SINGLE MOLECULE PULL DOWN OF DOUBLE STRAND RNA BINDING PROTEINS

BY

XINLEI WANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Bioengineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Advisor:

Professor Su-A Myong

ABSTRACT

RNA molecules are transcribed as single stranded naturally, but with most of them forming in to structures composed of duplex regions, loop, bulge or mismatches. RNAs with double stranded regions, or known as double-strand RNAs (dsRNAs). The class of proteins responsible for processing dsRNAs is termed double-stranded RNA binding proteins (dsRBP). In recent decades, an increasing number of reports have shown the role of dsRBP-dsRNA interaction as core strategy in various cellular regulation pathways, including RNA interference, anti-viral immunity, mRNA transport and alternative splicing. However, little is known about the molecular mechanisms underlying the interaction between dsRBPs and dsRNA.

Here we examined four human dsRBPs, ADAD2, TRBP, Staufen 1 and ADAR1 `which have various numbers of RNA binding domains expressed in mammalian cells. We applied single molecule pull-down (SiMPull) assay to investigate the intensity of various dsRNA-dsRBP interactions. Our results demonstrate that despite the highly conserved dsRNA binding domains, the dsRBPs exhibit diverse substrate specificy. While TRBP and ADAR1 have a preference for binding simple duplex RNA, ADAD2 and Staufen1 display higher affinity to imperfectly base-paired structured RNA substrates. We also demonstrate ATP-independent sliding activity of TRBP and Staufen probed by single molecule protein induced fluorescence enhancement (smPIFE), which demonstrates how single molecule approaches could be utilized to provide new insight into molecular mechanisms involved in protein-RNA interaction. Collectively, our study highlights the diverse nature of substrate specificity exhibited by dsRBPs that may be critical for their cellular function.

To Father, Mother and Ruiqing Lu

ACKNOWLEDGEMENTS

I would like to thank my advisor Su-A Myong for revising numerous drafts and clearing up several critical points. I would also thank Dr. Taekjip Ha for all the help I got from his group on techniques and instruments.

This project would not be possible without many other people, who have contribute their support over years both on bench work and discussion. I would like to thank Hye Ran Koh for her invaluable suggestion and mentorship. For training me on human cell culture, I would like to thank Hyang Yeon Lee. For the help on setting up human open reading frame library, I would like to thank Leslie, Benjamin James. Additionally, for helping me setting up fluorometry measurement, I would like to thank Alex Kreig. I would like to thank Ramreddy Tippana, Younghoon Kim and Yupeng Qiu for their participation in discussion and valuable ideas.

This work was supported by the U.S. National Science Foundation Physics Frontiers Center Program (0822613) through the Center for the Physics of Living Cells and NIH Director's new Innovator Award NIH 1DP2GM105453 for S.M. and X.W. and National Science Foundation grant NSF-PHY-1430124, the National Institutes of Health grant NIH 9P41GM104601 for K.S. and L.V.

Finally I would like to thank my family along with Ruiqing Lu for their love and support through this long but rewarding process.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Background

While all cellular RNA molecules are synthesized in single-stranded (ss) form, many can form into secondary structures that encompass segments of double stranded (ds) RNA. Hence, dsRNA molecules are common in cells and are recognized as critical regulatory factors in many biological processes (1-3). For example, dsRNA regions are present in the precursors of microRNAs, siRNAs, messenger RNA (mRNA), transfer RNA (tRNA), as well as in the genome of RNA viruses that can be released into cells upon infection.

The family of proteins responsible for processing dsRNA is called double stranded RNA binding proteins (dsRBP). In recent decades, an increasing number of reports have shown the employment of dsRBP-dsRNA interaction as core strategy in various cellular regulation pathways, including RNA interference, anti-viral immunity, mRNA transport and alternative splicing. Various dsRNAs serve as cargoes, activators and substrates of dsRBPs to carry out the function of this protein family (4,5). For example, certain dsRNA structures found in viruses activates protein kinase R (PKR), which in turn triggers the downstream antiviral immune pathways (6,7); pri-microRNAs are recognized and cleaved by Drosha-DGCR8 to produce pre-microRNA in the nucleus, which then get exported into cytoplasm and cleaved by Dicer-TRBP to transform into matured microRNA (8,9).

The dsRBP family is defined by the presence of one or more double-stranded RNA binding domains (dsRBD) (10). The dsRBDs are highly conserved in amino acid composition and domain structures and have been identified across various species (11-13). Despite the high degree of domain conservation, dsRBPs are involved in diverse biological functions where they

interact with variety of RNA substrates, which vary in secondary structure and differ in length of duplex. While the biological functions of dsRBPs are known, it remains uncertain if dsRBPs exhibit certain substrate specificity.

Two types of dsRBDs are found in dsRBPs; type-1 dsRBD (dsRBD-I) usually binds dsRNA while type-2 (dsRBD-II) is mainly involved in protein-protein interaction (10,14). The number of dsRBD-I present in each dsRBP is highly variable (5); for instance, while both function as RNA deaminase, ADAD2 contains only one dsRBD-1, whereas ADAR1 contains three. It is heretofore unknown why some dsRBPs need multiple units while others possess a single dsRBD-I, and if the number of the dsRBD-I is correlated with the protein's affinity to dsRNA. dsRBD-I adopts an "α-β-β-β-α" structure, which contacts dsRNA in three grooves (minor-major-minor) along a stem spanning 15 base pairs (bp) (11,12). This protein-RNA binding mode is structure- but not sequence-dependent since dsRBDs recognize the A-form helical axis of dsRNA rather than the specific RNA sequence (4,15). While dsRBD-I of ADAR2 recognizes and binds dsRNA at certain mismatch locations (16), it is not clear to what extent other dsRBDs contribute to binding dsRNA and imperfectly base-paired structured dsRNA.

1.2 Thesis organization

To address some of these outstanding questions, we examined dsRNA interaction with four dsRBPs: ADAD2, TRBP, Staufen1 and ADAR1. We chose dsRBPs that contain different number of dsRBD-I units and participate in various cellular functions. ADAD2 has only a single dsRBD-I, followed by Staufen1 and TRBP containing two, and ADAR1 possessing three units of dsRBD-Is. In terms of biological function, both ADAD2 and ADAR1 are RNA deaminases that edit adenosine to inosine (A to I) in mRNA and microRNA precursors (17). Staufen1 is responsible for mRNA transport to dendrites in neurons where its fourth dsRBD likely binds

microtubules along axons (18). In addition, human Staufen1 binds to the 19bp stem in ARF1 mRNA and also to intermolecular Alu element-Alu element duplexes for the purpose of Staufen-mediated mRNA decay (SMD)(19)(20); and intramolecular Alu element-Alu elemenr duplexes to compete with mRNA retention in paraspeckles (21). TRBP is a key player in the RNA Induced Silencing Complex (RISC) assembly (22) and also modulates the initiation of HIV-1 gene expression (23,24).

We investigated the binding affinity of the four dsRBPs toward six different dsRNA substrates with varying length of duplex and secondary structures. The length variants include 25bp, 40bp, and 55bp dsRNA whereas the imperfectly base-paired structured RNA includes pre-let7 (pre-microRNA), TAR RNA and tRNA. All proteins were overexpressed in mammalian cells (HEK 293) and pulled down to a single molecule imaging surface coated with the appropriate antibody (25). Fluorescence-labeled RNA substrates were added to test their binding affinity. Our study reveals that despite the presence of highly conserved dsRNA binding domains, the dsRBPs investigated display substantial differences in their substrate specificity on RNA substrates.

