The *bicoid* mRNA localization factor Exuperantia is an RNA-binding pseudonuclease

Daniela Lazzaretti[1], Katharina Veith[1,5], Katharina Kramer[2,3,6], Claire Basquin[4], Henning Urlaub[2,3], Uwe Irion[1] and Fulvia Bono[1,7]

[1]Max Planck Institute for Developmental Biology, Tübingen, Germany

[2]Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

[3]Bioanalytics, Institute for Clinical Chemistry, University Medical Center Göttingen, Germany

[4]Max Planck Institute of Biochemistry, Martinsried, Germany

[5]Present address: Institute for Biochemistry and Molecular Biology, University of Hamburg, Germany

[6]Present address: Max Planck Institute for Plant Breeding Research, Köln, Germany

[7]Correspondence should be addressed to F. B. (fulvia.bono@tuebingen.mpg.de)

Abstract

Anterior patterning in *Drosophila* is mediated by the localization of *bicoid* (*bcd*) mRNA at the anterior pole of the oocyte. Exuperantia (Exu) is a putative exonuclease (EXO) associated with *bcd* and required for its localization. We present the crystal structure of Exu that reveals a dimeric assembly with each monomer consisting of a 3'-5' EXO-like domain and a Sterile Alpha Motif (SAM)-like domain. The catalytic site is degenerated and inactive. Instead, the EXO-like domain mediates dimerization and RNA binding. We show that Exu binds RNA directly *in vitro,* that the SAM-like domain is required for RNA binding activity and that Exu binds to a structured element present in *bcd* 3'UTR with high affinity. Using structure-guided mutagenesis, we show that Exu dimerization is essential for *bcd* localization. Our data demonstrate that Exu is a non-canonical RNA binding protein with EXO-SAM-like domain architecture that interacts with its target RNA as a homodimer.

Introduction

Intracellular localization of mRNAs is a conserved mode of gene expression regulation in eukaryotes found in many cell types and processes[1-5]. mRNA localization targets the synthesis of specific proteins to their cytoplasmic site of function and provides fine temporal and spatial control of gene expression. As many copies of the encoded protein can be translated from a single mRNA molecule, the localization of specific transcripts can lead to a local enrichment of the protein and simultaneously prevents ectopic translation. mRNA localization is driven by the cooperative interactions between RNA signals on the localizing transcript, proteins that recognize these signals (mRNP components or *trans*-acting factors) and components of the cytoskeleton.

Drosophila melanogaster early development is a well-characterized model system for mRNA localization[6]. In the oocyte and early embryo, a substantial proportion of all mRNAs are localized[7,8]. The oocyte relies on nurse cells for the import of cytoplasmic proteins and RNAs[9,10]. In the oocyte, localized maternal mRNAs such as *bicoid (bcd)*, *gurken (grk)* and *oskar (osk)* determine the body axes of the embryo and future fly prior to fertilization[11].

Bcd, a transcription factor translated from the anteriorly localized *bcd* mRNA[12,13], generates a morphogen gradient along the embryonic anteroposterior axis, defining the head and thoracic segments[14]. Proper *bcd* localization involves several, partially redundant, steps[15]. In early oogenesis (stage 5-6), *bcd* is actively transported from nurse cells. By mid-oogenesis (stage 8), *bcd* is restricted to the anterior of the oocyte in a tight ring by an active mechanism that requires maternal Exuperantia (Exu)[15-17]. Finally, at stage 10b, *bcd* redistributes into a disc associated with the anterior cortex[18] and is translated after egg deposition and transcript polyadenylation[19-21].

Exu is thus known to be required for *bcd* localization from the earliest stages of oogenesis. In *exu*

mutants, *bcd* is transported to the oocyte but fails to accumulate at the anterior end, resulting in loss of head structures and an expanded thorax, a phenotype similar to weak *bcd* mutants[13,15,16,22]. Exu is required in nurse cells for the assembly of a *bcd* mRNP complex competent for correct localization[23]. Exu is enriched in the nurse cells at electron dense structures, the sponge bodies, and at tubular ER, and assembles into *bcd* mRNA containing particles that are transported to the oocyte in a dynein-dependent manner[17,23-28].

Exu co-purifies in a large ribonucleoprotein particle that includes the proteins Ypsilon schachtel (Yps), Me31B and Cup with *osk*, but not *bcd*, mRNA[29,30]. However, *exu* mutants show only transient mislocalization of *osk* mRNA that is compensated later in oogenesis[22,29].

Although *bcd* is a paradigm for mRNA localization studies and the requirement for Exu in this process has been known for almost 30 years, the molecular function of Exu is poorly understood.

Exu, which is conserved in insects, is predicted to have a 3'-5' exoribonuclease domain (EXO) from the DEDD exonuclease superfamily[31]. The EXO domain is found in several enzymes involved in RNA metabolism, including many essential in RNA and DNA maturation and degradation in prokaryotes and eukaryotes[32]. Whether exonucleolytic activity is important for Exu function is unknown.

Exu C-terminal region (CTR) is predicted to be natively unstructured and is phosphorylated by Par-1, a Ser/Thr protein kinase[33]. Phosphorylation site mutants have a weak *bcd* localization defect that is compensated later in oogenesis[33].

Exu has been shown to cross-link to RNA *in vitro*[34]. However, since it has no predicted canonical RNA binding motifs, it was proposed to associate with *bcd* via an adapter protein[34]. Neither the RNA-binding region nor the sequence specificity has been determined[25].

To further understand the role of Exu in mRNA localization we determined crystal structures of *Drosophila* Exu with and without the SAM-like domain. The structures show that the protein homodimerizes by interdigitating with itself. The EXO-like domain lacks the signature EXO sequence motifs and contributes instead to RNA binding. The SAM-like domain is also required for RNA binding and *in vivo* activity. We show that dimerization is essential for high affinity interactions with *bcd* 3'UTR motifs *in vitro* and Exu function *in vivo*.

Results

Structure determination and quality

Attempts to crystallize full-length Exu were unsuccessful. Limited proteolysis generates two stable fragments that are truncated from the C-terminal end of the protein. Therefore, deletion constructs of Exu were used that include the residues from 1 to 333 (Exu^{333}) and from 1 to 406 (Exu^{406}) (**Fig. 1a**). In the latter, a 60 amino acid (aa) non-conserved loop (loop 1), was mutated to a shorter (5 aa) loop present in the Exu ortholog from *Bombyx mori* (**Supplementary Fig. 1**). We determined the crystal structures of Exu^{333} and Exu^{406} at 2.37 Å and 2.80 Å resolution, respectively. The structure of SeMet-substituted Exu^{333} was solved by single wavelength anomalous dispersion (SAD). The refined Exu^{333} model has an *R*free of 22.3% and *R*factor of 20.4% with good stereochemistry (**Table 1**). In the final model, 25 residues at the N-terminus, 11 at the C-terminus and two loop regions, including loop1, are disordered (**Fig. 1b-c**). The Exu^{333} model was then used as molecular replacement (MR) model to obtain phases for native Exu^{406} crystals. Additional electron density observed in these crystals allowed the SAM-like C-terminal domain to be built. In the structure of Exu^{406}, the N- and C-termini and some loop regions, including loop1 in molecule A, are disordered (**Fig. 1d-e**). The asymmetric unit (ASU) of this

crystal form includes two molecules in a dimeric assembly. The refined Exu^{406} model has an *R*free of 27.6% and *R*factor of 25.2% with good stereochemistry (**Table 1**). In molecule B, the SAM-like domain could only be partially modeled, accounting for the relatively high *Rfree* and B factors for this chain (**Table 1**).

Overview of the structure

Exu is a dimer that resembles a Greek cross with all arms of similar length: 85 Å vertical width (distance between residue 335 of each SAM-like domain) and 86 Å horizontal width (distance between residue 278 of each EXO-like domain). The two domains in each monomer form orthogonal arms of the cross (**Fig. 1d**).

The EXO-like domain has a typical EXO fold, consistent with prior bioinformatic predictions[31]. A mixed α/β arrangement of secondary structure elements[35,36] is organized into a curved antiparallel β-sheet (made of 5 antiparallel β-strands) flanked on each side by 4 α-helices (**Fig. 1**). The β-sheet is capped on one side by a small α-helix and by an Exu-specific β-hairpin extension (**Fig. 1**).

The SAM-like domain is a 5 helix-bundle with structural similarity to the Sterile Alpha Motif (SAM)-like fold (SCOP 47768) **(Supplementary Fig. 2)**. A 32 Å-long helix connects the two domains. All independent observations of the EXO domain monomers (one molecule in Exu^{333} ASU and two molecules in Exu^{406} ASU) are highly similar (RMSD of 0.34 Å over 2441 atoms), suggesting that the EXO dimer acts as a rigid scaffold. Similarly, the linker helix packs tightly against the EXO domain and shows limited structural variation in the three monomers.

Exu is not an exonuclease

Database searches with PDBefold[37] detect high structural similarity of the EXO-like domain of Exu with the mouse 3'-5' DEDD exonuclease Trex1 (Z-score = 9.2) (**Fig. 2**). 3'-5' exonucleases digest nucleic acids from the 3' end hydrolyzing one nucleotide at a time using a two-metal ion mechanism[36,41-43]. The four conserved metal binding residues (DEDD) of the EXO family proteins and a general base residue are distributed across three identifiable motifs, Exo I, II and III (**Supplementary Fig. 3a**). Structure-based alignment with Trex1 and RNaseT and inspection of the putative active site of Exu reveal, however, that key catalytic residues are not conserved and the metal binding site is disorganized. This suggests that Exu is unlikely to perform nucleic acid degradation by the same mechanism as other EXO domain-containing proteins (**Fig. 2**; **Supplementary Fig. 3**). Two of the four acidic residues required for activity are present in Exu, Asp39 and Asp41, corresponding to Asp18 and Glu20 in mouse Trex1. However, residues equivalent to the remaining negatively charged residues (Asp130 and Asp200 in mouse Trex1) are a small hydrophobic (Ile144) and a positively-charged (Arg264) residue, respectively. Furthermore, the hydrophilic residue (His195 in mouse Trex1) involved in orienting the attacking water molecule is Gly259 in Exu. These alterations prevent metal ion coordination and hydrolysis as Arg264 interacts with Asp39 and Asp41; this renders Exu a pseudonuclease (**Fig. 2; Supplementary Fig. 4**). Consequently, we refer to this domain as EXO-like. The large non-conserved insertion at loop1 makes aligning primary sequences of Exu to other EXO proteins difficult, which is perhaps why we and others could not rule out Exu exonuclease activity based on primary sequence alone[31].

Exu forms a tight dimer via a finger-like β-hairpin loop

The extensive Exu dimer interface is made up of two separate regions with 2770 $Å^2$ of buried surface area from each monomer (16.8 % of monomer surface) (**Fig. 3**).

Dimerization is primarily mediated by the EXO-like domain, contributing 66% of the buried surface area (**Fig. 3**). In particular, a symmetrical interaction of the Exu-specific β-hairpin insertion in the EXO-like domain mediates dimer association. The inner surfaces of the β-hairpin loops face each other in an anti-parallel fashion and are bolstered by the helices on one side of the β-sheet (**Fig. 1, 3**). Contacts of the SAM-like domain to the EXO-like domain through the outer surface of the hairpin consolidate the interaction across the dimer (**Fig. 3a**). The homodimerization surfaces of Exu are distinct from those observed in other exonuclease dimers such as Trex1 or RNaseT (**Fig. 2a-b; Supplementary Fig. 3a**).

Conserved residues at the dimer interface stabilize the assembly mainly through hydrophobic interactions, hydrogen bonds and van der Waals contacts. In particular Arg92, conserved in all arthropods (**Supplementary Fig. 1**), forms a guanidinium π-stacked interaction with the symmetry-related Arg92 at the center of the interface (**Fig. 3b; Supplementary Fig. 4**). This interaction is stabilized by a hydrogen bond interaction with Asn72 and hydrophobic interactions with conserved Met68 and Met71 (**Fig. 3b**). Conserved Tyr70 and Tyr155 side chains in both monomers form a π-π interaction with each other (**Fig. 3d**). This interaction is consolidated by hydrogen bonding of Lys113 side chain to the backbone of Lys154.

A single point mutation, R92A, and a double mutant, M68A Y70A, are sufficient to convert the protein to a monomer in solution as demonstrated by multi-angle laser light scattering (MALLS) experiments coupled to size exclusion chromatography (SEC) (**Fig. 3e**). Wild-type (wt) Exu dimers elute as a monodisperse peak in SEC in the presence of 300 mM NaCl, indicating that the dimer is salt resistant and stable in solution, consistent with its hydrophobic character (**Fig. 3e**).

Exu is an RNA-binding protein

Co-localization studies *in vivo* and cross-linking experiments of cell lysates *in vitro* suggested that Exu associates with *bcd* mRNA[25,34]. Analysis of the charge distribution on the dimer surface reveals a high prevalence of positive electrostatic potential (**Fig. 4a-b**), consistent with nucleic acid binding. The positively-charged residues are grouped in two regions: a large surface on top of the dimer, spanning both monomers (**Fig. 4a**) and extending on the side of the EXO-like domain, and a smaller region at the bottom surface of the SAM-like domain (**Fig. 4b**). This surface charge distribution mirrors the distribution of conserved residues, underscoring their functional importance (**Fig. 4c-d**).

