, 5066-77.

Kumar, R., Singh, S.K., Koshkin, A.A., Rajwanshi, V.K., Meldgaard, M. and Wengel, J., 1998, The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA. Bioorg Med Chem Lett 8, 2219-22.

Lacadie, S.A. and Rosbash, M., 2005, Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell 19, 65-75.

LaCava, J., Chandramouli, N., Jiang, H. and Rout, M.P., 2013, Improved native isolation of endogenous Protein A-tagged protein complexes. Biotechniques 54, 213-6.

Larsen, H.J., Bentin, T. and Nielsen, P.E., 1999, Antisense properties of peptide nucleic acid. Biochim Biophys Acta 1489, 159-66.

Larson, D.R., Singer, R.H. and Zenklusen, D., 2009, A single molecule view of gene expression. Trends Cell Biol 19, 630-7.

Leach, D.R., Krummel, M.F. and Allison, J.P., 1996, Enhancement of antitumor immunity by CTLA-4 blockade. Science 271, 1734-6.

Lei, E.P. and Silver, P.A., 2002, Intron status and 3'-end formation control cotranscriptional export of mRNA. Genes & development 16, 2761-6.

Lewis, J.D., Gorlich, D. and Mattaj, I.W., 1996, A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. Nucleic Acids Res 24, 3332-6.

Lewis, J.D. and Izaurralde, E., 1997, The role of the cap structure in RNA processing and nuclear export. Eur J Biochem 247, 461-9.

Lim, F., Downey, T.P. and Peabody, D.S., 2001, Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. The Journal of biological chemistry 276, 22507-13.

Lindborg, M., Dubnovitsky, A., Olesen, K., Bjorkman, T., Abrahmsen, L., Feldwisch, J. and Hard, T., 2013, High-affinity binding to staphylococcal protein A by an engineered dimeric Affibody molecule. Protein Eng Des Sel 26, 635-44.

Lindmark, R., Thoren-Tolling, K. and Sjoquist, J., 1983, Binding of immunoglobulins to protein A and immunoglobulin levels in mammalian sera. J Immunol Methods 62, 1-13.

Ljungberg, U.K., Jansson, B., Niss, U., Nilsson, R., Sandberg, B.E. and Nilsson, B., 1993, The interaction between different domains of staphylococcal protein A and human polyclonal IgG, IgA, IgM and F(ab')2: separation of affinity from specificity. Mol Immunol 30, 1279-85.

Long, R.M., Gu, W., Lorimer, E., Singer, R.H. and Chartrand, P., 2000, She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to ASH1 mRNA. EMBO J 19, 6592-601.

Lopes, T.S., Klootwijk, J., Veenstra, A.E., van der Aar, P.C., van Heerikhuizen, H., Raue, H.A. and Planta, R.J., 1989, High-copy-number integration into the ribosomal DNA of Saccharomyces cerevisiae: a new vector for high-level expression. Gene 79, 199-206.

Lund, M.K. and Guthrie, C., 2005, The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. Mol Cell 20, 645-51.

Maass, D.R., Sepulveda, J., Pernthaner, A. and Shoemaker, C.B., 2007, Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs). J Immunol Methods 324, 13-25.

Maleki, L.A., Baradaran, B., Majidi, J., Mohammadian, M. and Shahneh, F.Z., 2013, Future prospects of monoclonal antibodies as magic bullets in immunotherapy. Hum Antibodies 22, 9-13.

Mariani, M., Camagna, M., Tarditi, L. and Seccamani, E., 1991, A new enzymatic method to obtain high-yield F(ab)2 suitable for clinical use from mouse IgGl. Mol Immunol 28, 69-77.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R., 2005, Recruitment of the human TREX complex to mRNA during splicing. Genes Dev 19, 1512-7.

Mattick, J.S. and Makunin, I.V., 2006, Non-coding RNA. Hum Mol Genet 15 Spec No 1, R17-29.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. and Lukyanov, S.A., 1999, Fluorescent proteins from nonbioluminescent Anthozoa species. Nature biotechnology 17, 969-73.