CHAPTER 2
METHODOLOGY

2.1 RNA labeling and annealing

The sequences of all RNA substrates are displayed in Table S1. Pre-let7, TAR and tRNA molecules were purchased from IDT as single strand RNA with fluorescent label at the 5' end. 25bp, 40bp and 55bp dsRNAs were purchased from Dharmacon as separate single strand RNA and 3'-DY547 was incorporated in the process of each RNA synthesis.

For dsRNA annealing, two complementary RNA strands were mixed in equal concentration in annealing buffer (100mM NaCl and 10mM Tris at pH 8) and heated at ~90°C for 2 min and gradually cooled to room temperature. U40 with 3' amine modification was purchased from IDT and labeled with Cy3 NHS ester dye from GE Healthcare1. Briefly, the dye was mixed at two-fold molar excess concentration with RNA containing 3' amine modifier, in a buffer containing 100mM $NaHCO_3$ at pH 8.5 and then incubated overnight. Unreacted dye was removed by two rounds of ethanol precipitation. The resulting labeling efficiency was about 90%.

2.2 Protein lysate preparation

ADAD2, TRBP, and Staufen1 were cloned from Human Open Reading Frame Library and a C-terminal EYFP was added to each protein sequence. Then, C-terminal EYFP-TRBP, C-terminal EYFP-Staufen1 and N-terminal EGFP-ADAR1 were overexpressed in human A549 cells, and C-terminal EYFP-ADAD2 was overexpressed in HEK293 cells. Cells were lysed using RIPA (Thermo Scientific RIPA Lysis and Extraction Buffer, Catalog number: 89900) 24 hours after transfection and cell lysates were collected and centrifuged; finally supernatants were collected for each protein. Cell lysates were stored in -80°C for later use.

2.3 Lysate protein concentration measurement

To measure the concentration of EYFP/EGFP tagged proteins and normalize them to the same level on single molecule surface, following things have been done. First, we used commercial cy5 dye with known concentration measured by spectrophotometry and did a serial dilution with cy5 and measure the corresponding fluorescent intensity with fluorometry at each diluted concentration to produce a standard concentration-fluorescent intensity line.

And then we measured the fluorescent intensity of the four dsRBPs in lysate with either EGFP or EYFP tags by fluorometry and normalized the intensity using extinct coefficient, quantum yield and absorption percentage at specific excitation wavelength of EGFP and EYFP, respectively to get the normalized fluorescence of each protein lysate. And we plot the normalized fluoresence along the standard concentration-fluorescent intensity line to read concentration for each individual lysate sample.

2.4 Single molecule pull down assay (SiMPull)

Polyethylene glycol (PEG)-coated quartz slides with flow chambers were obtained according to previously published protocol 2. A PEG surface was coated with Neutravidin (0.05mg/mL) followed by anti-GFP (RABBIT, 5μg/ml) antibody conjugated with biotin (Rockland 600-406-215) and incubated for another 5min in T50 (10 mM Tris pH 8 and 50 mM NaCl). About 400pM of C-terminal EYFP or N-terminal EGFP fused dsRBP cell lysates were added to the antibody coated surface and incubated for 5-10min.

To measure the level of dsRBPs on the PEG surface, we used the level of TRBP-EYFP as standard for all dsRBPs, since the concentration of four dsRBPs were calibrated and dilution factors were normalized to make sure they were applied at the same level on the PEG surface. To measure the level of TRBP after SM Pull-down using biotinylated GFP antibody, we used anti-TRBP antibody (Abcam [1D9](ab129325)) and Anti-mouse Alexa Fluor® 488 Conjugate (Cell Signaling #4408).

2.5 Tables

Table 2.1.

Name	Sequence
25bp RNA	5’- rGrCrUrUrGrUrCrGrGrGrArGrCrGrCrCrArCrCrCrUrCrUrGrC-3’ (up) 5’- rGrCrArGrArGrGrGrUrGrGrCrGrCrUrCrCrCrGrArCrArArGrC-DY547-3’ (down)
40bp RNA	5’- rGrCdTrUrArArCrArArCrCrArGrArUrCrArArArGrArArArArArArCrArGrArCrArUrU rGrUrCrA-3’ (up) 5’- rUrGrArCrArArUrGrUrCrUrGrUrUrUrUrUrUrCrUrUrUrGrArUrCrUrGrGrUrUrGrU rUrArArGrCrGrU-DY547-3’ (down)
55bp RNA	5’- rArCrGrCrUrUrArArCrArArCrCrArGrArUrCrArArArGrArArArArArArCrArGrArCrA rUrUrGrUrCrArArUrUrGrCrArArArGrCrArArArArA-3’ (up) 5’- rUrUrUrUrUrGrCrUrUrUrGrCrArArUrUrGrArCrArArUrGrUrCrUrGrUrUrUrUrUrU rCrUrUrUrGrArUrCrUrGrGrUrUrGrUrUrArArGrCrGrU-DY547-3’ (down)
Pre-let7	5’- cy3-rUrGrArGrGrUrArGrUrArGrGrUrUrGrUrArUrArGrUrUrUrUrArGrGrGrUrCrArCrA rCrCrCrArCrCrArCrUrGrGrGrArGrArUrArArCrUrArUrArCrArArUrCrUrArCrUrGrUrCrU rUrArCrC-3’
TAR	5’-cy3-rGrGrUrCrUrCrUrCrUrCrUrGrGrUrUrArGrCrCrArGrArUrCrUrGrArGrCrCrUrGrGrGrA rGrCrUrCrUrCrUrGrGrCrUrArArCrUrArGrGrGrArArCrC-3’
tRNA	5’-cy3- rGrGrGrArArGrCrCrCrGrGrArUrArGrCrUrCrArGrUrCrGrGrUrArGrArGrCrArUrC rArGrArCrUrUrUrUrArArUrCrUrGrArGrGrGrUrCrCrArGrGrGrUrUrCrArArGrUrCrC rCrUrGrUrUrCrGrGrGrCrGrCrCrA-3’
PolyU40	5’ –rUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrU rUrUrUrUrUrU-cy3-3’

Table 2.1. Sequences of dsRNA substrates involved in interaction with dsRBPs.

CHAPTER 3
RESULTS

3.1 Single molecule pull down (SiMPull) of dsRBPs

Four dsRBPs, ADAD2, TRBP, Staufen1 and ADAR1 were chosen because they all have at least one dsRBD-I (type I dsRBD) with highly conserved amino acid residues (Figure 3.1 A and B). However, the dsRBPs differ in total length (600 to 1,500 amino acids) as well as in overall domain composition. ADAD2 and ADAR1 possess a large deaminase domain whereas TRBP and Staufen1 encompass dsRBD-II (type II dsRBD) domains (Figure 3.1 A). We sought to set up a single molecule affinity testing platform in which one can immobilize the same number/density of dsRBP and apply same amount of RNA substrates with varying structures. To achieve this, we first need to normalize the concentration of expressed dsRBPs, followed by single molecule pull-down assay (25) to isolate dsRBPs directly from mammalian cells.