SEC experiments of Exu alone and in complex with nucleic acids show that Exu forms a stable complex with ssRNA **(Fig. 5a)**. Analysis of SEC-purified Exu in complex with $(U)_{20}$ by UV-cross-linking combined with mass-spectrometry (MS)[38] revealed a pattern of cross-linked sites broadly corresponding to the conserved, positively-charged surfaces of Exu. Many of the identified residues are basic and hydrophobic and are found on both the EXO-like domain and the SAM-like domain (**Fig. 4-5; Supplementary Table 1**).

Technical issues prevented the use of EMSA and ITC to assess Exu RNA binding affinities. We used fluorescence anisotropy (FA) to measure the affinity of RNA binding interactions. The affinity of wt Exu for $(U)_{20}$ ssRNA is in the sub-micromolar range (94±22 nM) (**Fig. 5d, f**). Exu^{406} binds RNA with a somewhat reduced affinity compared to full-length protein (226±30 nM) (**Fig. 5d, f**). This is consistent with the cross-linking of residues in the CTR **(Supplementary Table 1)**; these residues are not conserved nor embedded in any recognizable motif (**Supplementary Fig. 1**). Removal of the SAM-like domain (Exu^{333}), however, while not affecting dimerization (**Fig. 5i**), reduces the affinity of Exu for the RNA by almost 50-fold (4250±192 nM) (**Fig. 5d, f**). As the SAM-like domain is unstable in isolation and could not be purified, no direct affinity measurements could be made.

A known *exu* loss-of-function allele (*exu*^{PJ42} or *exu*^{1}) bears an R339S mutation at a conserved arginine in the SAM-like domain[39] (**Supplementary Fig. 1**). This substitution, and all others discussed below, does not impair protein folding (**Supplementary Fig. 5e**). However, it causes a small, reproducible decrease in RNA-binding affinity (**Fig. 5e-f**). Affinity is further decreased by a substitution with a more electronegative residue (R339E) **(Supplementary Fig. 5c-d)** and an additional point mutation at a cross-linked residue on the SAM-like domain (R339E R325E) (**Fig. 5e-f**). This is consistent with MS data that identified a UV-induced cross-link of Arg339 to $(U)_{20}$ RNA (**Fig. 5c**).

The SAM-like domain of Exu is structurally related to the SAM-like domain fold, including the CTD domains of bacterial RNA polymerases and the SAM domains of transcription regulators. Both superfamilies function in protein-protein and protein-nucleic acid interactions. A structure of Vts1 (the yeast ortholog of Smaug), a SAM domain-containing protein, in complex with an RNA hairpin suggested residues that might be involved in RNA binding in Exu. Point mutations to equivalent residues in Exu had no marked effect on RNA binding, showing that the RNA binding surface of Exu SAM-like domain differs from that of Vts1 (**Supplementary Fig. 5**). Moreover, other SAM-like domains lack the region where Arg339 resides. Therefore, Exu has unique RNA-binding features not observed in related structures (**Supplementary Fig. 2**).

In addition, mutants of cross-linked and conserved residues on the EXO domain (R84E K95E R141E) display reduced affinity for the RNA (**Fig. 5e-f**). Therefore, RNA binding is mediated by both domains of Exu, including a contribution from Arg339 on the SAM-like domain.

We characterized length requirements for ssRNA binding to Exu. A $(U)_8$ sequence binds with low affinity, whereas increasing the RNA length to 30 and 50 nt leads to a considerable increase in binding affinity without inducing Exu multimerization (**Supplementary Fig. 5f-g; 6**), indicating the engagement of additional contacts on Exu.

Taken together, the *in vitro* binding and cross-linking MS data suggest that Exu might bind RNA directly *in vivo*.

Exu dimer discriminates RNA secondary structure elements

To further characterize Exu nucleic acid binding, we measured binding affinities to 20 nt oligomers of different sequence composition **(Fig. 6)** by FA. Exu binds to $(U)_{20}$, $(UC)_{10}$, ssRNA and ssDNA of mixed sequence with a similar affinity. While it is unclear whether ssDNA is a biologically relevant substrate, this observation is consistent with the surface properties of Exu **(Fig. 4)**; that Exu binds weakly to dsDNA supports its cytoplasmic localization and function[25] (**Supplementary Fig. 7**). Exu has 6-fold lower affinity for $(UA)_{10}$ and 7-fold higher affinity for $(UG)_{10}$ as compared to $(U)_{20}$ **(Fig. 6a-d)**. While these differences in affinity may reflect a linear sequence preference, both UA- and UG-repeated oligomers are predicted to fold into secondary structures **(Supplementary Fig. 6a),** suggesting that Exu may bind preferentially to structured RNA.

Indeed, the 3'UTR of *bcd* is highly structured and subdivided into functional domains[40,41] **(Figure 6e)**. One domain, which includes stem loops IV and V, is sufficient to recapitulate early steps of *bcd* localization[42,43]. A sub-region, stem loop Vb (overlapping with the *bcd* localization element 1, BLE1), is essential for localization[44]. Perturbation of this region affects localization at a similar step as Exu requirement, consistent with a functional interaction[44,34,43].

As Exu has preferential binding for structured RNA sequences *in vitro* and is thought to associate with *bcd* mRNA *in vivo*, we measured Exu binding to the *bcd*-Vb element. Exu binds *bcd*-Vb with high affinity, 15±4 nM **(Fig. 5-6).** An antisense oligo, that retains similar secondary structure, binds with similar affinity to Exu **(Fig. 6**; **Supplementary Fig. 6a),** indicating that RNA conformation rather than primary sequence may be required in *bcd* BLE1 recognition. A

perfect dsRNA probe and the short ssRNA $(U)_8$, which are similar lengths to the *bcd*-Vb stem and loop respectively, bind poorly to Exu, indicating that both elements are important for the recognition (**Fig. 6**; **Supplementary Fig. 5**). The binding is selective, as the localization signal on *K10* mRNA (*K10* transport/localization element, TLS), recognized by Egalitarian (Egl), another EXO-containing protein, binds 6-fold less strongly[45,46] **(Fig. 6)**.

All Exu truncations and mutations to RNA cross-linked residues have decreased affinity for $(U)_{20}$ and for *bcd*-Vb compared with wt Exu, suggesting an overlap of binding surfaces on Exu **(Fig. 5, 6)**. However, two independent dimerization mutants (M68A Y70A and R92A) bind *bcd*-Vb RNA with reduced affinity of 7- and 20-fold, respectively. In contrast, the same dimerization mutants bind $(U)_{20}$ RNA with a similar affinity as wt (110±6 and 34±4 nM, respectively) (**Fig. 6f-h**), showing that Exu homodimerization is specifically required for high affinity binding of *bcd*-Vb.

Moreover, static light scattering experiments coupled to SEC suggest a stoichiometry of two *bcd*-Vb (or two $(U)_{20}$) molecules per Exu dimer, while 50 nt ssRNA binds with 1:1 stoichiometry (**Supplementary Fig. 6**).

These data suggest that Exu distinguishes RNA secondary structure elements required for *bcd* localization and that dimerization promotes binding to a physiological target.

Exu SAM-like domain and dimerization are required *in vivo*

To gain further insights into Exu function we set up a structure-guided genetic rescue assay *in vivo*. We modified a genomic fragment containing the *Exu* gene[25] to express an N-terminally Venus-tagged Exu at physiological levels. The constructs, encoding wt, mutant or truncated Exu proteins, were injected to generate transgenic flies and recombined into a landing site (*attP*) on the same position on chromosome 2, in a background lacking a 329 Kb genomic stretch on chromosome 2R encompassing the *exu* gene (*Df(2R)exu1,* here Df). In the rescue experiment,

these transgenic lines were crossed to flies heterozygous for a known *exu*-null allele, *exu*$^{VL\ 47}$, to obtain progeny devoid of endogenous Exu and expressing the different Venus-tagged Exu constructs at similar levels (**Fig. 7j**). Early embryos were collected from these females and *bcd* mRNA was visualized by *in situ* hybridization.

As expected, anterior localization of *bcd* is impaired in *exu*-null (Df/*exu*^{VL}) early embryos, and this phenotype is fully rescued by Venus-Exu wt expression (FL; **Fig. 7a-c**). No defects in *osk* mRNA localization were observed (**Supplementary Fig. 7**), in agreement with previous studies[22,29]. A hybrid Exu protein, containing the loop1 substitution of the corresponding *Bombyx mori* sequence (**Supplementary Fig. 1**), retains full rescue activity (**Fig. 7d**). Similarly, an Exu protein mutated in the two residues conserved in other exonucleases (**Fig. 2**) is also fully competent for *bcd* localization (**Fig. 7e**), confirming that catalytic activity is not important for function. A corresponding mutation in other systems generated catalytically dead exonucleases and was used to show that exonuclease activity is not relevant for Egl function *in vivo*[48]. In contrast, mutations that disrupt Exu dimerization (**Fig. 3**) impair its ability to localize *bcd* mRNA, showing that the dimer is the functional unit of Exu *in vivo* (**Fig. 7f-g**). A construct including the structured domains, and lacking the CTR (Exu^{410}), is sufficient to achieve proper *bcd* localization (**Fig. 7h**), indicating that our structure contains the essential functional features of Exu. In contrast, the EXO-like domain alone (Exu^{333}) does not rescue *bcd* localization (**Fig. 7i**), highlighting the functional importance of the SAM-like domain. A point mutation in this domain (R339S; **Fig. 5c**) also abolishes Exu function[39].

Exu has been shown to co-localize with *bcd* mRNA throughout early and mid-oogenesis[15,25,39]. We therefore examined the localization of our Venus-tagged Exu constructs at different stages of egg development (**Supplementary Fig. 7**). Venus-Exu FL is transported into the oocyte in pre-vitellogenic egg chambers, concentrates at both anterior and posterior poles of the oocyte at stage

8-10 and is enriched in cytoplasmic foci in the nurse cells, consistent with previous reports[25,39]. Most of the analyzed constructs did not show obvious defects in Exu distribution (**Supplementary Fig. 7**). However, Exu^{333} failed to localize anteriorly in more than half of the examined oocytes (**Supplementary Fig. 7**). This correlates with the severe loss in RNA-binding affinity observed *in vitro* (**Fig. 5**).

Discussion

Exu is a pseudonuclease

Our structure shows that Exu is a homodimeric multidomain protein consisting of an EXO-like domain coupled to a SAM-like domain.

The presence of an EXO domain in a protein involved in RNA metabolism generally implies a functional role in RNA degradation. However, the structural and *in vivo* rescue analyses reveal that the EXO-like domain lacks a classical exonucleolytic active site and that mutation of residues conserved between Exu and active exonucleases has no appreciable impact on function. Instead, the EXO-like domain provides a different essential function for Exu, that of homodimerization. Exu monomers interact via conserved, hydrophobic and electropositive surfaces, suggesting that the dimer acts as a unitary structural and functional platform. Structure-based mutants that prevent dimer formation are not functional *in vivo,* showing that Exu homodimerization is essential for *bcd* mRNA localization.

Exu is an RNA-binding protein

Our cross-linking MS and *in vitro* binding studies show that Exu is an RNA binding protein and that RNA binding residues lie within conserved, electropositive regions, suggesting a path of the bound RNA on the surface. Combining these studies with structure-guided mutations, we show

that both the EXO-like and SAM-like domains contribute to RNA binding. The dramatic reduction in RNA binding affinity on removal of the SAM-like domain highlights its major contribution to RNA binding, either alone or as a composite surface with the EXO-like domain. An Exu mutant lacking the SAM-like domain dimerizes *in vitro* but is non-functional *in vivo*.

A single point mutation in the SAM-like domain, present in a previously described *exu* allele (R339S), impairs *bcd* localization[39]. This residue directly cross-links to RNA and its mutation consistently reduces RNA binding by 2-fold *in vitro*. While it is possible that this reduction would account for the loss-of-function phenotype *in vivo*, we cannot rule out that interactors apart from *bcd* mRNA are affected by this mutation.

Exu discriminates secondary structure elements and functions as a dimer in vivo

Here we show direct, high affinity binding of Exu to the *bcd* 3'UTR localization signal, BLE1, in agreement with data in flies showing that Exu and BLE1 are required in early stages of *bcd* localization[44,34,43]. The low affinity of Exu for an unrelated mRNA localization signal further implies that Exu-BLE1 recognition is specific, although additional factors might be involved.

We also observe that Exu binds ssRNA, albeit with lower affinity, and that Exu binds preferentially particular sequences. Whether these sequences are recognized directly or have some intrinsic propensity to form secondary structures requires further investigation. In light of our findings that Exu has high affinity for the *bcd*-Vb element, the latter explanation seems more likely. However, we observe that long ssRNAs can also associate with Exu with high affinity. This might reflect an additional role of Exu in ssRNA packaging in the context of RNP transport particle formation. This mode of binding could be exploited by other RNAs that lack a structured signal (**Fig. 7k**).