Merz, K., Hondele, M., Goetze, H., Gmelch, K., Stoeckl, U. and Griesenbeck, J., 2008, Actively transcribed rRNA genes in S. cerevisiae are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. Genes Dev 22, 1190-204.

Moks, T., Abrahmsen, L., Nilsson, B., Hellman, U., Sjoquist, J. and Uhlen, M., 1986, Staphylococcal protein A consists of five IgG-binding domains. European journal of biochemistry / FEBS 156, 637-43.

Mumberg, D., Muller, R. and Funk, M., 1995, Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156, 119-22.

Muyldermans, S., 2013, Nanobodies: Natural Single-Domain Antibodies. Annual Review of Biochemistry 82, 775-97.

Muyldermans, S., Baral, T.N., Retamozzo, V.C., De Baetselier, P., De Genst, E., Kinne, J., Leonhardt, H., Magez, S., Nguyen, V.K., Revets, H., Rothbauer, U., Stijlemans, B., Tillib, S., Wernery, U., Wyns, L., Hassanzadeh-Ghassabeh, G. and Saerens, D., 2009, Camelid immunoglobulins and nanobody technology. Vet Immunol Immunopathol 128, 178-83.

Neri, D., Momo, M., Prospero, T. and Winter, G., 1995, High-affinity antigen binding by chelating recombinant antibodies (CRAbs). Journal of molecular biology 246, 367-73.

Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L. and Weissman, J.S., 2006, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441, 840-6.

Niedner, A., Muller, M., Moorthy, B.T., Jansen, R.P. and Niessing, D., 2013, Role of Loc1p in assembly and reorganization of nuclear ASH1 messenger ribonucleoprotein particles in yeast. P Natl Acad Sci USA 110, E5049-58.

Nisonoff, A., Wissler, F.C., Lipman, L.N. and Woernley, D.L., 1960, Separation of univalent fragments from the bivalent rabbit antibody molecule by reduction of disulfide bonds. Arch Biochem Biophys 89, 230-44.

Obado, S.O., Brillantes, M., Uryu, K., Zhang, W., Ketaren, N.E., Chait, B.T., Field, M.C. and Rout, M.P., 2016, Interactome Mapping Reveals the Evolutionary History of the Nuclear Pore Complex. PLoS Biol 14, e1002365.

Oda, Y., Huang, K., Cross, F.R., Cowburn, D. and Chait, B.T., 1999, Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci U S A 96, 6591-6.

Oeffinger, M., 2012, Two steps forward--one step back: advances in affinity purification mass spectrometry of macromolecular complexes. Proteomics 12, 1591-608.

Oeffinger, M., Wei, K.E., Rogers, R., DeGrasse, J.A., Chait, B.T., Aitchison, J.D. and Rout, M.P., 2007, Comprehensive analysis of diverse ribonucleoprotein complexes. Nat Methods 4, 951-6.

Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. and Nishimune, Y., 1997, 'Green mice' as a source of ubiquitous green cells. FEBS Lett 407, 313-9.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A. and Ahn, N.G., 2005, Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4, 1487-502.

Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F. and Chartrand, P., 2005, Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. Mol Cell Biol 25, 4752-66.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M., 2002, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1, 376-86.

Oricchio, E., Nanjangud, G., Wolfe, A.L., Schatz, J.H., Mavrakis, K.J., Jiang, M., Liu, X., Bruno, J., Heguy, A., Olshen, A.B., Socci, N.D., Teruya-Feldstein, J., Weis-Garcia, F., Tam, W., Shaknovich, R., Melnick, A., Himanen, J.P., Chaganti, R.S. and Wendel, H.G., 2011, The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. Cell 147, 554-64.

Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remington, S.J., 1996, Crystal structure of the Aequorea victoria green fluorescent protein. Science 273, 1392-5.

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S.G., Wohlkonig, A., Ruf, A., Muyldermans, S., Hol, W.G., Kobilka, B.K. and Steyaert, J., 2014, A general protocol for the generation of Nanobodies for structural biology. Nature protocols 9, 674-93.