To normalize the concentration of these four proteins, we did following things First the full length of each dsRBP fused with an EYFP or EGFP tag was over-expressed in HEK293 cells (Figure 3.1 A). The cell lysate was obtained and the intensity of EYFP/EGFP was imaged by confocal microscope to test the dsRBP expression level. Second, we used commercial cy5 dye with known concentration measured by spectrophotometry (Figure 3.2 A) and serially diluted this cy5 dye to measure the corresponding fluorescent intensity by fluorometry to produce a standard concentration-fluorescent intensity line (Figure 3.2 B). And then we measured the fluorescent intensity of the four dsRBPs with either EGFP or EYFP tags by fluorometry and normalized these fluorescent intensity using extinct coefficient, quantum yield and absorption percentage at specific excitation wavelength of EGFP and EYFP, respectively.

And we plot the normalized fluorescence of these four dsRBPs along the standard concentration-fluorescent intensity line (Figure 3.2 B) to read each individual concentration.

Upon determining the concentration of each protein based on the fluorescence intensity of EYFP/EGFP, the same concentration of each dsRBP was applied on a single molecule surface coated with anti-GFP/anti-YFP antibody (24) (Figure 3.3 A). The specificity of dsRBP binding to the surface was confirmed by adding serially diluted cell lysate to anti-GFP coated surface and applying primary and fluorescence (A488) labeled secondary antibody for the corresponding dsRBP, where we used TRBP lysate as an example (Figure 3.3. B). We confirmed the dsRBP binding specificity and obtained an accurate count of TRBP molecules on the surface and the number of countable TRBP molecules on the surface was saturated as the concentration of cell lysate increased (Figure 3.3 C). And this also provided us the level of total individual dsRBP on surface, which is critical to obtain the ratio of dsRBP bound by dsRNA in the experiments afterwards.

3.2 Relative binding affinity of dsRBPs to various RNAs

To investigate the RNA binding specificity of the four dsRBPs, Cy3-labeled RNA molecules were applied to the dsRBP-immobilized imaging surface (Figure 3.4 A). The fluorescence of EYFP or EGFP on dsRBPs does not interfere with the detection of Cy3 signal due to the extremely fast photobleaching of both fluorescent proteins. The RNA substrates that differ in duplex length and secondary structure were prepared. The length variants, 25bp, 40bp and 55bp dsRNA were categorized as “non-structured” and the structure variants, pre-let7, TAZR and tRNA as “structured” RNA (Figure 3.4 B). We note that all structured RNA still retained a long dsRNA stem sufficient for dsRBP binding (Figure 3.4. C). The single strand (ss) RNA composed of 40-uracil (U40) was included as a negative control.

After checking the density of dsRBP molecules on the imaging surface, we applied Cy3 labeled RNA, washed out the unbound RNA and visualized the fluorescence signal by home-built total internal reflection fluorescence (TIRF) microscope (Figure 3.4 A). The single molecule images from multiple areas were taken to count the number of dsRNA bound to proteins on the surface. For each RNA interaction case, the average number of RNA bound per field of view was used as a proxy for the comparative binding affinity of the corresponding RNA-dsRBP pair. The nonspecific binding of dsRNA was checked by omitting cell lysates and by applying a negative control, U40 ssRNA. Both showed negligible binding, suggesting that the fluorescence signals arise from the specific binding of dsRNA to dsRBP molecules (Figure 3.6A).

Taking TRBP as an example, we further confirmed that the overexpressed dsRBPs are interacting with applied dsRNAs on themselves without any other collaborator or protein complex, and that the terminal EGFP/EYFP tagging would not affect dsRBPs binding to dsRNA substrates (Figure 3.5 A and B). We fixed the concentration of applied dsRBPs from lysate at around 300pM. At this concentration, we verified that Staufen1 would exist dominantly as monomer (Figure 3.5 C) (27). Due to limitation of bulk measurement, previous literature mostly focused on function of Staufen1 homodimer, hence single molecule pull down provides us a unique approach to study the activity of monomer Staufen1.

We quantified the fraction of RNA-bound dsRBP to total dsRBP molecules (sum of bound- and unbound molecules) for each dsRBP with six dsRNAs and one ssRNA. Amongst the four dsRBPs, TRBP showed the highest relative binding affinity to all substrates, mostly ranging between 80 to 90%, except for the reduced binding to tRNA (~50%). Staufen1, ADAR1, and ADAD2 exhibit substantially lower affinities to all RNAs in general; on average, 20%-40% of protein was occupied by RNA (Figure 3.4 C). Notably, the comparative binding affinities of dsRBPs toward RNA are not correlated with the number of dsRBD-I. For instance, ADAR1

with three dsRBD-Is and ADAD2 with only one dsRBD-I displayed similar binding affinities for RNA substrates whereas TRBP with two dsRBD-Is exhibited the highest affinity. Together, our data indicates that the number of type-1 dsRBDs is not a major factor for determining dsRBP-dsRNA affinity.

To test if TRBP alone is primarily responsible for binding dsRNA, we performed EMSA where we subjected three Cy3 labeled dsRNA substrates to both TRBP overexpressed cell lysate and purified TRBP. The result shows that TRBP-RNA complex from both are comparable (Figure 3.5 A and B). This also confirms that the EYFP tagging does not interfere with dsRNA binding (43). Together, our data indicates that the number of type-1 dsRBDs may not be a major factor for determining dsRBP-dsRNA affinity. In agreement with previous report (44, 45), Staufen1 interacts with itself at high concentration range (>100 nM) as shown by EMSA assay (Figure 3.5 C).

The four dsRBPs also displayed different binding affinities to various structural features of RNAs. To compare the binding propensity toward structured RNA, we obtained the average of all bound fractions corresponding to structured (pre-let7, TAR and tRNA) and non-structured RNA (25, 40 and 55 bp) substrates for each dsRBP (Figure 3.4 C) and calculated the ratio between structured versus non-structured (S/N), termed here affinity ratio (Figure 3.4 D); ratio of 1 indicates no bias to either type whereas a ratio >1 reflects a preference toward structured RNAs. Interestingly, ADAD2 and Staufen1 showed significantly higher affinity for the structured RNAs with the affinity ratio approaching 2:3 (Figure 3.4 D). On the other hand, TRBP and ADAR1 have similar affinities for almost all the RNA substrates (Figure 3.4 D). Unexpectedly, all the proteins displayed substantial affinity to highly structured tRNA-like RNA, which can be in part due to tRNA, which has a complex L-shape with only a short dsRNA portion of 13–14 bp and the mixed population of misfolded RNA. (46)

To further test the binding of tRNA, we performed two types of competitive binding assays. First, we applied equimolar concentration of Cy3 labeled tRNA and Cy5 labeled 27bp dsRNA to single molecule surface coated with individual dsRBPs. We observed that the binding affinity and preference for tRNA versus dsRNA exhibited the same pattern as our previous assay shown in Figure 3C b (Figure 3.6 B). This result confirms that tRNA binds to dsRBPs, albeit to varying degrees and that Staufen1tossay shown in Figur for structured RNA is retained even in the presence of dsRNA. Second, we added tRNA to dsRBPs pre-bound with dsRNA and observed that tRNA still exhibited sufficient level of competitive binding (Figure 3.6 C). Furthermore, dsRBD–tRNA binding model is plausible: in one possible binding mode, dsRBD2 of TRBP fits in the duplex-like region of tRNA, and the contact area between dsRBD and tRNA is comparable to values observed for regular dsRNA, even slightly higher, as the ssRNA tail of tRNA can also bind to the dsRBD on the side (Figure 3.6 D).

In conclusion, the four tested dsRBPs exhibit different substrate specificities in our single molecule platform, which may arise from different binding affinities toward RNA substrates with varying secondary structures. Furthermore, RNA binding affinity of dsRBP does not seem to depend on the number of its dsRBD-I.