Monomeric Exu mutants bind short RNA sequences with similar affinity to wild-type dimers but are specifically impaired in *bcd*-Vb stem loop recognition and are not functional *in vivo*. Our observations that RNA binding extends across both Exu monomers (from cross-linking and MS) and that Exu dimers likely bind two molecules of *bcd*-Vb RNA supports a role of Exu as a unitary platform for RNA binding. Our *in vivo* data show that RNA binding, like dimerization, is essential and these activities of Exu are coupled.

Further studies are required to determine other physiological RNA targets of Exu, to characterize RNA binding and to explore its connection to the localization machinery.

Evolution of Exu

From sequence analysis, we found Exu homologs where conserved catalytic residues are retained in some lophotrocozoans (molluscs) and vertebrate sequences, indicating that these are active exonucleases (**Supplementary Fig. 1**). This suggests that Exu evolved from active exonucleases and lost its catalytic activity along the arthropod stem lineage. Other EXO domain proteins have been shown to bind RNA[45,49,50]. In particular, Maelstrom was recently shown to lack canonical EXO activity and features a Zn^{2+}-binding insertion that plays a functional role in RNA binding and the piRNA pathway[49,51]. Furthermore, in the *Drosophila* mRNA localization pathway, Egl is predicted to contain an EXO domain. Egl function in dynein-dependent transport of localized mRNAs does not require an intact catalytic site[45,48,52]. Therefore, the loss of EXO activity and acquisition of RNA-binding function appear to have occurred independently in distantly related families of EXO-containing proteins.

The presence of a SAM-like domain C-terminal to the EXO-like domain is unique to arthropods, suggesting that loss of the catalytic site coincided with recruitment of a SAM-like domain. We

also found that the β-hairpin loop that mediates dimerization in *Drosophila* Exu is conserved in arthropods but not outside arthropods. In Exu evolution as an RNA binding scaffold, we can identify three steps: loss of canonical EXO catalytic activity, acquisition of a SAM-like domain and insertion of a loop to mediate dimerization. Despite *bcd* being present only in high dipterans[53,54], all insect lineages appear to have a conserved, catalytically inactive Exu ortholog, suggesting a broader function in the regulation of RNA expression.

Accession codes

The coordinates and structure factors have been deposited in the Macromolecular Structure Database of European Bioinformatic Institute (EBI) with ID code 5L7Z and 5L80 for Exu^{333} and Exu^{406}, respectively.

Acknowledgements

We wish to thank the MPI-Martinsried Crystallization Facility. We also thank the staff at the Swiss Light Source synchrotron for assistance during data collection, S. Grüner, E. Khazina and V. Ahl for assistance with MALLS measurements and D. Hildebrand and N. Weiss for help with Ab production. We thank A. Cook, E. Lorentzen and E. Conti for discussion and critical reading of the manuscript. This project has received funding from the Max Planck Gesellschaft, the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC grant agreement n° 310957 and the Deutsche Forschungsgemeinschaft (SFB860 to K.K. and H.U. and BO3588/2-1 to F.B.).

Author Contributions

Biochemical, biophysical and crystallization work were done by D.L. and K.V.; fly work by D.L and U.I.; K.K. and H.U. carried out the MS analysis; D.L. and C.B. analyzed FA data; F.B. solved the structures and supervised the project. F.B., D.L and U.I. wrote the paper.

Competing financial interests

The authors declare no competing financial interests.

References

1. Gonsalvez, G.B. & Long, R.M. Spatial regulation of translation through RNA localization. *F1000 biology reports* **4**, 16 (2012).
2. Sutton, M.A. & Schuman, E.M. Dendritic protein synthesis, synaptic plasticity, and memory. *Cell* **127**, 49-58 (2006).
3. Bramham, C.R. & Wells, D.G. Dendritic mRNA: transport, translation and function. *Nature reviews Neuroscience* **8**, 776-789 (2007).
4. Besse, F. & Ephrussi, A. Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nature Reviews Molecular Cell Biology* **9**, 971-980 (2008).
5. Jung, H., Gkogkas, C.G., Sonenberg, N. & Holt, C.E. Remote control of gene function by local translation. *Cell* **157**, 26-40 (2014).
6. Kugler, J.-M. & Lasko, P.F. Kugler, Jan-Michael and Lasko Paul. *Fly*, 1-14 (2009).
7. Lécuyer, E. et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174-187 (2007).
8. Jambor, H. et al. Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. *eLife* **4**(2015).
9. Spradling, A.C. Germline cysts: communes that work. *Cell* **72**, 649-651 (1993).
10. Mahajan-Miklos, S. & Cooley, L. The villin-like protein encoded by the Drosophila quail gene is required for actin bundle assembly during oogenesis. *Cell* **78**, 291-301 (1994).
11. Ephrussi, A. & St Johnston, D. Seeing is believing: the bicoid morphogen gradient matures. *Cell* **116**, 143-152 (2004).
12. Frohnhöfer, H.N.-V., C Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid. *Nature* **334**, 120-125 (1986).
13. Nüsslein-Volhard, C., Frohnhöfer, H.G. & Lehmann, R. Determination of anteroposterior polarity in Drosophila. *Science (New York, NY)* **238**, 1675-1681 (1987).
14. Driever, W. & Nüsslein-Volhard, C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**, 95-104 (1988).

15. St Johnston, D., Driever, W., Berleth, T., Richstein, S. & Nüsslein-Volhard, C. Multiple steps in the localization of bicoid RNA to the anterior pole of the Drosophila oocyte. *Development (Cambridge, England)* **107 Suppl**, 13-19 (1989).
16. Berleth, T. et al. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *The EMBO Journal* **7**, 1749-1756 (1988).
17. Mische, S., Li, M., Serr, M. & Hays, T.S. Direct observation of regulated ribonucleoprotein transport across the nurse cell/oocyte boundary. *Molecular biology of the cell* **18**, 2254-2263 (2007).
18. Schnorrer, F., Luschnig, S., Koch, I. & Nüsslein-Volhard, C. Gamma-tubulin37C and gamma-tubulin ring complex protein 75 are essential for bicoid RNA localization during drosophila oogenesis. *Developmental cell* **3**, 685-696 (2002).
19. Driever, W. & Nüsslein-Volhard, C. A gradient of bicoid protein in Drosophila embryos. *Cell* **54**, 83-93 (1988).
20. Driever, W., Ma, J., Nüsslein-Volhard, C. & Ptashne, M. Rescue of bicoid mutant Drosophila embryos by bicoid fusion proteins containing heterologous activating sequences. *Nature* **342**, 149-154 (1989).
21. Sallés, F.J., Lieberfarb, M.E., Wreden, C., Gergen, J.P. & Strickland, S. Coordinate initiation of Drosophila development by regulated polyadenylation of maternal messenger RNAs. *Science (New York, NY)* **266**, 1996-1999 (1994).
22. Schüpbach, T. & Wieschaus, E. Germline autonomy of maternal-effect mutations altering the embryonic body pattern of Drosophila. *Developmental Biology* **113**, 443-448 (1986).
23. Cha, B.J., Koppetsch, B.S. & Theurkauf, W.E. In vivo analysis of Drosophila bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway. *Cell* **106**, 35-46 (2001).
24. Pokrywka, N.J., Payne-Tobin, A., Raley-Susman, K.M. & Swartzman, S. Microtubules, the ER and Exu: new associations revealed by analysis of mini spindles mutations. *Mechanisms of development* **126**, 289-300 (2009).
25. Wang, S. & Hazelrigg, T. Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature* **369**, 400-403 (1994).
26. Wilsch-Bräuninger, M., Schwarz, H. & Nüsslein-Volhard, C. A sponge-like structure involved in the association and transport of maternal products during Drosophila oogenesis. *The Journal of Cell Biology* **139**, 817-829 (1997).
27. Theurkauf, W.E. & Hazelrigg, T. In vivo analyses of cytoplasmic transport and cytoskeletal organization during Drosophila oogenesis: characterization of a multi-step anterior localization pathway. *Development (Cambridge, England)* **125**, 3655-3666 (1998).
28. Nakamura, A., Amikura, R., Hanyu, K. & Kobayashi, S. Me31B silences translation of oocyte-localizing RNAs through the formation of cytoplasmic RNP complex during Drosophila oogenesis. *Development (Cambridge, England)* **128**, 3233-3242 (2001).
29. Wilhelm, J.E. et al. Isolation of a ribonucleoprotein complex involved in mRNA localization in Drosophila oocytes. *The Journal of Cell Biology* **148**, 427-440 (2000).
30. Wilhelm, J.E., Hilton, M., Amos, Q. & Henzel, W.J. Cup is an eIF4E binding protein required for both the translational repression of oskar and the recruitment of Barentsz. *The Journal of Cell Biology* **163**, 1197-1204 (2003).
31. Moser, M.J., Holley, W.R., Chatterjee, A. & Mian, I.S. The proofreading domain of Escherichia coli DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic acids research* **25**, 5110-5118 (1997).
32. Ibrahim, H., Wilusz, J. & Wilusz, C.J. RNA recognition by 3'-to-5' exonucleases: the substrate perspective. *Biochimica et biophysica acta* **1779**, 256-265 (2008).

33. Riechmann, V. & Ephrussi, A. Par-1 regulates bicoid mRNA localisation by phosphorylating Exuperantia. *Development (Cambridge, England)* **131**, 5897-5907 (2004).
34. Macdonald, P.M., Leask, A. & Kerr, K. exl protein specifically binds BLE1, a bicoid mRNA localization element, and is required for one phase of its activity. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 10787-10791 (1995).
35. de Silva, U. et al. The crystal structure of TREX1 explains the 3' nucleotide specificity and reveals a polyproline II helix for protein partnering. *The Journal of biological chemistry* **282**, 10537-10543 (2007).
36. Hsiao, Y.-Y. et al. Structural basis for RNA trimming by RNase T in stable RNA 3'-end maturation. *Nature chemical biology* **7**, 236-243 (2011).
37. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica Section D, Biological crystallography* **60**, 2256-2268 (2004).
38. Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods* **11**, 1064-1070 (2014).
39. Marcey, D., Watkins, W.S. & Hazelrigg, T. The temporal and spatial distribution pattern of maternal exuperantia protein: evidence for a role in establishment but not maintenance of bicoid mRNA localization. *The EMBO Journal* **10**, 4259-4266 (1991).
40. Macdonald, P.M. bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure. *Development (Cambridge, England)* **110**, 161-171 (1990).
41. Brunel, C. & Ehresmann, C. Secondary structure of the 3′ UTR of bicoid mRNA. *Biochimie* **86**, 91-104 (2004).
42. Macdonald, P.M. & Struhl, G. cis-acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. *Nature* **336**, 595-598 (1988).
43. Macdonald, P.M. & Kerr, K. Redundant RNA recognition events in bicoid mRNA localization. *RNA (New York, NY)* **3**, 1413-1420 (1997).
44. Macdonald, P.M., Kerr, K., Smith, J.L. & Leask, A. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development (Cambridge, England)* **118**, 1233-1243 (1993).
45. Dienstbier, M., Boehl, F., Li, X. & Bullock, S.L. Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes and Development* **23**, 1546-1558 (2009).
46. Bullock, S.L., Ringel, I., Ish-Horowicz, D. & Lukavsky, P.J. A′-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. *Nature Publishing Group* **17**, 703-709 (2010).
47. Hazelrigg, T. et al. The exuperantia gene is required for Drosophila spermatogenesis as well as anteroposterior polarity of the developing oocyte, and encodes overlapping sex-specific transcripts. *Genetics* **126**, 607-617 (1990).
48. Navarro, C., Puthalakath, H., Adams, J.M., Strasser, A. & Lehmann, R. Egalitarian binds dynein light chain to establish oocyte polarity and maintain oocyte fate. *Nature cell biology* **6**, 427-435 (2004).
49. Chen, K.-M. et al. Metazoan Maelstrom is an RNA-binding protein that has evolved from an ancient nuclease active in protists. *RNA (New York, NY)* **21**, 833-839 (2015).
50. Yang, Z. et al. PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway. *Molecular Cell* **61**, 138-152 (2016).
51. Matsumoto, N. et al. Crystal Structure and Activity of the Endoribonuclease Domain of the piRNA Pathway Factor Maelstrom. *Cell reports* **11**, 366-375 (2015).

52. Mach, J.M. & Lehmann, R. An Egalitarian-BicaudalD complex is essential for oocyte specification and axis determination in Drosophila. *Genes and Development* (1997).
53. Lynch, J. & Desplan, C. 'De-evolution' of Drosophila toward a more generic mode of axis patterning. *The International journal of developmental biology* **47**, 497-503 (2003).
54. Lemke, S. et al. Bicoid occurrence and Bicoid-dependent hunchback regulation in lower cyclorrhaphan flies. *Evolution & development* **10**, 413-420 (2008).
55. Glaser, F. et al. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163-164 (2003).