Parham, P., 1983, On the fragmentation of monoclonal IgG1, IgG2a, and IgG2b from BALB/c mice. J Immunol 131, 2895-902.

Pasero, C. and Olive, D., 2013, Interfering with coinhibitory molecules: BTLA/HVEM as new targets to enhance anti-tumor immunity. Immunol Lett 151, 71-5.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M. and Guthrie, C., 2007, Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. PLoS Biol 5, e90.

Porter, R.R., 1959, The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. Biochem J 73, 119-26.

Poser, I., Sarov, M., Hutchins, J.R., Heriche, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M.M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K.P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A.F., Peters, J.M., Buchholz, F. and Hyman, A.A., 2008, BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods 5, 409-15.

Pratt, J.M., Simpson, D.M., Doherty, M.K., Rivers, J., Gaskell, S.J. and Beynon, R.J., 2006, Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protoc 1, 1029-43.

Prochasson, P., Florens, L., Swanson, S.K., Washburn, M.P. and Workman, J.L., 2005, The HIR corepressor complex binds to nucleosomes generating a distinct protein/DNA complex resistant to remodeling by SWI/SNF. Genes Dev 19, 2534-9.

Querido, E. and Chartrand, P., 2008, Using fluorescent proteins to study mRNA trafficking in living cells. Methods Cell Biol 85, 273-92.

Rappsilber, J., Ishihama, Y. and Mann, M., 2003, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem 75, 663-70.

Renier, N., Wu, Z., Simon, D.J., Yang, J., Ariel, P. and Tessier-Lavigne, M., 2014, iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. Cell 159, 896-910.

Revets, H., De Baetselier, P. and Muyldermans, S., 2005, Nanobodies as novel agents for cancer therapy. Expert Opin Biol Ther 5, 111-24.

Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. and Ewers, H., 2012, A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. Nature Methods 9, 582-4.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B., 1999, A generic protein purification method for protein complex characterization and proteome exploration. Nature biotechnology 17, 1030-2.

Rodriguez-Navarro, S., Fischer, T., Luo, M.J., Antunez, O., Brettschneider, S., Lechner, J., Perez-Ortin, J.E., Reed, R. and Hurt, E., 2004, Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. Cell 116, 75-86.

Rodriguez, A.J., Condeelis, J., Singer, R.H. and Dictenberg, J.B., 2007, Imaging mRNA movement from transcription sites to translation sites. Semin Cell Dev Biol 18, 202-8.

Romer, T., Leonhardt, H. and Rothbauer, U., 2011, Engineering antibodies and proteins for molecular in vivo imaging. Curr Opin Biotechnol 22, 882-7.

Roovers, R.C., Vosjan, M.J., Laeremans, T., el Khoulati, R., de Bruin, R.C., Ferguson, K.M., Verkleij, A.J., van Dongen, G.A. and van Bergen en Henegouwen, P.M., 2011, A biparatopic anti-EGFR nanobody efficiently inhibits solid tumour growth. Int J Cancer 129, 2013-24.

Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyldermans, S., Cardoso, M.C. and Leonhardt, H., 2006, Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Methods 3, 887-9.

Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y. and Chait, B.T., 2000, The yeast nuclear pore complex: composition, architecture, and transport mechanism. The Journal of cell biology 148, 635-51.

Roy, A., Kucukural, A. and Zhang, Y., 2010, I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5, 725-38.

Roy, A., Yang, J. and Zhang, Y., 2012, COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res 40, W471-7.

Saavedra, C.A., Hammell, C.M., Heath, C.V. and Cole, C.N., 1997, Yeast heat shock mRNAs are exported through a distinct pathway defined by Rip1p. Genes & development 11, 2845-56.

Sachs, A.B., Davis, R.W. and Kornberg, R.D., 1987, A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol Cell Biol 7, 3268-76.

Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H. and Vogel, J., 2009, In vivo expression and purification of aptamer-tagged small RNA regulators. Nucleic Acids Res 37, e133.

Sasso, E.H., Silverman, G.J. and Mannik, M., 1991, Human IgA and IgG F(ab')2 that bind to staphylococcal protein A belong to the VHIII subgroup. J Immunol 147, 1877-83.

Schechter, I. and Berger, A., 1967, On the size of the active site in proteases. I. Papain. Biochemical and biophysical research communications 27, 157-62.

Scheid, J.F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T.Y., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., Hurley, A., Myung, S., Boulad, F., Poignard, P., Burton, D.R., Pereyra, F., Ho, D.D., Walker, B.D., Seaman, M.S., Bjorkman, P.J., Chait, B.T. and Nussenzweig, M.C., 2011, Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. Science 333, 1633-7.

Schmid, M., Olszewski, P., Pelechano, V., Gupta, I., Steinmetz, L.M. and Jensen, T.H., 2015, The Nuclear PolyA-Binding Protein Nab2p Is Essential for mRNA Production. Cell Rep 12, 128-39.

Schmid, M., Poulsen, M.B., Olszewski, P., Pelechano, V., Saguez, C., Gupta, I., Steinmetz, L.M., Moore, C. and Jensen, T.H., 2012, Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. Mol Cell 47, 267-80.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M., 2011, Global quantification of mammalian gene expression control. Nature 473, 337-42.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Luhrmann, R. and Hurt, E., 1997, Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. EMBO J 16, 3256-71.

SenGupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M., 1996, A three-hybrid system to detect RNA-protein interactions in vivo. Proc Natl Acad Sci U S A 93, 8496-501.

Seraphin, B. and Rosbash, M., 1989, Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell 59, 349-58.

Shagin, D.A., Barsova, E.V., Yanushevich, Y.G., Fradkov, A.F., Lukyanov, K.A., Labas, Y.A., Semenova, T.N., Ugalde, J.A., Meyers, A., Nunez, J.M., Widder, E.A., Lukyanov, S.A. and Matz, M.V., 2004, GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. Mol Biol Evol 21, 841-50.

Shaner, N.C., Steinbach, P.A. and Tsien, R.Y., 2005, A guide to choosing fluorescent proteins. Nat Methods 2, 905-9.

Sharp, P.M. and Li, W.H., 1987, The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15, 1281-95.

Sheth, U. and Parker, R., 2006, Targeting of aberrant mRNAs to cytoplasmic processing bodies. Cell 125, 1095-109.

Shi, Y., Pellarin, R., Fridy, P.C., Fernandez-Martinez, J., Thompson, M.K., Li, Y., Wang, Q.J., Sali, A., Rout, M.P. and Chait, B.T., 2015, A strategy for dissecting the architectures of native macromolecular assemblies. Nat Methods 12, 1135-8.

Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J., 2006, Novel chromophores and buried charges control color in mFruits. Biochemistry 45, 9639-47.

Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P. and Geromanos, S.J., 2006, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 5, 144-56.

Silverman, J., Liu, Q., Bakker, A., To, W., Duguay, A., Alba, B.M., Smith, R., Rivas, A., Li, P., Le, H., Whitehorn, E., Moore, K.W., Swimmer, C., Perlroth, V., Vogt, M., Kolkman, J. and Stemmer, W.P., 2005, Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. Nature biotechnology 23, 1556-61.

Simon, D.N. and Rout, M.P., 2014, Cancer and the nuclear pore complex. Adv Exp Med Biol 773, 285-307.

Skerra, A. and Pluckthun, A., 1988, Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. Science 240, 1038-41.

Slobodin, B. and Gerst, J.E., 2010, A novel mRNA affinity purification technique for the identification of interacting proteins and transcripts in ribonucleoprotein complexes. RNA 16, 2277-90.

Smith, T.C., Fridy, P.C., Li, Y., Basil, S., Arjun, S., Friesen, R.M., Leszyk, J., Chait, B.T., Rout, M.P. and Luna, E.J., 2013, Supervillin binding to myosin II and synergism with anillin are required for cytokinesis. Mol Biol Cell 24, 3603-19.

Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A. and Vermeulen, M., 2013, Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res 41, e28.

Snay-Hodge, C.A., Colot, H.V., Goldstein, A.L. and Cole, C.N., 1998, Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. EMBO J 17, 2663-76.

Spivak, M., Weston, J., Bottou, L., Kall, L. and Noble, W.S., 2009, Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. J Proteome Res 8, 3737-45.

Srisawat, C. and Engelke, D.R., 2001, Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. RNA 7, 632-41.

Strasser, K., Bassler, J. and Hurt, E., 2000, Binding of the Mex67p/Mtr2p heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. J Cell Biol 150, 695-706.

Strasser, K. and Hurt, E., 2000, Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. EMBO J 19, 410-20.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R. and Hurt, E., 2002, TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417, 304-8.

Tackett, A.J., DeGrasse, J.A., Sekedat, M.D., Oeffinger, M., Rout, M.P. and Chait, B.T., 2005, I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. J Proteome Res 4, 1752-6.

Tardiff, D.F., Lacadie, S.A. and Rosbash, M., 2006, A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Molecular cell 24, 917-29.

Taylor, M.S., LaCava, J., Mita, P., Molloy, K.R., Huang, C.R., Li, D., Adney, E.M., Jiang, H., Burns, K.H., Chait, B.T., Rout, M.P., Boeke, J.D. and Dai, L., 2013, Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Cell 155, 1034-48.

Thompson, M.K., Fridy, P.C., Keegan, S., Chait, B.T., Fenyo, D. and Rout, M.P., 2016, Optimizing selection of large animals for antibody production by screening immune response to standard vaccines. J Immunol Methods 430, 56-60.

Tollervey, D., 1987, A yeast small nuclear RNA is required for normal processing of pre-ribosomal RNA. EMBO J 6, 4169-75.

Trcek, T., Larson, D.R., Moldon, A., Query, C.C. and Singer, R.H., 2011, Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell 147, 1484-97.

Tsai, B.P., Wang, X., Huang, L. and Waterman, M.L., 2011, Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. Mol Cell Proteomics 10, M110 007385.

Tseng, S.S., Weaver, P.L., Liu, Y., Hitomi, M., Tartakoff, A.M. and Chang, T.H., 1998, Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. EMBO J 17, 2651-62.

Tuck, A.C. and Tollervey, D., 2013, A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 154, 996-1009.

Udayakumar, D., Zhang, G., Ji, Z., Njauw, C.N., Mroz, P. and Tsao, H., 2011, EphA2 is a critical oncogene in melanoma. Oncogene 30, 4921-9.

Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L., 1983, Gene fusion vectors based on the gene for staphylococcal protein A. Gene 23, 369-78.

Ulrichts, H., Silence, K., Schoolmeester, A., de Jaegere, P., Rossenu, S., Roodt, J., Priem, S., Lauwereys, M., Casteels, P., Van Bockstaele, F., Verschueren, K., Stanssens, P., Baumeister, J. and Holz, J.B., 2011, Antithrombotic drug candidate ALX-0081 shows superior preclinical efficacy and safety compared with currently marketed antiplatelet drugs. Blood 118, 757-65.

Upadhyay, A., Dixit, U., Manvar, D., Chaturvedi, N. and Pandey, V.N., 2013, Affinity capture and identification of host cell factors associated with hepatitis C virus (+) strand subgenomic RNA. Mol Cell Proteomics 12, 1539-52.

Van Bockstaele, F., Holz, J.B. and Revets, H., 2009, The development of nanobodies for therapeutic applications. Curr Opin Investig Drugs 10, 1212-24.

Vanlandschoot, P., Stortelers, C., Beirnaert, E., Ibanez, L.I., Schepens, B., Depla, E. and Saelens, X., 2011, Nanobodies(R): new ammunition to battle viruses. Antiviral Res 92, 389-407.