3.3 Figures

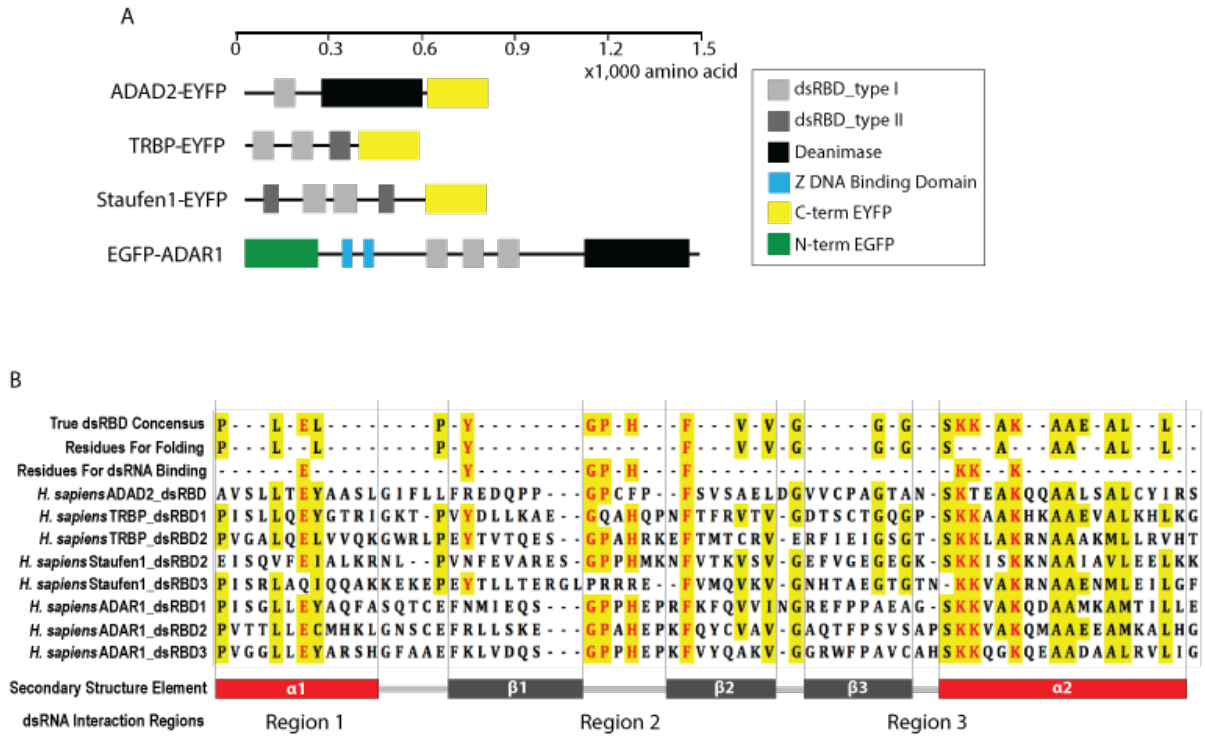

Figure 3.1. Single molecule pull down of four dsRBPs.

A) Functional domains of four dsRBPs. ADAD2, TRBP2 and Stau1 were fused with EYFP at C-terminal, while ADAR1 was fused with GFP at N-terminal.

B) Amino acid sequence information of the type-1 dsRBDs in ADAD2, TRBP2, Stau1 and ADAR1. Top line shows the consensus residues of dsRBDs, which are highlighted in yellow shade. Second line shows the key residues critical for folding into dsRBD conserved αββββα structure. The third line shows the residues necessary for dsRNA binding, of which the font color is red. Secondary structure α-β-β-β-α sections are labeled out, together with dsRNA binding regions.

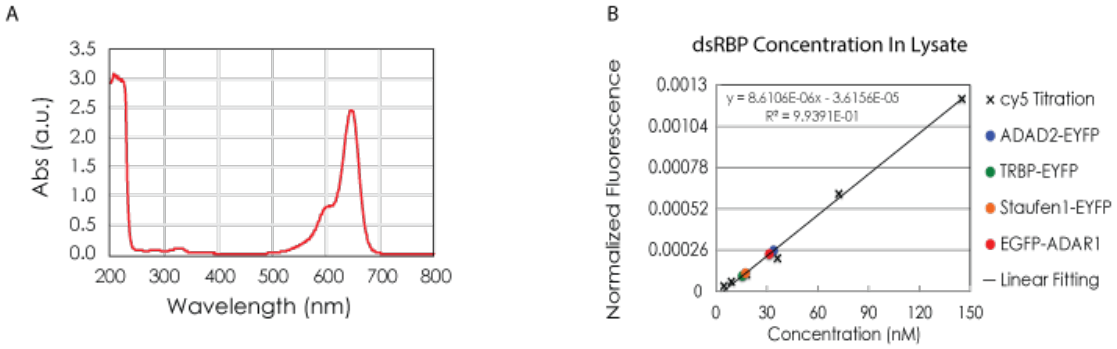

Figure 3.2. Characterization of four dsRBPs concentration from cell lysate.

A) Spectrophotometry absorption measurement of free Cy5 dye sample.

B) Determination of dsRBP concentration based on the calibration curve generated from Cy5 dye. Cy5 dye sample was serially diluted to generate a standard fluorescence-to-concentration line. Plotting normalized fluorescence intensity of four dsRBPs from lysate onto the calibration line enables estimation of concentration for each dsRBPs in lysate.