Figure 1

Structure of Exu dimer and constructs used.

a) Schematic representation of the architecture of the proteins used in this study. Solid and dashed boxes represent folded domains and loop1 region, respectively. Thin lines are regions predicted to be unstructured. The constructs used are indicated in gray.

b-e) Structure of Exu^{333} (**b-c**) and Exu^{406} (**d-e**) homodimers in two views related by a 90° rotation about the horizontal axis with monomer A (red), monomer B (gray) and the linker helix (teal) shown as cartoons. Dotted lines represent loop connections that are not visible in the electron density. Highlighted in orange is loop1, substituted in Exu^{406} structure and not ordered in the Exu^{333} structure. The EXO-like domains form a continuous horizontal bar with the β-hairpin loops facing each other. **c, e**) Lateral view of the dimeric complex showing the interaction interface between the helices of the EXO-like domains.

These and all other protein structure figures were generated using PyMOL (http://www.pymol.org).

Figure 2

The EXO-like domain of Exu lacks residues required for exonuclease activity.

a) Catalytic center of Trex1 (PDB: 2IOC[35,36]) bound to AMP in a post-hydrolytic state. An overview of the dimer assembly is displayed on the left; one of the monomers (dashed box) is enlarged on the right with a black box indicating the region enlarged in the representation below. D18, E20, D130 and D200 coordinate the two catalytic Mn^{2+} ions (purple spheres) (**c**). **b**) Putative catalytic center of Exu[333], in a similar representation to (**a**) with the corresponding residues in the EXO-like domain of Exu labeled. Dimer assembly is displayed on the left, with Exu monomer A (in red) in the same orientation as Trex1 monomer A (dark gray, **a**). **c-d**) Detailed views of the catalytic site of Trex1 and Exu, respectively. Acidic residues that match the DEDD consensus in the putative active site of Exu are indicated with black arrowheads.

Figure 3

Conserved residues mediate Exu dimerization.

a, c) View of the monomer-monomer interface. Exu is shown in the same orientation as in **Fig. 1d** and **1e**, respectively. Molecule B is rendered as a gray surface, with interacting residues highlighted in teal. The interaction site is boxed in black. **b, d**) Zoomed-in views of the dimerization interface (in the same orientation as in **a** and **c**, respectively). Residues mutated in **e** are indicated with red and pink arrowheads, respectively. **e**) SEC and calculated molecular weight (obtained from MALLS) of Exu wt and of two dimerization mutants.

Figure 4

Surface properties of the Exu dimer.

Surface representation of Exu in the same orientation as in **Fig. 1d** (**a, c**) and rotated 180º along the x axis (**b, d**). **a-b**) The surface is colored according to electrostatic potential with positively charged residues in blue. **c-d**) Exu surface rendering coloured according to conservation with a

gradient from white to green indicating increasingly conserved residues. Conservation scores were calculated from the arthropod sequences included in the multiple sequence alignment shown in **Supplementary Fig. 1**, using ConSurf[55]. The surface representations were generated from the coordinates of molecule A with all the side chains modelled and duplicated using the dimer symmetry.

Figure 5

Multiple residues on Exu surface are involved in RNA binding.

a) SEC profiles of purified Exu wt (left), of Exu wt pre-incubated with $(U)_{20}$ RNA (middle) and of $(U)_{20}$ RNA alone (right). The position of the peak containing the free RNA is indicated by a dotted line and the elution volume of Exu wt by a gray line.

b, c) Surface rendering of the Exu homodimer with the monomers in two shades of gray. The surface representations were generated as in **Fig. 4**, and in the same orientation as in **Fig. 1d** (**b**) or rotated 180º along the x axis (**c**). Residues UV-cross-linked to the RNA and identified by MS are highlighted in red and labeled.

d-h) K_d values determined by FA. A constant amount of 5'-fluorescein labelled $(U)_{20}$ (**d-f**) or *bcd*-Vb hairpin (**f-h**) was incubated with increasing concentrations of the indicated recombinantly purified Exu constructs. The FA data were fitted to the Hill equation to obtain the dissociation constants (K_d); mean K_d and standard deviation are reported in table (**f**). Data are from three independent experiments, apart from R325E R339E + $(U)_{20}$ (four); Exu^{333} and R339S + $(U)_{20}$ (five); Exu wt + $(U)_{20}$ (nine), Exu wt + *bcd*-Vb (fifteen independent experiments). **d-e, g-h**) Data from representative FA measurements, with the best fit plotted as a solid line. **i**) SEC and calculated molecular weight (obtained from MALLS) of Exu^{333}.

Figure 6

Specificity of Exu binding to the RNA.

a-d, f-h) K_d values determined by FA. A constant amount of 5'-fluorescein labeled oligonucleotide of the indicated sequence (**c, h**) was incubated with increasing concentrations of recombinantly purified Exu wt (**a-d**) or mutants (**f-h**). The FA data were fitted to the Hill equation to obtain the dissociation constants (K_d); mean K_d and standard deviation are reported in tables (**c, h**). Data are from three independent experiments, apart from Exu wt + $(UA)_{10}$, ssDNA and *K10* TLS (four); Exu wt + *bcd*-Vb antisense (five); Exu wt + $(U)_{20}$ (nine), Exu wt + *bcd*-Vb (fifteen independent experiments). **a-b, f-g**) Data from representative FA measurements, with the best fit plotted as a solid line. **d**) Column graph of the data in table (**c**), plotting mean K_d (bars), standard deviation (black lines) and K_d values obtained for each measurement (gray dots). Numbers above the bars indicate the number of independent experiments performed. The dotted line is drawn at the mean K_d of Exu for *bcd*-Vb RNA (red bar). **e**) Schematic representation of *bcd* 3'UTR secondary structure, as described[41]. The BLE1 is highlighted in red and the *bcd*-Vb hairpin used in the FA measurements is boxed.

Figure 7

***Drosophila* phenotypes and *bicoid* localization.**

a-i) *bcd in situ* hybridizations of *Drosophila* early embryos (0-2 h). Numbers at the top right corner indicate the number of embryos displaying the illustrated phenotype vs the total number of embryos examined. The genotype of the analyzed embryos is reported at the bottom right of each image. All alleles are in *trans*-heterozygous combination with exu^{VL}: **a**) wt; **b**) *exu* deficiency *Df(2R)exu1* (Df); **c-i**) embryos expressing Venus-Exu transgenic constructs: **c**) full length (FL); **d**) loop1 sequence substituted with the corresponding residues from *Bombyx mori* (LDADS; see

also **Supplementary Fig. 1**); **e**) putative catalytic site mutant (D39A D41A); **f-g**) two dimerization mutants (M68A Y70A and R92A); **h-i**) C-terminal truncations containing both EXO-like and SAM-like domain (Exu^{410}) or EXO-like domain alone (Exu^{333}). Scale bar: 50 μm.

j) Western Blot showing the expression levels of endogenous Exu (in exu^{VL} heterozygous flies) and of the different transgenic Venus-Exu constructs (as in **c**-**i**) in ovaries. The blot was revealed with anti-GFP (top panel) and anti-Exu (middle panel) antibodies; anti-α-tubulin (bottom panel) was used as a loading control. Original images of the blots can be found in **Supplementary Data Set 1**.

k) Proposed model of Exu RNA binding.

Table 1

Data collection and refinement statistics of the crystal structures of Exu^{333} and of Exu^{406}.

	Native Exu^{333}	SeMet Exu^{333}		Exu^{406}
Data collection				
Space group	$P6_122$	$P6_122$		$P2_1$
Cell dimensions				
a, *b*, *c* (Å)	79.9, 79.9, 237.3	79.8, 79.8, 238.2		79.5, 66.6, 81.3
α, β, γ (°)	90, 90, 120	90, 90, 120		90, 90, 113.6
		Peak	*Inflection*	
Wavelength	1.000	0.980	1.040	1.000
Resolution (Å)	50-2.37 (2.47-2.37)	50-2.5 (3.00-2.5)	50-3.0 (3.3-3.0)	50-2.8 (2.9-2.8)
R_{sym} or R_{merge}	0.085 (0.972)	0.124 (0.693)	0.142 (0.653)	0.078 (1.045)
$I / \sigma I$	23.98 (3.2)	26.5 (7.3)	22.3 (6.2)	19.4 (2.2)
Completeness (%)	100 (100)	100 (100)	100 (100)	99.5 (99.2)
Redundancy	19.0 (19.6)	41.8 (42.0)	20.8 (20.1)	6.7 (6.6)
Refinement				
Resolution (Å)	40-2.37			47.3-2.8
No. reflections	18142			18391
R_{work} / R_{free}	20.4/22.3			25.2/27.6
No. atoms	3610			8456
Protein	3543			8445
Water	67			11
B-factors (A^2)	65.5			80.7
Protein	70.2			Chain A 76.7 Chain B 99.5
Water	60.9			66.0

R.m.s deviations		
Bond lengths (Å)	0.002	0.003
Bond angles (°)	0.468	0.658

One native crystal for each construct and one SeMet crystal were used for data collection. Values in parentheses are for the highest-resolution shell.

Online Methods

Protein expression and purification

Exu from *Drosophila melanogaster* was cloned in a pET-MCN vector, derived from the pET Novagen series[56]. All Exu constructs were cloned in-frame with an N-terminal glutatione S-transferase (GST) tag; Exu wt was also cloned in-frame with a hexahistidine (His) tag. The protein was expressed in *E. coli* BL21 Star™ (Life technologies) cells in auto-inducing medium[57], overnight at 20°C. GST-tagged constructs were affinity-purified on glutathione resin in lysis buffer (20 mM Tris-HCl pH 7.5 at 4°C, 1.2 M NaCl, 10% glycerol) containing 1 mM DTT. After washing 3 times in buffer A (20 mM Tris-HCl pH 7.5 at 4°C, 300 mM NaCl, 10% glycerol, 1 mM DTT), the recombinant protein was eluted from the resin by incubating with TEV protease overnight at 4°C, in buffer A. His-tagged Exu wt was purified by cobalt affinity chromatography in lysis buffer containing 20 mM imidazole and 1 mM beta-mercaptoethanol (β-me), and eluted with a gradient to 250 mM imidazole. The His-tag was subsequently removed by dialysis in the presence of TEV protease overnight at 4°C, in buffer A. In both cases, the protein was applied to a heparin resin, eluted with a gradient to 1M NaCl, and further purified by SEC on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare), in buffer A. Mutants and truncations were generated by site-directed mutagenesis from the GST-tagged construct, and purified as described above.

To produce Exu^{333} selenomethionine (SeMet) labeled protein, the construct was expressed in an *E. coli* methionine auxotroph strain, DL41, in M9 medium and starved before addition of SeMet. The protein was purified as described above for the unmodified protein.

Crystallization, data collection and analysis

Crystals of Exu^{333} were obtained in 6% MPD, 0.05 M MES, pH 6.5 and grown at 4°C by vapor diffusion. For data collection, crystals were cryo-protected with mother liquor supplemented with gradual addition of glycerol up to 30-40% and flash frozen in liquid nitrogen. The crystals diffracted to 2.37 Å resolution, belong to the space group $P6_122$ with cell dimensions of a=b=79.9 Å, c=273.3 Å, α=β=90°, γ=120° and contain one molecule in the ASU. Data were processed and scaled using XDS[58]. The structure was solved by SAD phasing of the SeMet substituted Exu^{333}. 7 Selenium sites and initial phases were obtained with SHELX[59]. Refinement was carried out using iterative cycles of model building in COOT[60] and restrained refinement in PHENIX[61].

Crystals of Exu^{406} grew in 250 mM sodium malonate, 18% PEG 1000, 50 mM MES, pH 6 at 4°C by vapor diffusion and were snap frozen as above. The crystals diffracted to 2.80 Å and belonged to space group $P2_1$ with cell dimensions of a=79.5 Å, b=66.6 Å, c=81.3 Å and α=β=90°, γ=113.6°. The structure was solved by MR using PHASER[62] and the Exu^{333} structure as a search model. Iterative cycles of model building and restrained refinement were carried out in COOT and PHENIX[60,61]. For both native datasets, the resolution cutoff was chosen to at least include I/σ>2.

All diffraction data were collected at the PXII beamline of the Swiss Light Source (Villigen, Switzerland). 100% of the residues in each structure fall within the allowed regions of the Ramachandran plot.

We refer to the higher resolution structure of Exu^{333} for molecular details and to the structure of Exu^{406} for overviews. Unless stated otherwise, we refer to molecule A throughout the manuscript. Residues are numbered according to Exu wt sequence.

Laser light scattering

For MALLS, purified Exu wt and two dimerization mutants (M68A Y70A and R92A) were loaded onto a Superdex 200 10/300 GL column (GE Healthcare) connected to a miniDAWN TREOS MALLS detector and Optilab rEX differential refractometer (Wyatt Technologies). Each run was performed with 0.5 mg of protein, in buffer A, at 12°C. Molecular weight calculations were performed using ASTRA software (Wyatt Technologies).

For Right-Angle Light Scattering (RALS), samples were loaded onto a Superdex 200 5/150 column (GE Healthcare) connected to a TDA302 detector array (Viscotek), in buffer H (20 mM Hepes pH 7.5, 300 mM NaCl and 10% glycerol), at room temperature (RT, 21-23ºC). Runs were performed with 10 mg/ml of purified Exu wt alone, or 1 mg/ml of Exu wt pre-incubated for 10' at RT with 2.5 molar excess of 5'-6-carboxy-fluorescein (6-FAM)-labeled RNA. Sample volume was 20 µl. Data were analyzed using OmniSEC 4.5 software.