Vasudevan, S. and Steitz, J.A., 2007, AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. Cell 128, 1105-18.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Jr., Hieter, P., Vogelstein, B. and Kinzler, K.W., 1997, Characterization of the yeast transcriptome. Cell 88, 243-51.

Venema, J. and Tollervey, D., 1999, Ribosome synthesis In Saccharomyces cerevisiae. Ann. Rev. Gen. 33, 261-311.

Vinciguerra, P., Iglesias, N., Camblong, J., Zenklusen, D. and Stutz, F., 2005, Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. EMBO J 24, 813-23.

Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S. and Conrath, K., 2009, General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. J Biol Chem 284, 3273-84.

von Pawel-Rammingen, U., Johansson, B.P. and Bjorck, L., 2002, IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. The EMBO journal 21, 1607-15.

Walzthoeni, T., Leitner, A., Stengel, F. and Aebersold, R., 2013, Mass spectrometry supported determination of protein complex structure. Curr Opin Struct Biol 23, 252-60.

Wang, Z., Gerstein, M. and Snyder, M., 2009, RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57-63.

Ward, E.S., Gussow, D., Griffiths, A.D., Jones, P.T. and Winter, G., 1989, Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. Nature 341, 544-6.

Warner, J.R., 1999, The economics of ribosome biosynthesis in yeast. Trends Biochem Sci 24, 437-40.

Weirich, C.S., Erzberger, J.P., Flick, J.S., Berger, J.M., Thorner, J. and Weis, K., 2006, Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. Nat Cell Biol 8, 668-76.

Westblade, L.F., Minakhin, L., Kuznedelov, K., Tackett, A.J., Chang, E.J., Mooney, R.A., Vvedenskaya, I., Wang, Q.J., Fenyo, D., Rout, M.P., Landick, R., Chait, B.T., Severinov, K. and Darst, S.A., 2008, Rapid isolation and identification of bacteriophage T4-encoded modifications of Escherichia coli RNA polymerase: a generic method to study bacteriophage/host interactions. Journal of proteome research 7, 1244-50.

Will, C.L. and Luhrmann, R., 2011, Spliceosome structure and function. Cold Spring Harb Perspect Biol 3.

Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. and Lockhart, D.J., 1997, Genome-wide expression monitoring in Saccharomyces cerevisiae. Nat Biotechnol 15, 1359-67.

Worn, A. and Pluckthun, A., 2001, Stability engineering of antibody single-chain Fv fragments. Journal of molecular biology 305, 989-1010.

Xia, N.S., Luo, W.X., Zhang, J., Xie, X.Y., Yang, H.J., Li, S.W., Chen, M. and Ng, M.H., 2002, Bioluminescence of Aequorea macrodactyla, a common jellyfish species in the East China Sea. Mar Biotechnol (NY) 4, 155-62.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W. and Steinmetz, L.M., 2009, Bidirectional promoters generate pervasive transcription in yeast. Nature 457, 1033-7.

Yamaguchi, Y., Kim, H., Kato, K., Masuda, K., Shimada, I. and Arata, Y., 1995, Proteolytic fragmentation with high specificity of mouse immunoglobulin G. Mapping of proteolytic cleavage sites in the hinge region. J Immunol Methods 181, 259-67.

Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M.V. and Remington, S.J., 2001, Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. P Natl Acad Sci USA 98, 462-7.

Zacharias, D.A., Violin, J.D., Newton, A.C. and Tsien, R.Y., 2002, Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. Science 296, 913-6.

Zenklusen, D., Larson, D.R. and Singer, R.H., 2008, Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol 15, 1263-71.

Zenklusen, D., Vinciguerra, P., Wyss, J.C. and Stutz, F., 2002, Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Mol Cell Biol 22, 8241-53.

Zhang, Y., 2008, I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40.

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E. and Reed, R., 2000, The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature 407, 401-5.

Zid, B.M. and O'Shea, E.K., 2014, Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature 514, 117-21.

Zuiderweg, E.R., 2002, Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry 41, 1-7.