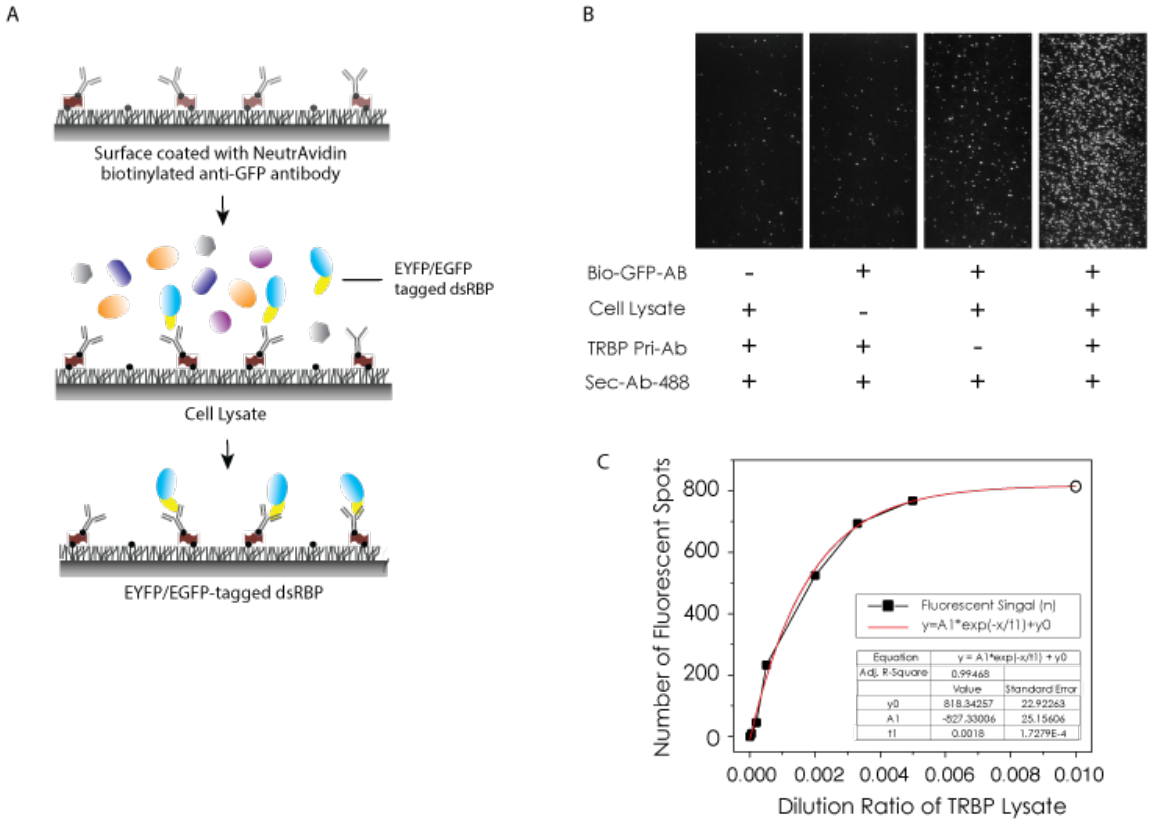

Figure 3.3. Single molecule pull down of normalized dsRBP lysate.

A) Schematic explanation for single molecule pull down of dsRBP from cell lysate using biotinylated antibody against EYFP/GFP onto bio-PEG coated surface, with non-specific protein being washed away. **B)** Measurement of TRBP level on single molecule surface using primary and Alexa488 labeled secondary antibody against TRBP. Control experiment showed the non-specific binding of TRBP, primary or secondary antibody.

C) Quantification of number of fluorescent spot from the right most panel in **B)** for increasing lysate amount, with the count saturated at highest lysate amount we tested.

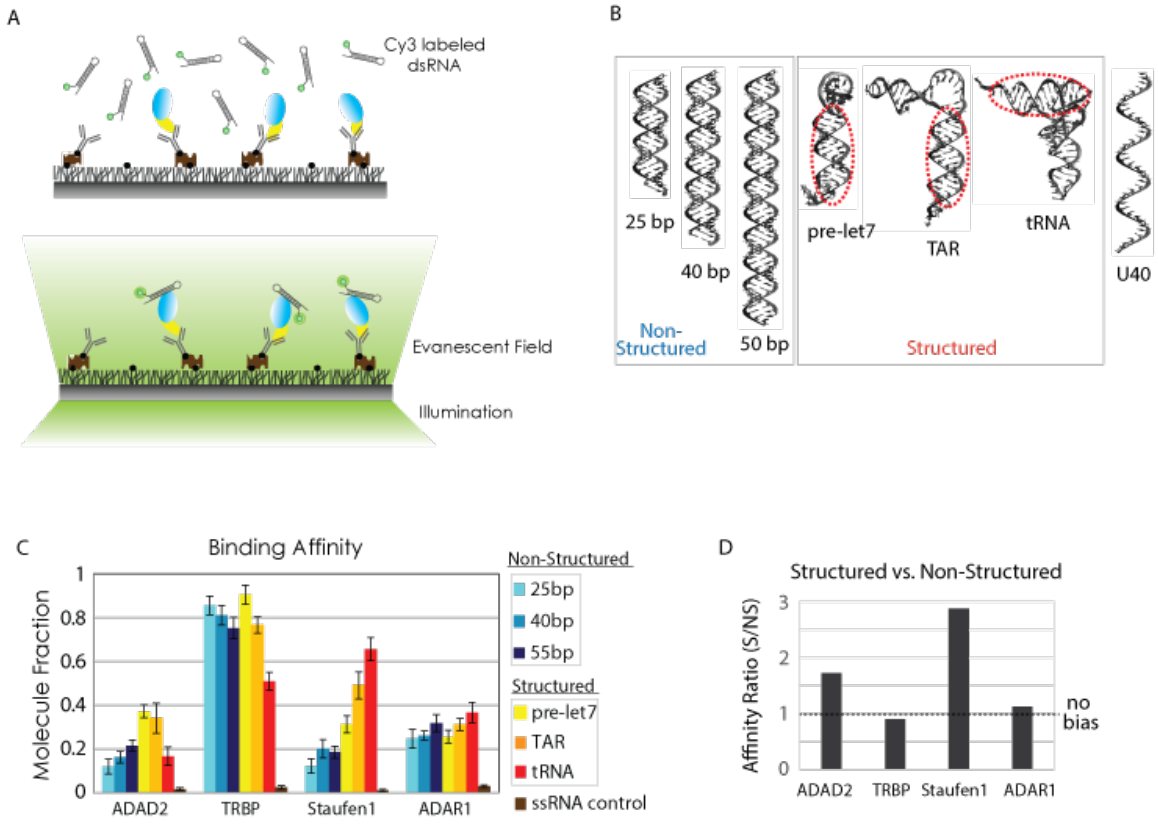

Figure 3.4. Relative binding affinity of dsRBPs to various dsRNAs.

(A) Schematic illustration of fluorescence labeled dsRNA bound by dsRBP on the PEG- coated surface. 1 nM dsRNAs were used in all cases.

(B) Secondary structure of six dsRNAs tested and the ssRNA of U40 tested as a negative control. Regions marked by red dotted circles involve a set of minor–major–minor grooves, which suffice for one dsRBD interaction with dsRNA.

(C) The relative binding affinity of each dsRBP to six dsRNAs shown in Figure 3B. The fraction was calculated as the ratio of the number of dsRNA molecule measured by cy3 detection and the number of dsRBP molecules detected by immuno-fluorescence measurement against EYFP/GFP at single molecule level.

(D) dsRBP preference to structured or non-structured dsRNA in terms of relative binding affinity. The categories of structured and non-structured RNA are shown in (B).