Analytical size exclusion chromatography

Exu (12.5 nmol) was incubated alone or with a 1.2 molar excess of RNA in buffer A, overnight at 4°C. The samples were analyzed on a Superdex 200 10/300 GL column (GE Healthcare), in buffer A, monitoring UV absorbance at 280 and 260 nm.

Fluorescence anisotropy

FA measurements were performed with 5'-6-FAM-labeled RNA at RT on an Infinite F200 plate reader (Tecan). The RNA, at a concentration of 10 nM, was incubated with different concentrations of purified Exu wt or mutants, in buffer H. The final reaction volume was 50 µl. Each titration point was measured three times, with an integration time of 40 µs, using 485 nm and 535 nm as excitation and emission wavelength, respectively. The data were analyzed by nonlinear regression fitting using the Prism 6 software (GraphPad).

Nucleic acid sequences are as follows (5' to 3'): AGGCAGUUUCUGGUACUCAG (50% GC RNA; 50% GC DNA has the same sequence; perfectly complementary RNA or DNA unlabeled oligonucleotides were used to generate dsRNA or dsDNA, respectively); CCCAAAAUGAAAAAUGUUUCUCUUGGGCGUAAUCUCAUACAAUGAUU ACCCUUAAAGAUCGAACAUUUAAACAAUAAUAUUUGGG (*bcd-*Vb); CGCUUGAUUGUAUUUUUAAAUUAAUUCUUAAAAACUACAAAUUAAGC (*K10* TLS).

Circular Dichroism (CD)

CD spectra were recorded on a Jasco J-810 spectropolarimeter, at 20ºC. Measurements were performed using 200 µl of recombinantly purified Exu (wt or mutants), at a concentration of 0.2 mg/ml, in buffer A. Each spectrum was recorded 5 times, and averaged.

Mass spectrometry

UV-induced protein-RNA cross-linking and enrichment of cross-linked peptides. UV-cross-linking and enrichment of cross-linked peptides was performed according to established protocols[38]. Briefly, Exu was incubated with a 1.2 molar excess of $(U)_{20}$ RNA in buffer A' (20 mM Tris-HCl pH 7.5 at 4°C, 300 mM NaCl), overnight at 4°C, and analyzed on a HiLoad 16/600

Superdex 200 pg column (GE Healthcare), in buffer A'. The peak corresponding to the Exu:RNA complex was collected and the concentration adjusted to 8.6 μM. The sample (100 μl) was transferred to black polypropylene microplates (Greiner Bio-One) and irradiated at 254 nm for 10 min; an equal volume of sample was kept as non-irradiated control. After ethanol precipitation, the samples were denatured in 4 M urea and 50 mM Tris-HCl, pH 7.9, diluted to 1 M urea in the same buffer and digested for 2 h at 52°C with 5 μg RNase A (Ambion, Applied Biosystems). After RNA digestion, proteolysis with trypsin (Promega) was performed overnight at 37°C with an enzyme to protein ratio of 1:20. The sample was desalted on an in-house–prepared C18 (Dr. Maisch GmbH) column, and the cross-linked peptides were enriched on an in-house–prepared TiO_2 (GL Sciences) column as previously described[38]. The samples were dried and then resuspended in 10 μl sample solvent (5% v/v ACN and 1% v/v Formic Acid) for MS analysis.

Nano–liquid chromatography and MS analysis. 8 μL of the above sample was injected onto a nano–liquid chromatography system (EASY nLC II, Thermo Scientific) including a C18 trapping column of length ~4 cm and inner diameter 100 μm, in line with a C18 analytical column of length ~10 cm and inner diameter 50 μm (both packed in house; C18 AQ 120 Å 5 μm trapping column or 3 μm analytical column). Analytes were loaded on the trapping column at a maximum pressure of 300 bar in buffer AF (0.1% v/v Formic Acid), washed with 25 μl buffer and subsequently eluted and separated on the analytical column with a gradient of 5–35% buffer BF (95% v/v acetonitrile and 0.1% v/v Formic Acid) with an elution time of 37 min (0.77%/min) and a flow rate of 300 nL/min. Online ESI-MS was performed with an Q Exactive instrument (Thermo Scientific), operated in data-dependent mode with a TOP10 method. MS scans were recorded in the m/z range of 350–1,600 at a resolution of 70,000 FWHM and for subsequent MS/MS the top ten most-intense ions were selected. Both precursor ions as well as fragment ions were scanned in the Orbitrap. Fragment ions were generated by higher-energy collision

dissociation (HCD) activation (normalized collision energy = 25) and recorded from m/z = 100 with a resolution of 7500 FWHM.

Database search with cross-linking MS. Data analysis was performed as previously described[38]. Briefly, raw data was converted into the open mzML format[63] with msconvert (part of ProteoWizard[64]) and processed with OpenMS[65,66] with OMSSA[67] as database search engine. The data from the cross-linking experiment and the non-irradiated control was centroided and aligned to correct for retention time shifts. A first search against a target-decoy database containing the sequences of full-length Exu lacking loop1 (as distributed with MaxQuant[68]) identified confident peptide-spectrum-matches (FDR 1%) that were excluded from further analysis. Next, fragment spectra were filtered from the cross-linking data if the precursor was observed in the non-irradiated control at comparable intensity (fold change < 2) or if it originated from a small RNA oligonucleotide (M < 1750 Da and fractional mass < 0.2). For all remaining precursors, variants were generated by subtracting calculated RNA masses of one to four uridines with several RNA modifications ($-H_2O$, $-HPO_3$, $-H_3PO_4$, $-H_2O + 152$, $-HPO_3 + 152$, $-H_3PO_4 + 152$.). The fragment spectra were searched with the original precursor and all its variants against a database containing full-length Exu wt and LDADS (see **Supplementary Fig. 1**). All reported cross-links were validated by close manual inspection of the fragment spectra.

Fly stocks

The following fly stocks were used: OregonR (wild-type), *Df(2R)exu1, cn^1 bw^1 sp^1 / CyO, P{hs-hid}4 / exu^{VL}, cn^1 bw^1 / CyO, P{hs-hid}4 / exu^{QR}, cn^1 bw^1 / CyO, P{hs-hid}4*[47,69]. For the transgenes a genomic fragment containing the *exu* gene[25] was modified to yield N-terminal Venus-StrepTagII fusions of wt, mutant or truncated Exu proteins. The constructs were cloned into the pUAST-attB vector via BamHI, thereby deleting the UAS-sites and the SV40 poly(A) signal. The

purified vectors were injected into embryos from a recombinant stock with the genotype: y^1 *M{vas-int.Dm}ZH-2A w*; M{3xP3-RFP.attP}ZH-22A* cn^1 *Df(2R)exu1* bw^1 sp^1 */ SM6a* and transgenic flies were identified in the F1 generation by the presence of orange eyes. In the following generation the *CyO, P{hs-hid}4* balancer chromosome was introduced into the stocks. Ovaries and embryos used for *in situ* hybridization, Western Blots and imaging were obtained from these transgenic flies crossed to exu^{VL}, cn^1 bw^1 */ CyO, P{hs-hid}4* flies, and selected against the presence of the balancer chromosome.

In situ hybridization and antibody staining

Ovary dissection was performed as described previously[70]. Ovaries were fixed in PBS containing 0.2% Tween 20, 4% paraformaldehyde, and stained with rhodamine-phalloidin (Molecular Probes R415, 1:1000) for 20' at RT. Samples were mounted in Fluoromount-G (SouthernBiotech), and imaged on an Olympus FluoView1000 confocal microscope using a UPLSAPO 60x oil objective (NA 1.35). Images were processed with ImageJ[71].

Early embryos were collected according to standard protocols[72]. For *in situ* hybridizations, full-length *bcd* and *osk* anti-sense RNA probes were labeled with Digoxigenin-UTP (Roche) and detected with Alkaline Phosphatase-conjugated anti-digoxigenin antibody (Roche 11093274910, 1:2000). Images were acquired on a Zeiss AxioImager Z.1 microscope, using a Plan-Apochromat 10x objective (NA 0.45) and a differential interference contrast (DIC) condenser.

The following antibodies were used for Western Blots: anti-GFP polyclonal antibody (Life Technologies A11122, 1:2000), anti-α-tubulin monoclonal antibody (Sigma T6074, 1:5000), anti-Exu serum (1:1000). The anti-Exu serum was developed in house in rabbits using purified Exu wt (see **Fig. 7j** for validation). Exu^{410} and Exu^{333} are poorly recognized by our anti-Exu polyclonal serum suggesting that the CTR of Exu is more immunogenic than the folded domains

of the protein. The C-terminally truncated constructs migrate as a single band, consistently with the absence of phosphorylation sites.

56. Diebold, M.-L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *Journal of structural biology* **175**, 178-188 (2011).
57. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**, 207-234 (2005).
58. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800 (1993).
59. Schneider, T.R. & Sheldrick, G.M. Substructure solution with SHELXD. *Acta crystallographica Section D, Biological crystallography* **58**, 1772-1779 (2002).
60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography* **60**, 2126-2132 (2004).
61. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography* **66**, 213-221 (2010).
62. McCoy, A.J. et al. Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674 (2007).
63. Martens, L. et al. mzML--a community standard for mass spectrometry data. *Molecular & cellular proteomics* **10**, R110.000133 (2011).
64. Chambers, M.C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature biotechnology* **30**, 918-920 (2012).
65. Kohlbacher, O. et al. TOPP--the OpenMS proteomics pipeline. *Bioinformatics* **23**, e191-7 (2007).
66. Sturm, M. et al. OpenMS - an open-source software framework for mass spectrometry. *BMC bioinformatics* **9**, 163 (2008).
67. Geer, L.Y. et al. Open mass spectrometry search algorithm. *Journal of proteome research* **3**, 958-964 (2004).
68. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372 (2008).
69. Schüpbach, T. & Wieschaus, E. Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. *Genetics* **121**, 101-117 (1989).
70. Palacios, I.M. & St Johnston, D. Kinesin light chain-independent function of the Kinesin heavy chain in cytoplasmic streaming and posterior localisation in the Drosophila oocyte. *Development (Cambridge, England)* **129**, 5473-5485 (2002).
71. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).
72. Harlow, E. & Lane, D. Preparing early whole-mount Drosophila embryos for immunostaining. *CSH protocols* pdb.prot4524-pdb.prot4524 (2006).

Figure 1

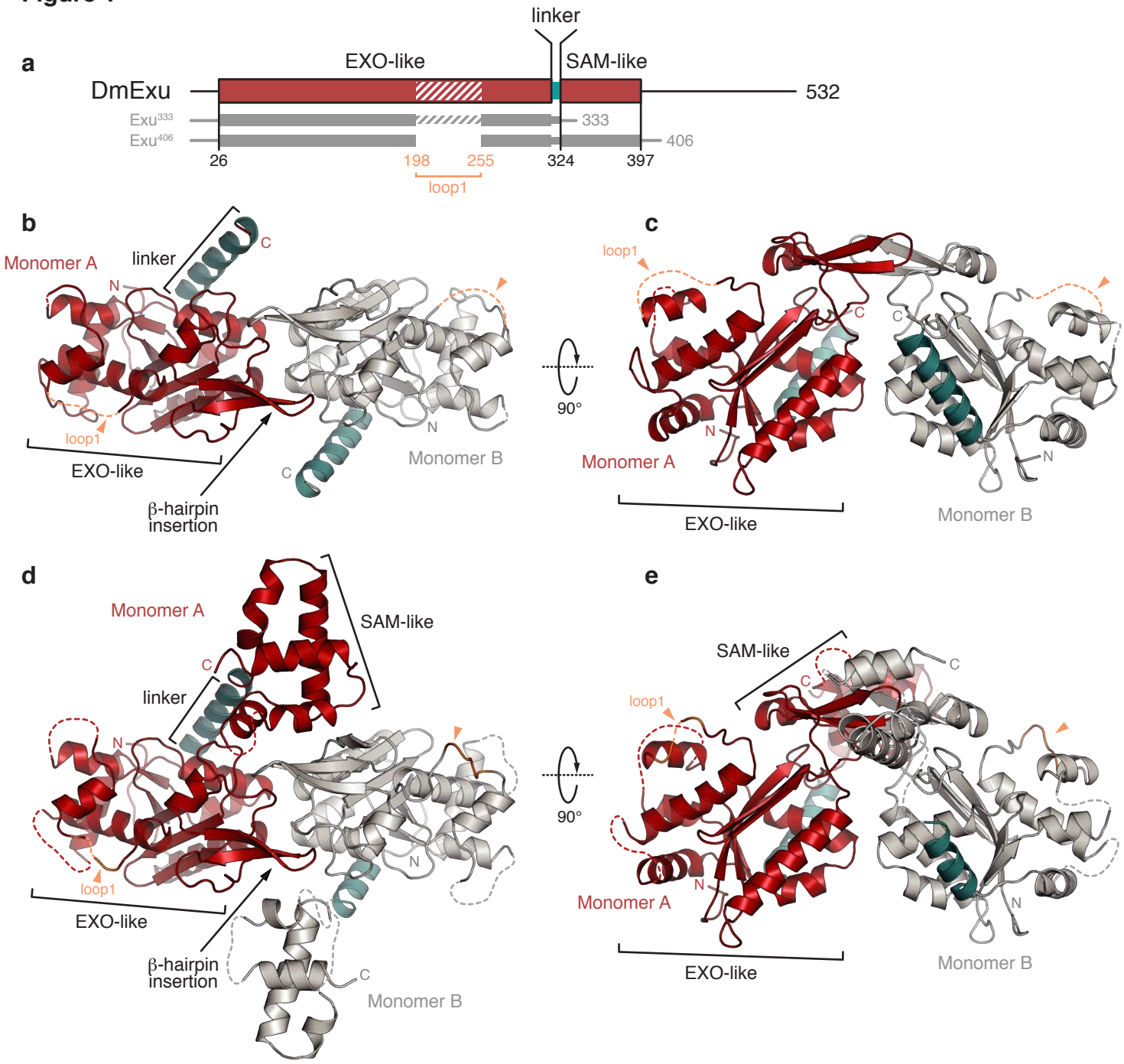

a
linker
EXO-like
SAM-like
DmExu
532
Exu333
333
Exu406
406
26
198
255
324
397
loop1
b
Monomer A
linker
C
N
loop1
EXO-like
β-hairpin
insertion
C
N
Monomer B
c
loop1
C
C
90°
N
Monomer A
EXO-like
N
Monomer B
d
Monomer A
SAM-like
C
linker
N
loop1
EXO-like
β-hairpin
insertion
N
C
Monomer B
e
SAM-like
C
C
loop1
90°
N
Monomer A
EXO-like
N
Monomer B

Figure 2

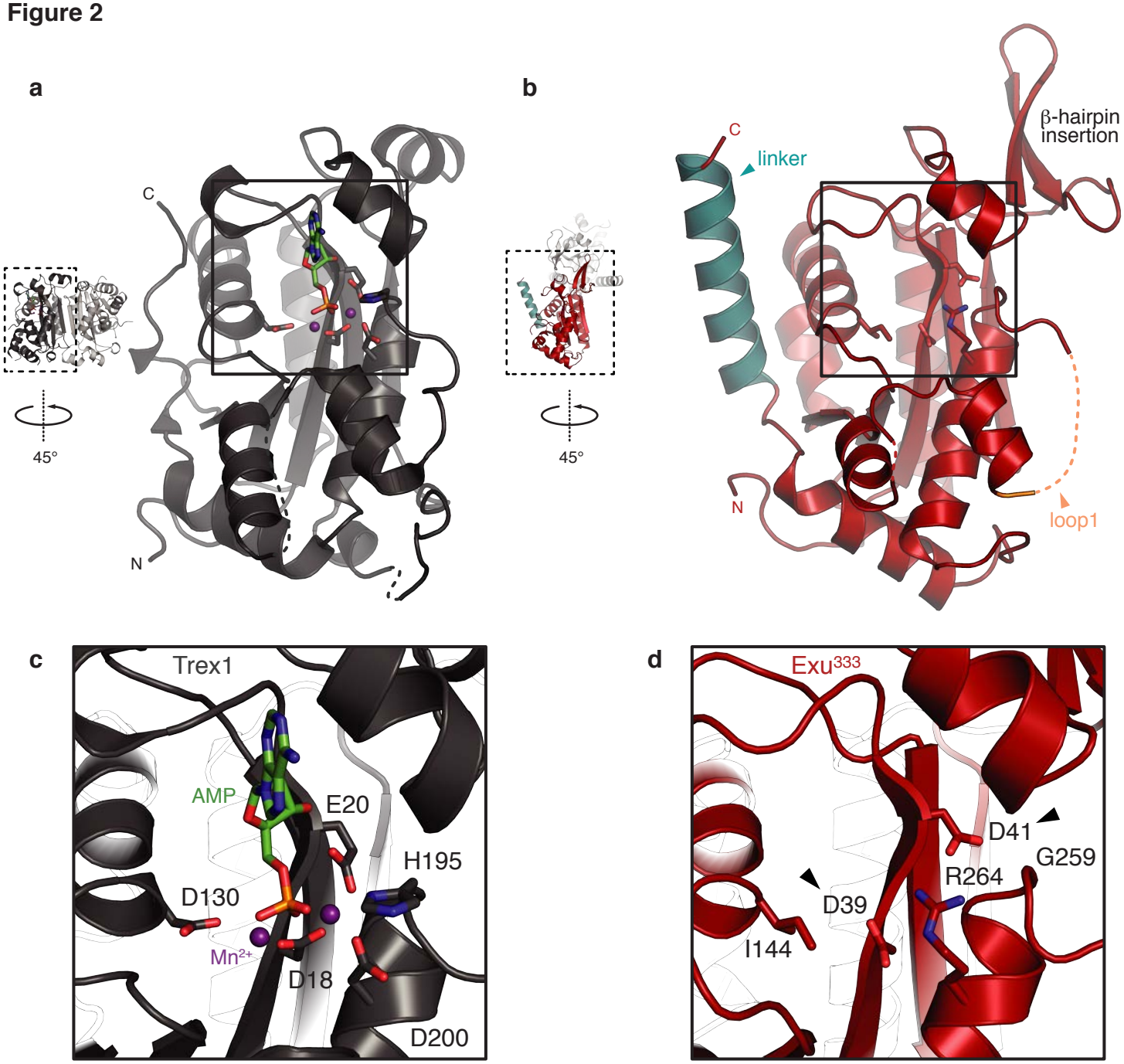
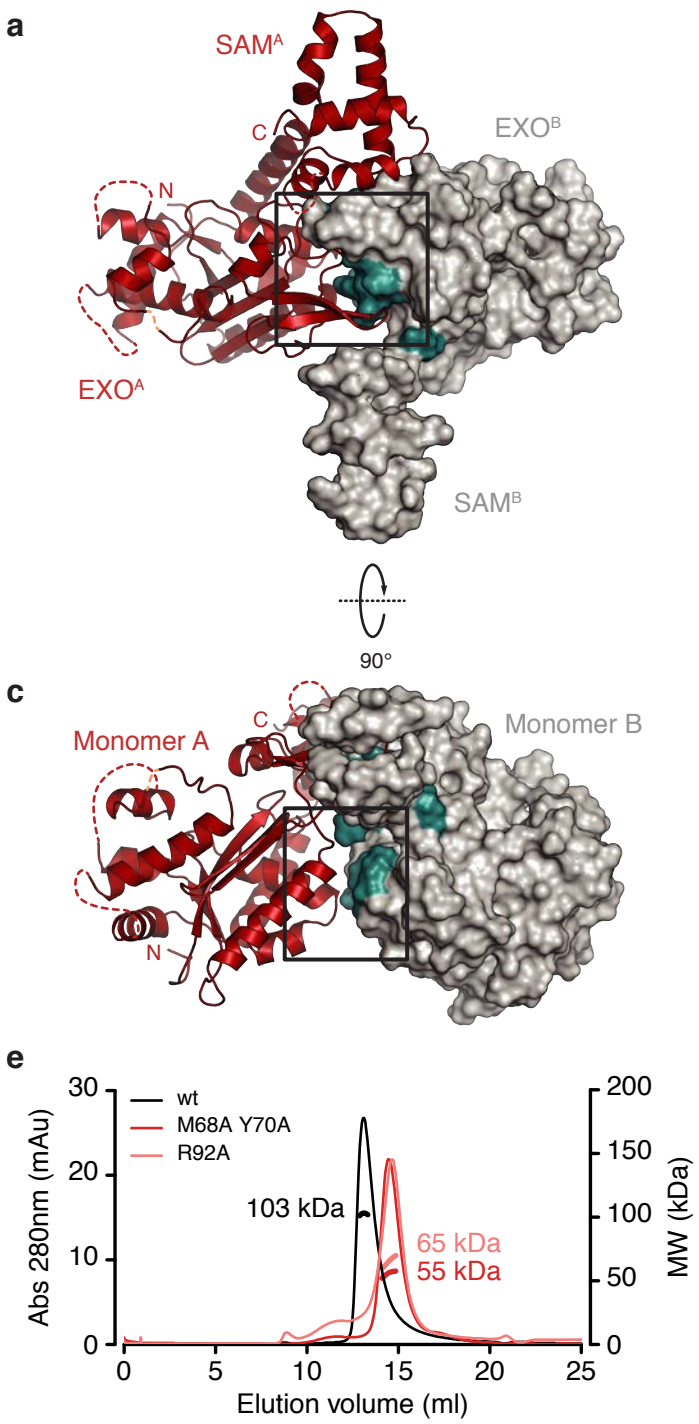

a
b
C
N
45°
45°
C
linker
β-hairpin
insertion
N
loop1
c
Trex1
AMP
E20
H195
D130
Mn2+
D18
D200
d
Exu333
D41
G259
R264
D39
I144

Figure 3


a
SAMA
C
EXOB
N
EXOA
SAMB
90°
c
Monomer A
C
Monomer B
N
e
wt
M68A Y70A
R92A
103 kDa
65 kDa
55 kDa
Abs 280nm (mAu)
MW (kDa)
Elution volume (ml)
0
5
10
15
20
25
30
20
10
0
200
150
100
50
0