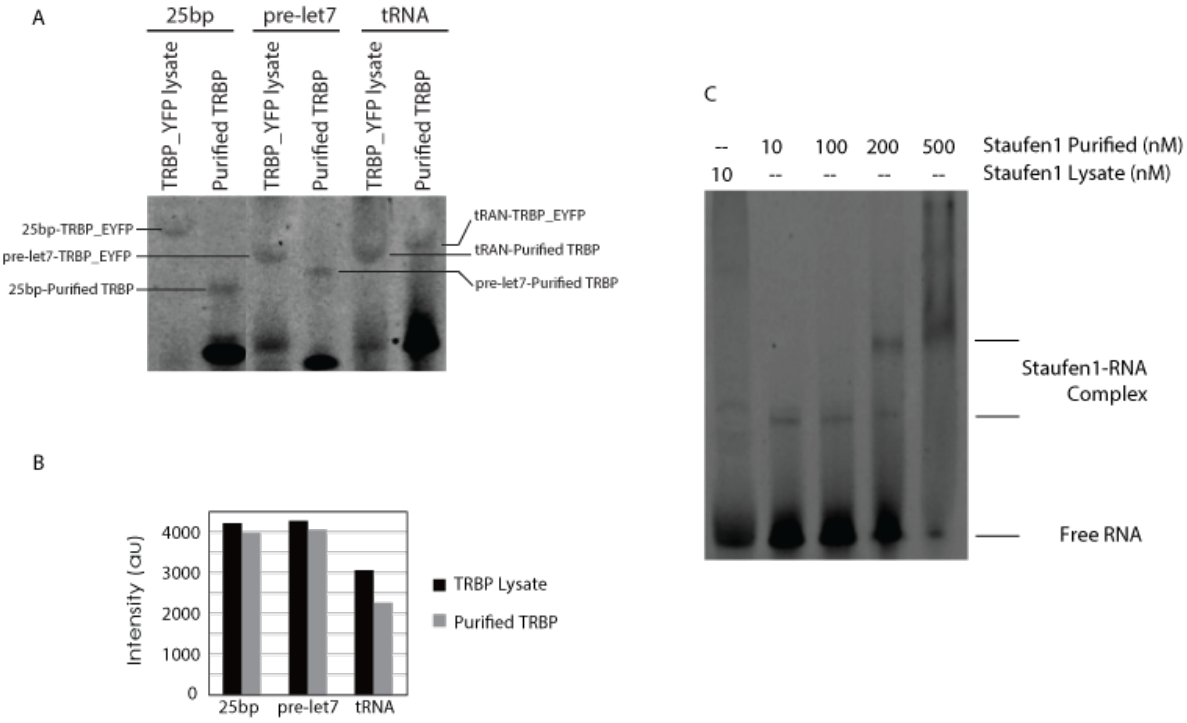

Figure 3.5. Characterization of protein purity and dimer/monomer status from cell lysate.

A) EMSA assay to test the affinity of TRBP_YFP from lysate vs. purified TRBP to three representative RNA substrates labeled with Cy3 labeled.

B) Quantification of shifted protein-RNA complex bands in A).

C) EMSA assay to test concentration dependent oligomerization of purified Staufen1 interacting with Cy3-tRNA in parallel with Staufen1 lysate.

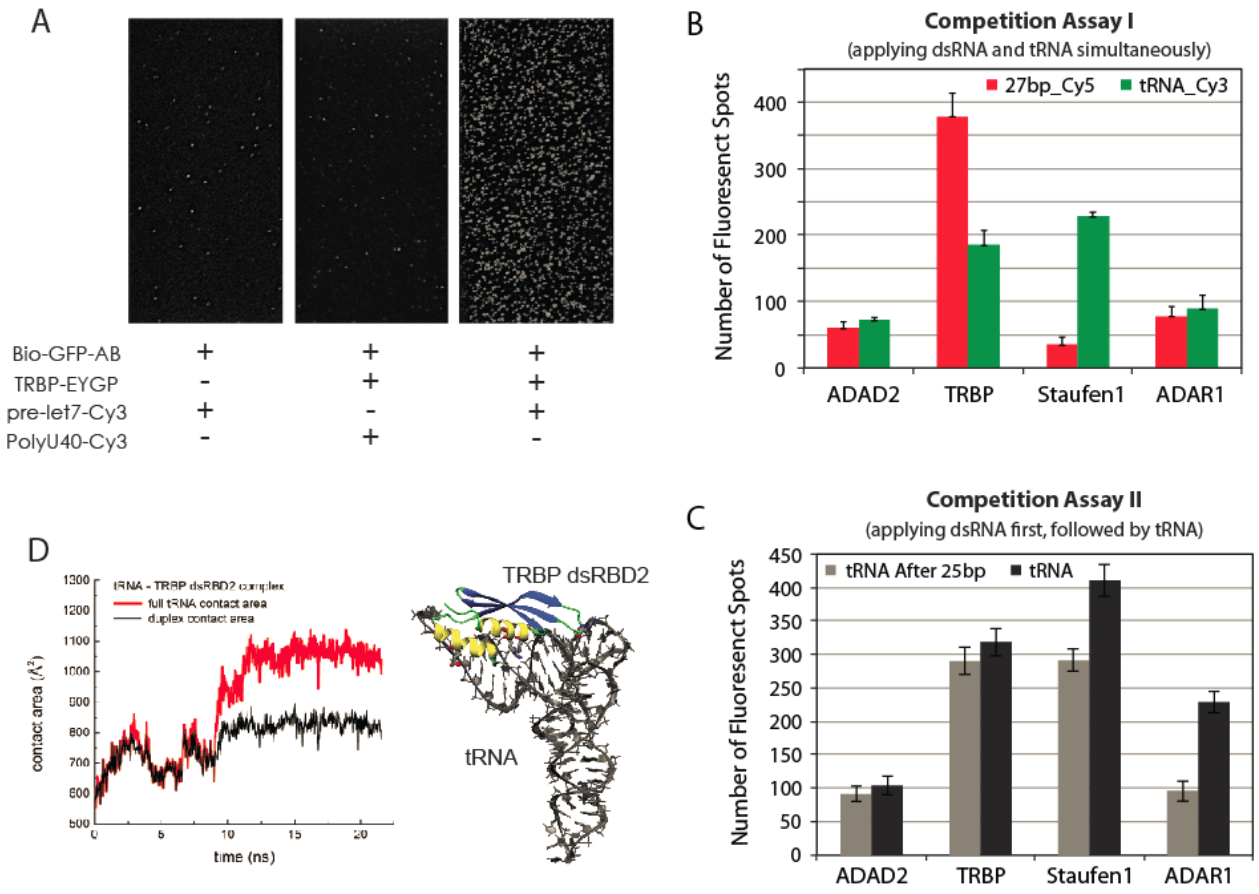

Figure 3.6. Nonspecific binding test and competitive binding affinity of dsRBPs towards tRNS vs. dsRNA

A) Testing nonspecific binding by omitting cell lysate, pre-let7 RNA and adding U40 ssRNA.

B) Competitive binding assay between duplex Cy3-tRNA and Cy5-dsRNA (27 bp) on single molecule platform by applying the same concentration (500pM each) of the two RNAs at each of the four dsRBPs immobilized surface.

C) Competitive binding assay in which dsRNA pre-bound to dsRBP was competed away by labeled tRNA (gray bar). The tRNA binding result is put in as a comparison (black bar). The tRNA was able to compete against dsRNA to sufficient degree for all four proteins tested.

D) MD simulation of dsRBD2 of TRBP to binding to dsRNA vs. tRNA. The interaction to tRNA is comparable to dsRNA.

CHAPTER 4
DISCUSSION

4.1 Protein–RNA interaction examined by single molecule fluorescence

The functions of dsRBPs are implicated in diverse cellular pathways including micro RNA, RNA editing, antiviral signaling, mRNA transport and alternative splicing. We sought to profile a set of dsRBPs in binding specificity of dsRBPs to different lengths and structures of dsRNAs that are relevant to cellular RNAs. The single molecule pull-down assay enabled us to measure the relative affinity of dsRNA substrates toward dsRBP. This platform provides several main advantages over other methods such as electrophoretic mobility shift assay (EMSA). The proteins can be directly pulled down from cell lysate without being processed through purification steps, hence preserving the native context of the protein in cells. Second, it enables one to detect not only stable binding but also weak or transient binding events, which is not possible with gel-based assays. Third, it allows observation of single dsRBP-dsRNA interaction with unlabeled protein to avoid potential disruption by the fluorescent dye that may perturb the protein activity.