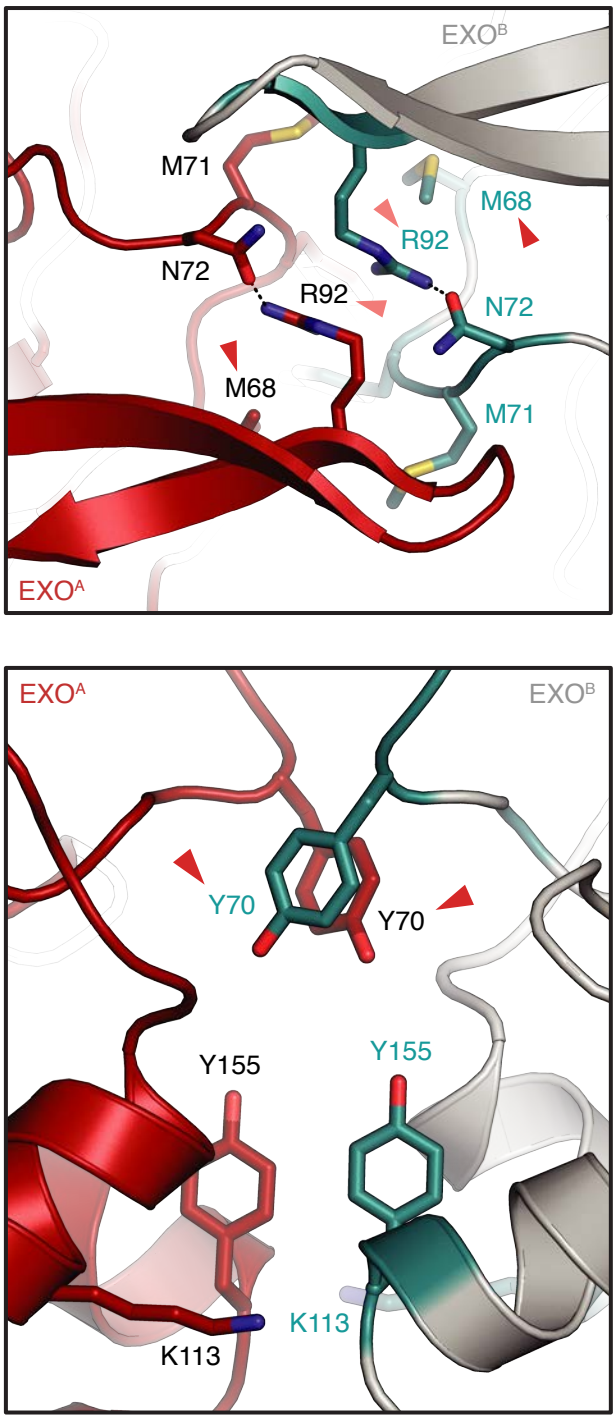

b
EXOB
M71
M68
R92
N72
R92
N72
M68
M71
EXOA
d
EXOA
EXOB
Y70
Y70
Y155
Y155
K113
K113

Figure 4

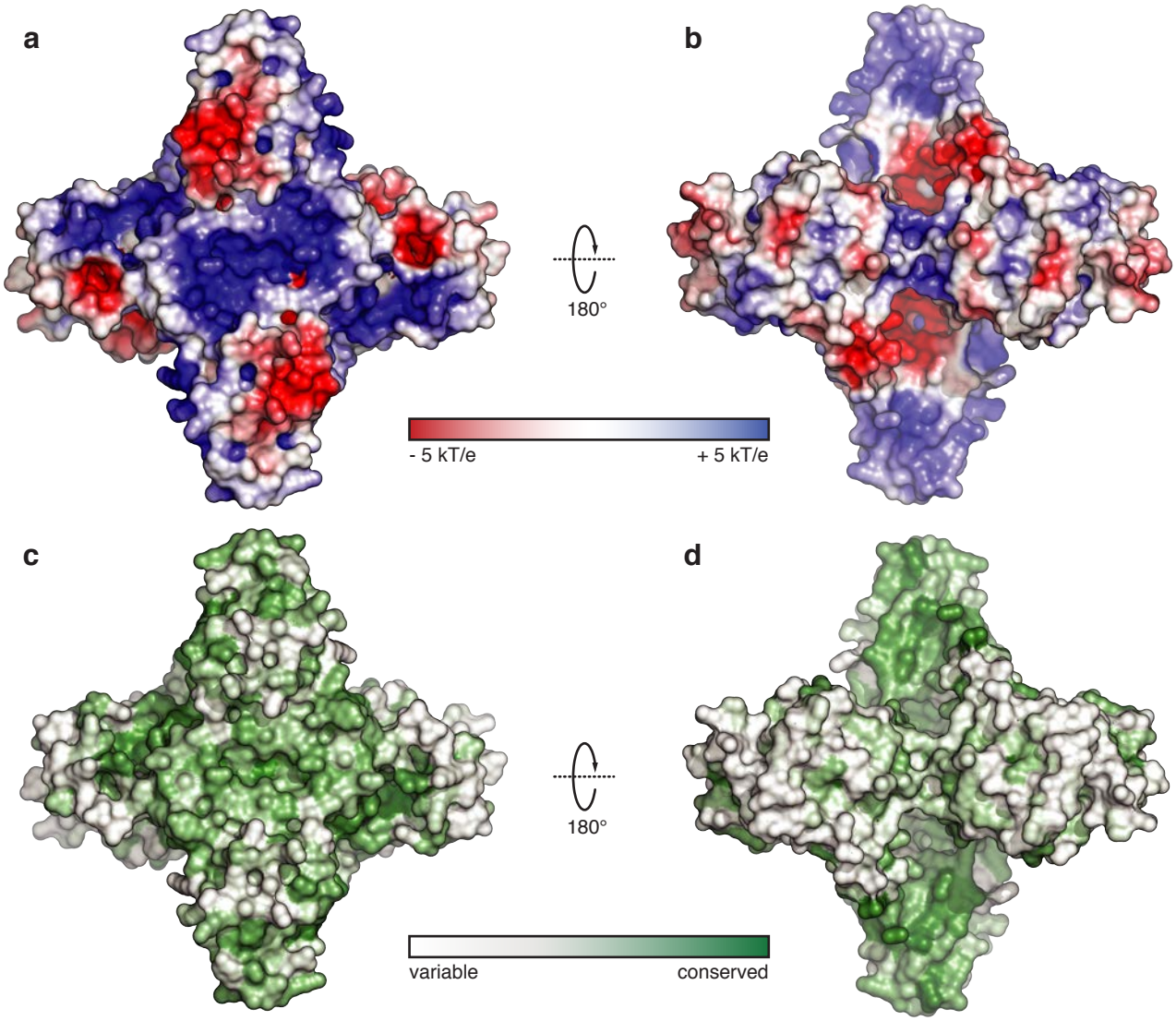


Figure 5

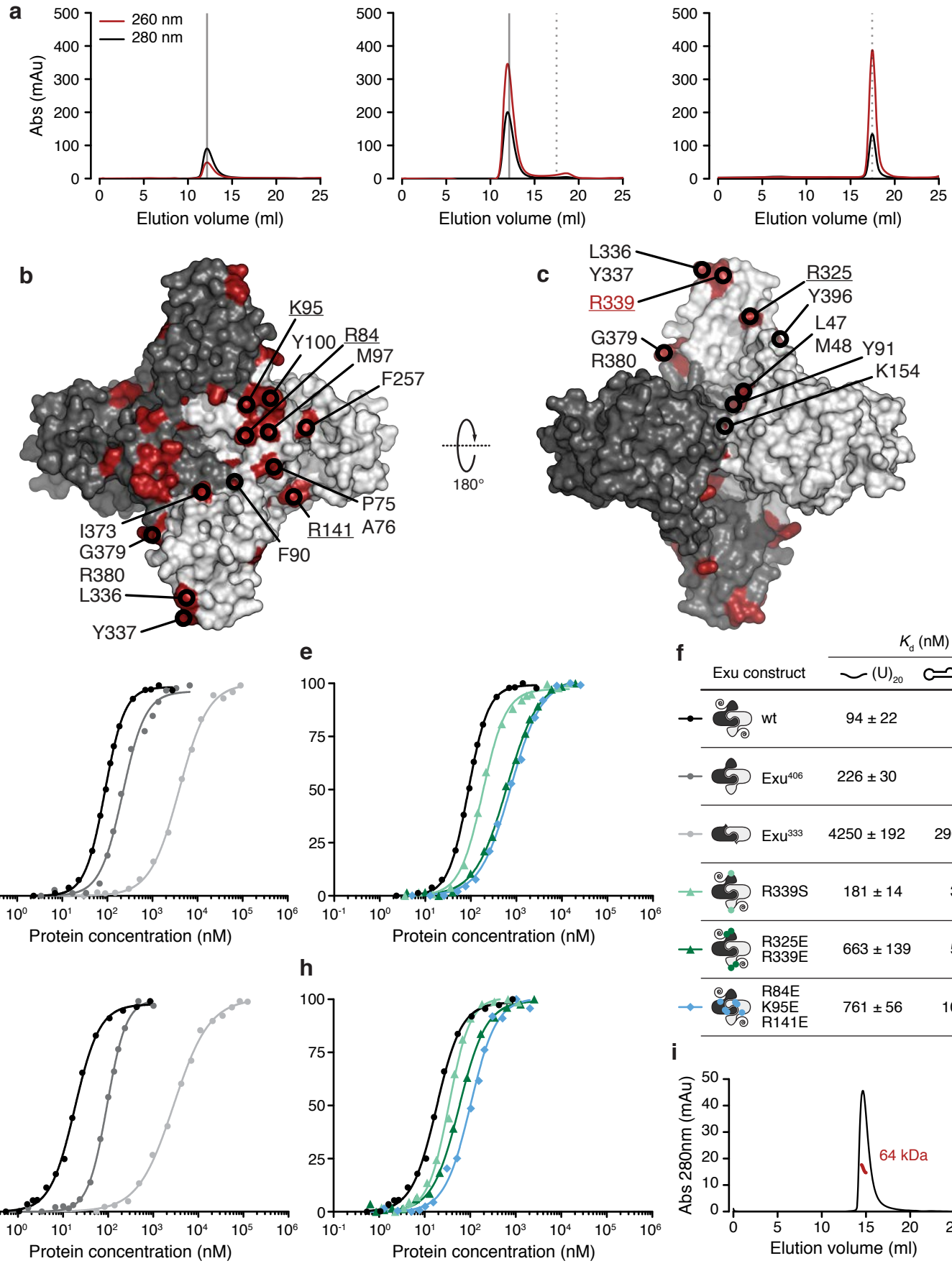
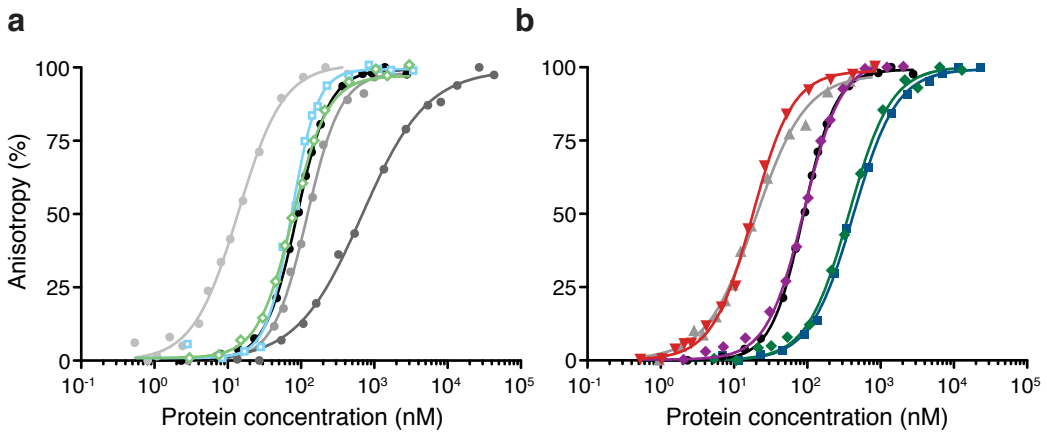
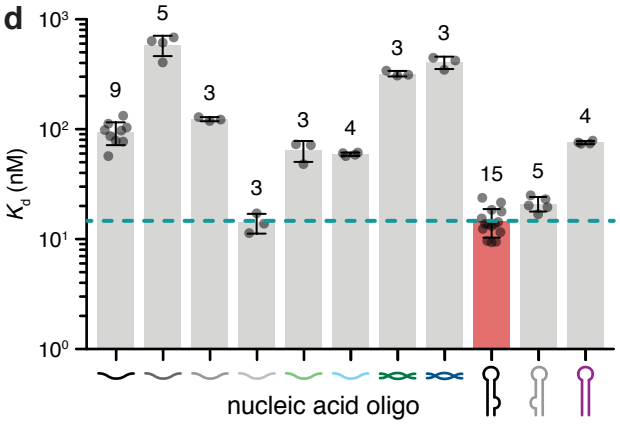

a
260 nm
280 nm
Abs (mAu)
Elution volume (ml)
b
K95
Y100
R84
M97
F257
P75
A76
R141
F90
I373
G379
R380
L336
Y337
180°
c
L336
Y337
R339
G379
R380
R325
Y396
L47
M48
Y91
K154
d
e
Anisotropy (%)
Protein concentration (nM)
f
Exu construct
K_d (nM)
$(U)_{20}$
bcd-Vb
wt
94 ± 22
15 ± 4
Exu406
226 ± 30
94 ± 2
Exu333
4250 ± 192
2904 ± 328
R339S
181 ± 14
32 ± 3
R325E
R339E
663 ± 139
53 ± 12
R84E
K95E
R141E
761 ± 56
102 ± 10
g
h
i
Abs 280nm (mAu)
MW (kDa)
64 kDa

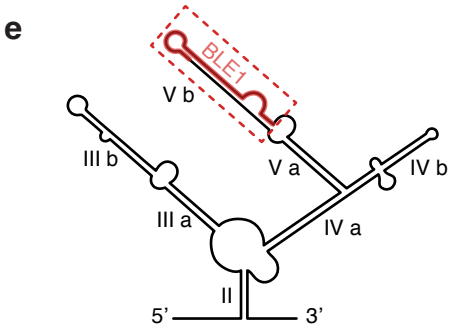
Figure 6

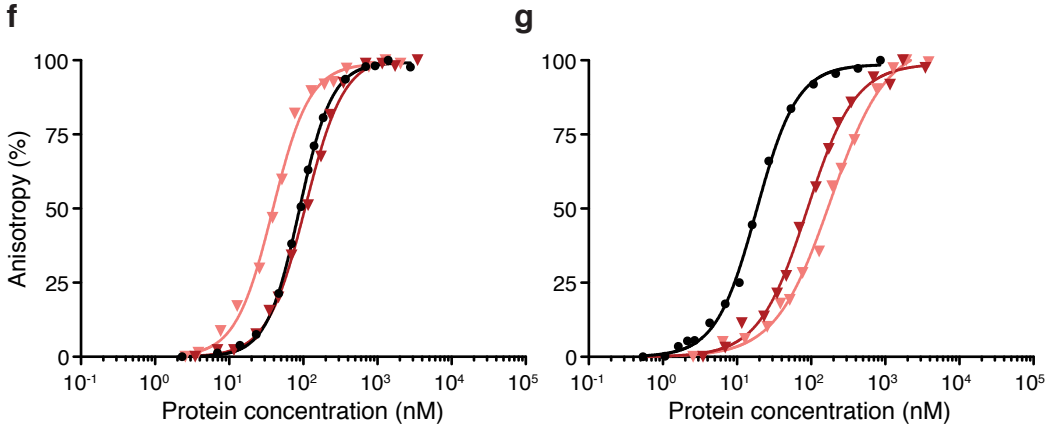




c

Exu wt +	K_d (nM)
$(U)_{20}$	94 ± 22
$(UA)_{10}$	598 ± 124
$(UC)_{10}$	125 ± 5
$(UG)_{10}$	14 ± 3
ssRNA (50% GC)	65 ± 14
ssDNA (50% GC)	60 ± 2
dsRNA (50% GC)	323 ± 19
dsDNA (50% GC)	409 ± 53
bcd-Vb	15 ± 4
bcd-Vb antisense	21 ± 3
K10 TLS	88 ± 3