4.2 dsRBPs' binding affinity to dsRNAs varies despite highly conserved dsRBD

There are ~30 known dsRBD-containing proteins in human cells. Most of the dsRBDs, including dsRBD1 and dsRBD2 of TRBP, dsRBD2 of Staufen1 and dsRBD3 of ADAR1, possess the canonical dsRNA binding motifs that contact the minor – major – minor pattern on A-form dsRNA helix. One exception is ADAD2 dsRBD, which is missing one K residue from the KKxxK motif and the H residue on the loop connecting _1 and _2 strands. The loop contains only hydrophobic residues, making it unlikely to form hydrogen bond contact with

the minor groove of dsRNA. The lack of K and H amino acids make ADAD2 dsRBD interactions with dsRNA substantially weaker than for other dsRBDs. In agreement, we obtained the lowest overall binding affinity of ADAD2 to all types of RNAs tested (Figure 3.4 C). On the other hand, dsRBD3 of ADAR1 has all the consensus dsRBD residues, yet ADAR1 exhibits weak binding to most RNAs tested, which is possibly due to the fact that ADAR1 represents a relatively large protein. Taken together, the structural model analysis and relative affinity results suggest that the binding affinity of dsRBP to dsRNA may be correlated to the predicted binding affinity of single dsRBDs in some cases, but not in all cases.

Our binding affinity results indicate a plausible distinction between the strong (TRBP) and weak binders (all others) to dsRNA, although most of the dsRBDs contain all the dsRNA binding consensus residues (besides ADAD2 and Staufen1-dsRBD3). A possible reason for this discrepancy in RNA binding strength might be the presence of a basic residue adjacent to the KKxxK motif (either K or R) observed in several dsRBDs. Alternatively, other structural features in ADAD2, ADAR1 and Staufen1 proteins can modulate their dsRNA binding and dynamic properties; for example, dsRBDs could be sterically occluded from interacting with RNA, have competitive binding to other protein do- mains, or require dimerization for high-affinity binding to dsRNA. In fact, ADAR1 is known to bind to 19-bp siRNA with high affinity in the dimer form (Kd = 0.21 nM), or with lower affinity when dimerization is prevented due to a single mutation (Kd = 2.2 nM) (51).

4.3 Number of dsRBDs may not correlate with the RNA binding

It is interesting and puzzling that the number of type-1 dsRBD per protein is highly variable in the dsRBP family, ranging from one to five. It is not known if the number of dsRBDs determines its binding strength to dsRNA or if they contribute to substrate specificity of the dsRBP. We characterized four dsRBPs with a different number of dsRBDs. Our data

shows that the binding affinity of dsRBPs toward RNA substrates is not necessarily correlated with the number of dsRBD-Is (Figure 3.1 A). ADAR1 with the highest number (3) of dsRBD-Is did not show the highest affinity whereas TRBP with only two dsRBD-Is exhibited the strongest affinity to all substrates. Our data also indicates that the number of dsRBDs may not contribute to binding preference to structured versus unstructured dsRNA or to length of dsRNA. For instance, ADAR1 with three dsRBDs showed the least variance in binding affinity toward studied RNA substrates, while Staufen1 with two dsRBDs showed preferred binding to structured RNA (Figure 3.4 D). We conclude that multiple dsRBDs may not be responsible for discriminating dsRNA length or structure on their own. We note that our conclusion is based on the sub- set of structural variants chosen for this study and may not reflect a situation in cells.

Based on the experimentally determined *Kd* values for dsRNA binding to several TRBP and ADAR1 constructs (49,51), we can further examine the effect of multiple dsRBDs on protein–dsRNA binding strength (Figure 3.6 D). The *Kd* values of TRBP–dsRBD2 and TRBP–dsRBD1 are 113 and 220 nM (49), respectively, indicating that TRBP–dsRBD2 has stronger binding to dsRNA. These *Kd* values can be used to estimate a lower bound on the *Kd* value of the TRBP–RBD1+2 construct, in which TRBP–RBD1 and TRBP–RBD2 are linked by a long flexible linker of 61 amino acids. If we assume that the dsRBDs act independently when flexibly bound, we can predict the *Kd* by obtaining a product of *Kd* values of individual dsRBDs (Figure 3.6 D). However, the experimental *Kd* value of TRBP–RBD1+2 (250 pM) is an order of magnitude higher than the theoretically estimated lower-bound *Kd* (24.9 pM), which indicates that dsRBDs are not completely independent of each other in TRBP. Therefore, the presence of multiple dsRBDs in a protein is likely to strengthen protein–dsRNA binding and lower the *Kd* value, but not to the strongest possible binding associated with the predicted lower-bound Kd value.

CHAPTER 5
CONCLUSION

By applying single molecule pull down (SiMPull) of four human cell expressed double strand RNA binding proteins (dsRBPs) and measurement of the interaction between dsRBPs and various dsRNAs, our results showed that the four tested dsRBPs exhibit different substrate specificities on our single molecule platform. Specifically, TRBP and ADAR1 have a preference for binding simple duplex RNA, while ADAD2 and Staufen1 display higher affinity to imperfectly base-paired structured RNA substrates, containing loop, bulge and mismatches. Furthermore, RNA binding affinity of dsRBP does not seem to depend on the number of its dsRBD-I. This showed an example of how single molecule approaches could be utilized to provide new insight into molecular mechanisms involved in protein-RNA interaction. Collectively, our study highlights the diverse nature of substrate specificity exhibited by dsRBPs that may be critical for their cellular function.

REFERENCES

1. Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B. *et al.* (2013) A long noncoding RNA mediates both activation and repression of immune response genes. *Science*, **341**, 789-792.
2. Han, B.W. and Chen, Y.Q. (2013) Potential pathological and functional links between long noncoding RNAs and hematopoiesis. *Science Signaling*, **6**.
3. Shukla, G.C., Singh, J. and Barik, S. (2011) MicroRNAs: Processing, maturation, target recognition and regulatory functions. *Molecular and Cellular Pharmacology*, **3**, 83-92.
4. Ryter, J.M. and Schultz, S.C. (1998) Molecular basis of double-stranded RNA-protein interactions: Structure of a dsRNA-binding domain complexed with dsRNA. *EMBO Journal*, **17**, 7505-7513.
5. Saunders, L.R. and Barber, G.N. (2003) The dsRNA binding protein family: Critical roles, diverse cellular functions. *FASEB Journal*, **17**, 961-983.
6. Nanduri, S., Carpick, B.W., Yang, Y., Williams, B.R.G. and Qin, J. (1998) Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO Journal*, **17**, 5458-5465.
7. Zheng, X. and Bevilacqua, P.C. (2004) Activation of the protein kinase PKR by short double-stranded RNAs with single-stranded tails. *Rna*, **10**, 1934-1945.
8. Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H. and Kim, V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes and Development*, **18**, 3016-3027.
9. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S. *et al.* (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*, **425**, 415-419.
10. Johnston, D.S.T., Brown, N.H., Gall, J.G. and Jantsch, M. (1992) A conserved double-stranded RNA-binding domain. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 10979-10983.
11. Tian, B., Bevilacqua, P.C., Diegelman-Parente, A. and Mathews, M.B. (2004) The double-stranded-RNA-binding motif: Interference and much more. *Nature Reviews Molecular Cell Biology*, **5**, 1013-1023.
12. Gan, J., Shaw, G., Tropea, J.E., Waugh, D.S., Court, D.L. and Ji, X. (2008) A stepwise model for double-stranded RNA processing by ribonuclease III. *Molecular Microbiology*, **67**, 143-154.
13. Masliah, G., Barraud, P. and Allain, F.H.T. (2013) RNA recognition by double-stranded

RNA binding domains: A matter of shape and sequence. *Cellular and Molecular Life Sciences*, **70**, 1875-1895.

14. Krovat, B.C. and Jantsch, M.F. (1996) Comparative mutational analysis of the double-stranded RNA binding domains of Xenopus laevis RNA-binding protein A. *Journal of Biological Chemistry*, **271**, 28112-28119.
15. Manche, L., Green, S.R., Schmedt, C. and Mathews, M.B. (1992) Interactions between double-stranded RNA regulators and the protein kinase DAI. *Molecular and Cellular Biology*, **12**, 5238-5248.
16. Stefl, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B. *et al.* (2010) The Solution Structure of the ADAR2 dsRBM-RNA Complex Reveals a Sequence-Specific Readout of the Minor Groove. *Cell*, **143**, 225-237.
17. Bass, B.L. and Weintraub, H. (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, **55**, 1089-1098.
18. Ferrandon, D., Elphick, L., Nüsslein-Volhard, C. and St Johnston, D. (1994) Staufen protein associates with the 3′ UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell*, **79**, 1221-1232.
19. Gong, C. and Maquat, L.E. (2011) LncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 39 UTRs via Alu eleme. *Nature*, **470**, 284-290.
20. Gong, C., Tang, Y. and Maquat, L.E. (2014) Erratum: MRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay(Nature Structural and Molecular Biology). *Nature Structural and Molecular Biology*, **21**, 1106.
21. Elbarbary, R.A., Li, W., Tian, B. and Maquat, L.E. (2013) STAU1 binding 3′ UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes and Development*, **27**, 1495-1510.
22. Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. and Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740-744.
23. Gatignol, A., Buckler-White, A., Berkhout, B. and Jeang, K.T. (1991) Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science*, **251**, 1597-1600.
24. Dorin, D., Bonnet, M.C., Bannwarth, S., Gatignol, A., Meurs, E.F. and Vaquero, C. (2003) The TAR RNA-binding protein, TRBP, stimulates the expression of TAR-containing RNAs in vitro and in vivo independently of its ability to inhibit the dsRNA-dependent kinase PKR. *Journal of Biological Chemistry*, **278**, 4440-4448.
25. Jain, A., Liu, R., Ramani, B., Arauz, E., Ishitsuka, Y., Ragunathan, K., Park, J., Chen, J., Xiang, Y.K. and Ha, T. (2011) Probing cellular protein complexes using single-molecule pull-down. *Nature*, **473**, 484-488.
26. Hwang, H., Kim, H. and Myong, S. (2011) Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proceedings of the National*

Academy of Sciences of the United States of America, 108, 7414-7418.

27. Gleghorn, M.L., Gong, C., Kielkopf, C.L. and Maquat, L.E. (2013) Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nature Structural and Molecular Biology*, **20**, 515-524.
28. Butcher, S.E. and Pyle, A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of chemical research*, **44**, 1302-1311.
29. Koh, H.R., Kidwell, M.A., Ragunathan, K., Doudna, J.A. and Myong, S. (2013) ATP-independent diffusion of double-stranded RNA binding proteins. *Proc Natl Acad Sci U S A*, **110**, 151-156.
30. Koh, H.R., Kidwell, M.A., Ragunathan, K., Doudna, J.A. and Myong, S. (2013) ATP-independent diffusion of double-stranded RNA binding proteins. *Proceedings of the National Academy of Sciences*, **110**, 151-156.
31. Hwang, H. and Myong, S. (2014) Protein induced fluorescence enhancement (PIFE) for probing protein-nucleic acid interactions. *Chemical Society reviews*, **43**, 1221-1229.
32. Yamashita, S., Nagata, T., Kawazoe, M., Takemoto, C., Kigawa, T., Güntert, P., Kobayashi, N., Terada, T., Shirouzu, M., Wakiyama, M. *et al.* (2011) Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Science*, **20**, 118-130.
33. Vukovic, L., Koh, H.R., Myong, S. and Schulten, K. (2014) Substrate recognition and specificity of double-stranded RNA binding proteins. *Biochemistry*, **53**, 3457-3466.
34. Yang, W., Wang, Q., Howell, K.L., Lee, J.T., Cho, D.S.C., Murray, J.M. and Nishikura, K. (2005) ADAR1 RNA deaminase limits short interfering RNA efficacy in mammalian cells. *Journal of Biological Chemistry*, **280**, 3946-3953.
35. Nie, Y., Ding, L., Kao, P.N., Braun, R. and Yang, J.H. (2005) ADAR1 interacts with NF90 through double-stranded RNA and regulates NF90-mediated gene expression independently of RNA editing. *Molecular and Cellular Biology*, **25**, 6956-6963.
36. Han,W., Wan,C.-K., Jiang,F. and Wu,Y.-D. (2010) PACE force field for protein simulations. 1. Full parameterization of version 1 and verification. J. Chem. Theory Comput., 6, 3373–3389.
37. Han,W. and Schulten,K. (2012) Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: improved backbone hydration and interactions between charged side chains. J. Chem. Theory Comput., 8, 4413–4424. 38. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E.,
38. Villa,E., Chipot,C., Skeel,R.D., Kale´,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. J. Comput. Chem., 26, 1781–1802.
39. Hornak,V., Abel,R., Okur,A., Strockbine,B., Roitberg,A. and Simmerling,C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins: Struct. Funct. Genet., 65, 712–725.

40. Pe´rez,A., Marcha´n,I., Svozil,D., Sponer,J., Cheatham Iii,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of conformers. Biophys. J., 92, 3817–3829.
41. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.,* **98,** 10089–10092.
42. Desterro,J.M., Keegan,L.P., Lafarga,M., Berciano,M.T., O'Connell,M. and Carmo-Fonseca,M. (2003) Dynamic association of RNA-editing enzymes with the nucleolus. *J. Cell Sci.,* **116,** 1805–1818.
43. Gleghorn,M.L., Gong,C., Kielkopf,C.L. and Maquat,L.E. (2013) Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat. Struct. Mol. Biol.,* **20,** 515–524.
44. Gong,C. and Maquat,L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature,* **470,** 284–288.
45. Gong,C., Tang,Y. and Maquat,L.E. (2013) mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat. Struct. Mol. Biol.,* **20,** 1214–1220.
46. Butcher,S.E. and Pyle,A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.,* **44,** 1302–1311.
47. Koh,H.R., Kidwell,M.A., Ragunathan,K., Doudna,J.A. and Myong,S. (2013) ATP-independent diffusion of double-stranded RNA binding proteins. *Proc. Natl. Acad. Sci. U.S.A.,* **110,** 151–156.
48. Yamashita,S., Nagata,T., Kawazoe,M., Takemoto,C., Kigawa,T., Gu¨ntert,P., Kobayashi,N., Terada,T., Shirouzu,M., Wakiyama,M. *et al.* (2011) Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci.,* **20,** 118–130.
49. Vukovic,L., Koh,H.R., Myong,S. and Schulten,K. (2014) Substrate recognition and specificity of double-stranded RNA binding proteins. *Biochemistry,* **53,** 3457–3466.
50. Yang,W., Wang,Q., Howell,K.L., Lee,J.T., Cho,D.S.C., Murray,J.M. and Nishikura,K. (2005) ADAR1 RNA deaminase limits short interfering RNA efficacy in mammalian cells. *J. Biol. Chem.,* **280,** 3946–3953.
51. Nie,Y., Ding,L., Kao,P.N., Braun,R. and Yang,J.H. (2005) ADAR1 interacts with NF90 through double-stranded RNA and regulates NF90-mediated gene expression independently of RNA editing. *Mol. Cell. Biol.,* **25,** 6956–6963.