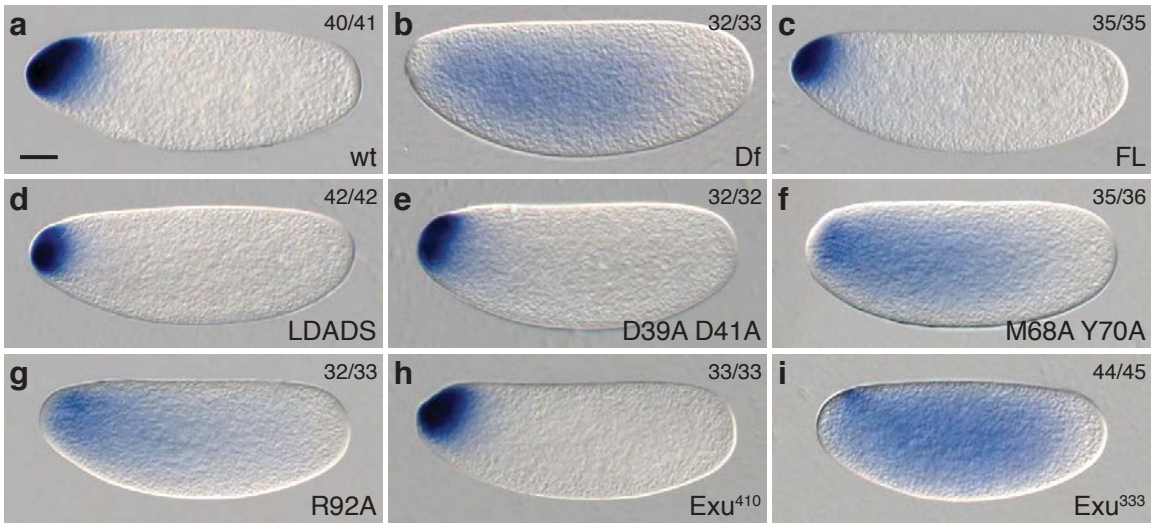


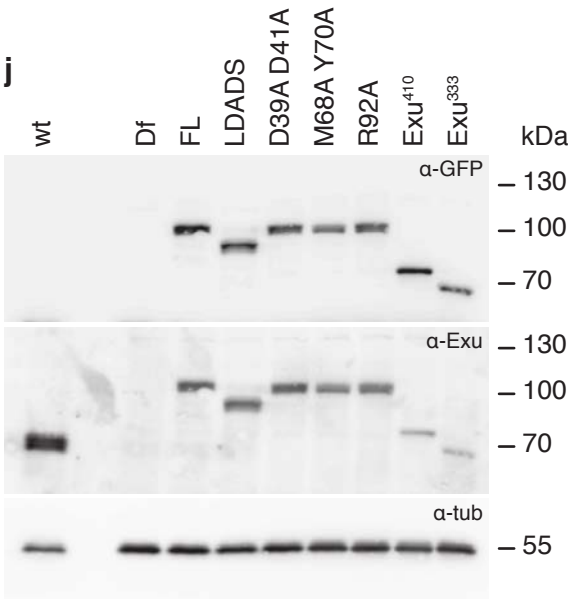


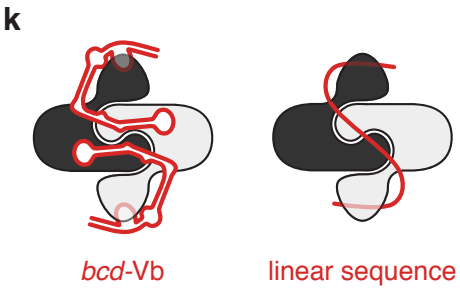
h

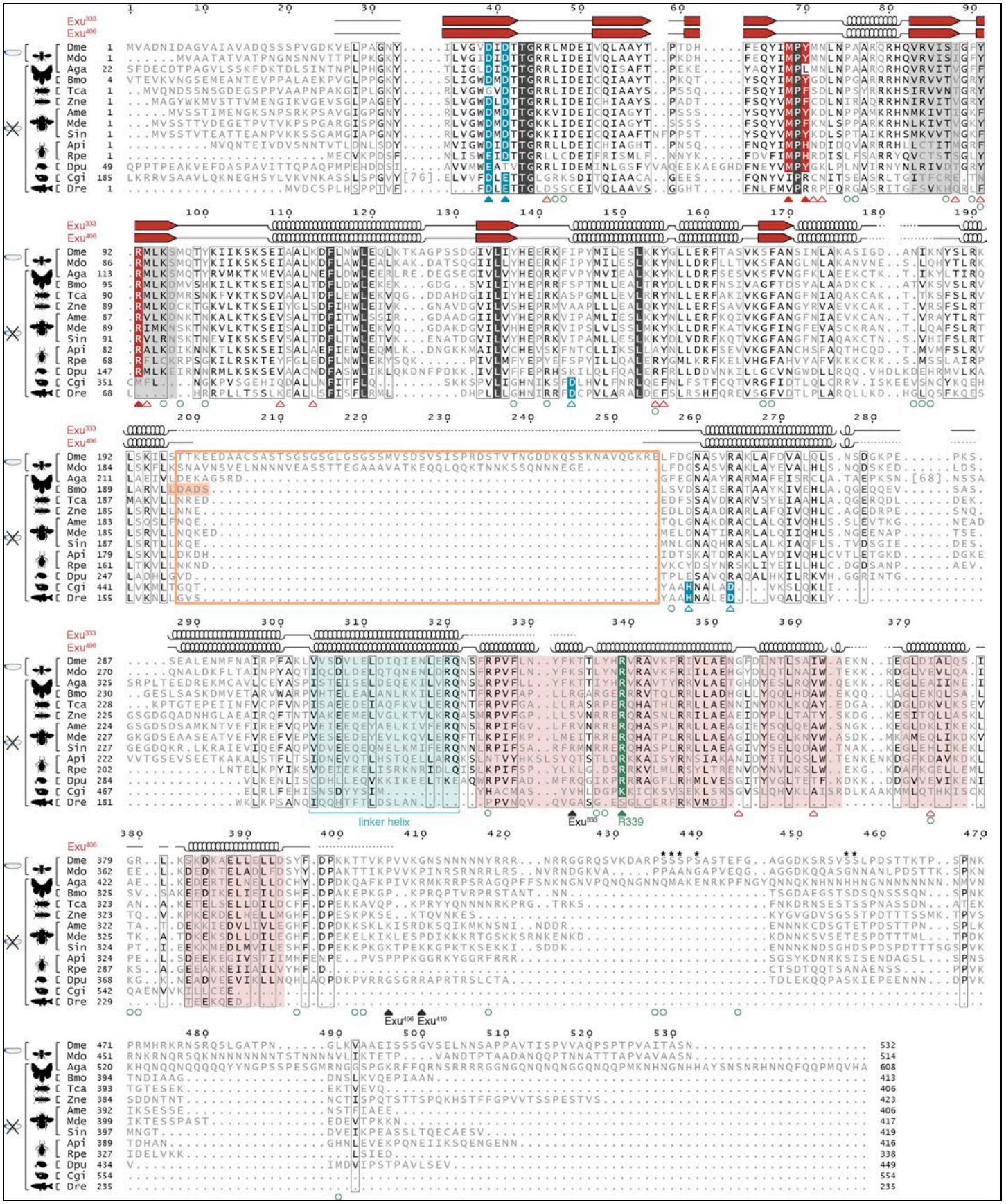
Exu mutant	K_d (nM) $(U)_{20}$	K_d (nM) *bcd*-Vb
wt	94 ± 22	15 ± 4
M68A Y70A	110 ± 6	81 ± 6
R92A	34 ± 4	177 ± 6

Figure 7


a
40/41
wt
b
32/33
Df
c
35/35
FL
d
42/42
LDADS
e
32/32
D39A D41A
f
35/36
M68A Y70A
g
32/33
R92A
h
33/33
Exu410
i
44/45
Exu333


j
wt
Df
FL
LDADS
D39A D41A
M68A Y70A
R92A
Exu410
Exu333
kDa
α-GFP
130
100
70
α-Exu
130
100
70
α-tub
55


k
bcd-Vb
linear sequence

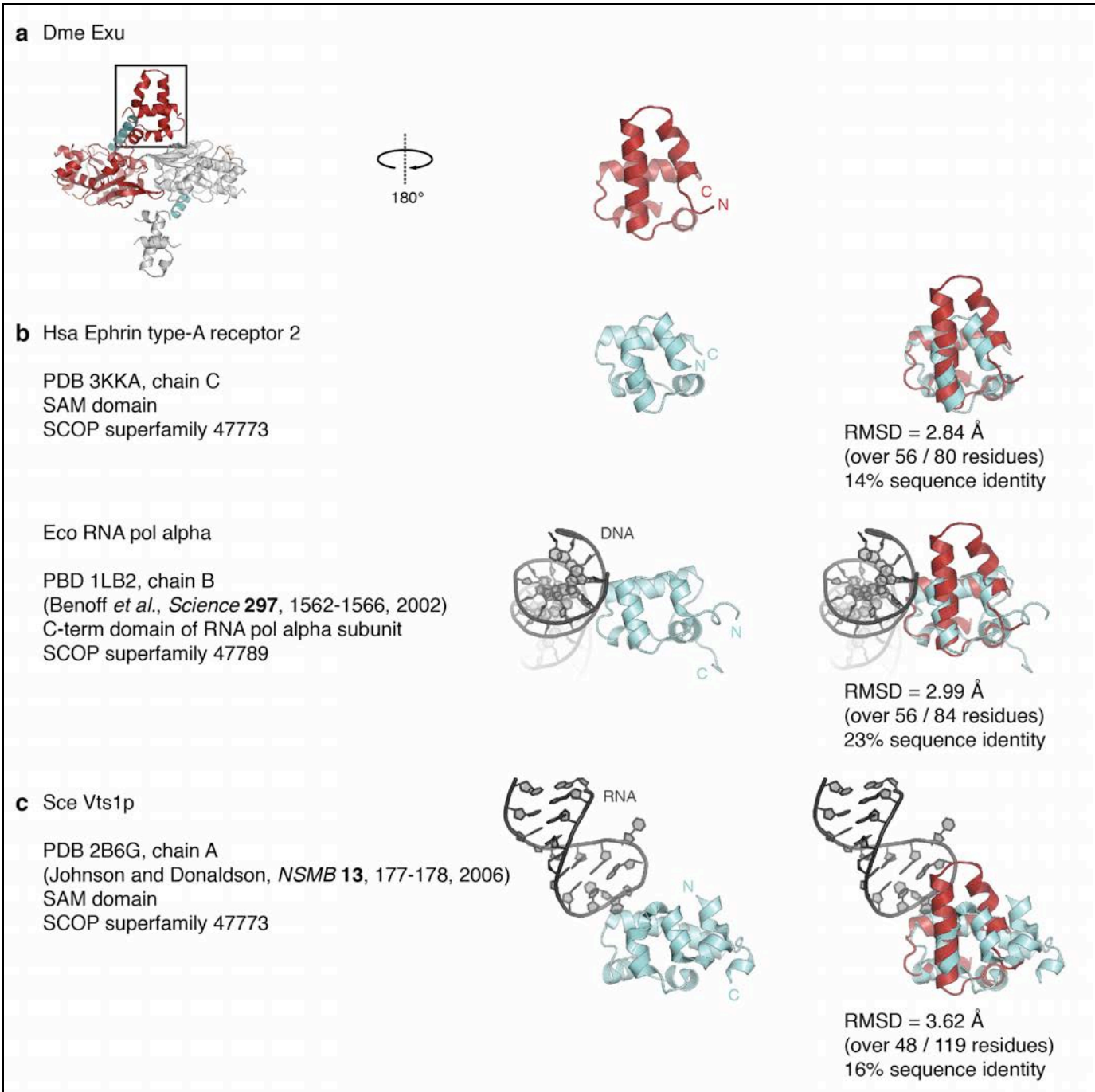

Exu333
Exu406
Dme
Mdo
Aga
Bmo
Tca
Zne
Ame
Mde
Sin
Api
Rpe
Dpu
Cgi
Dre
linker helix
Exu333
R339
Exu406
Exu410

Supplementary Figure 1

Alignment of Exu homologs.

Sequences used to generate the alignment include: [Insecta, order Diptera] *D. melanogaster* (Dme), *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. virilis*, *D. mojavensis*, *D. willistoni*, *D. pseudoobscura pseudoobscura*, *D. persimilis*, *D. grimshawi*, *D. miranda*, *D. affinis*, *B. cucurbitae*, *B. dorsalis*, *C. capitata*, *M. domestica* (Mdo), *A. gambiae* (Aga), *A. sinensis*, *A. darlingi*, *C. quinquefasciatus*, *A. aegypti*; [Lepidoptera] *B. mori* (Bmo), *D. plexippus*, *P. aegeria*; [Coleoptera] *D. ponderosae*, *T. castaneum* (Tca); [Phthiraptera] *P. humanus corporis*; [Isoptera] *Z. nevadensis* (Zne), *M. rotundata*; [Hymenoptera] *B. terrestris*, *B. impatiens*, *A. mellifera* (Ame), *A. florea*, *N. vitripennis*, *M. demolitor* (Mde), *S. invicta* (Sin), *C. biroi*, *C. floridanus*, *A. echinatior*, *H. saltator*; [Hemiptera] *A. pisum* (Api), *D. citri*, *R. pedestris* (Rpe); [Crustacea] *D. pulex* (Dpu); [Mollusca] *C. gigas* (Cgi); [Chordata] *D. rerio* (Dre). Numbering refers to the Dme sequence. Only the suborder Cyclorrapha (Diptera) has a *Bicoid* ( ) gene.

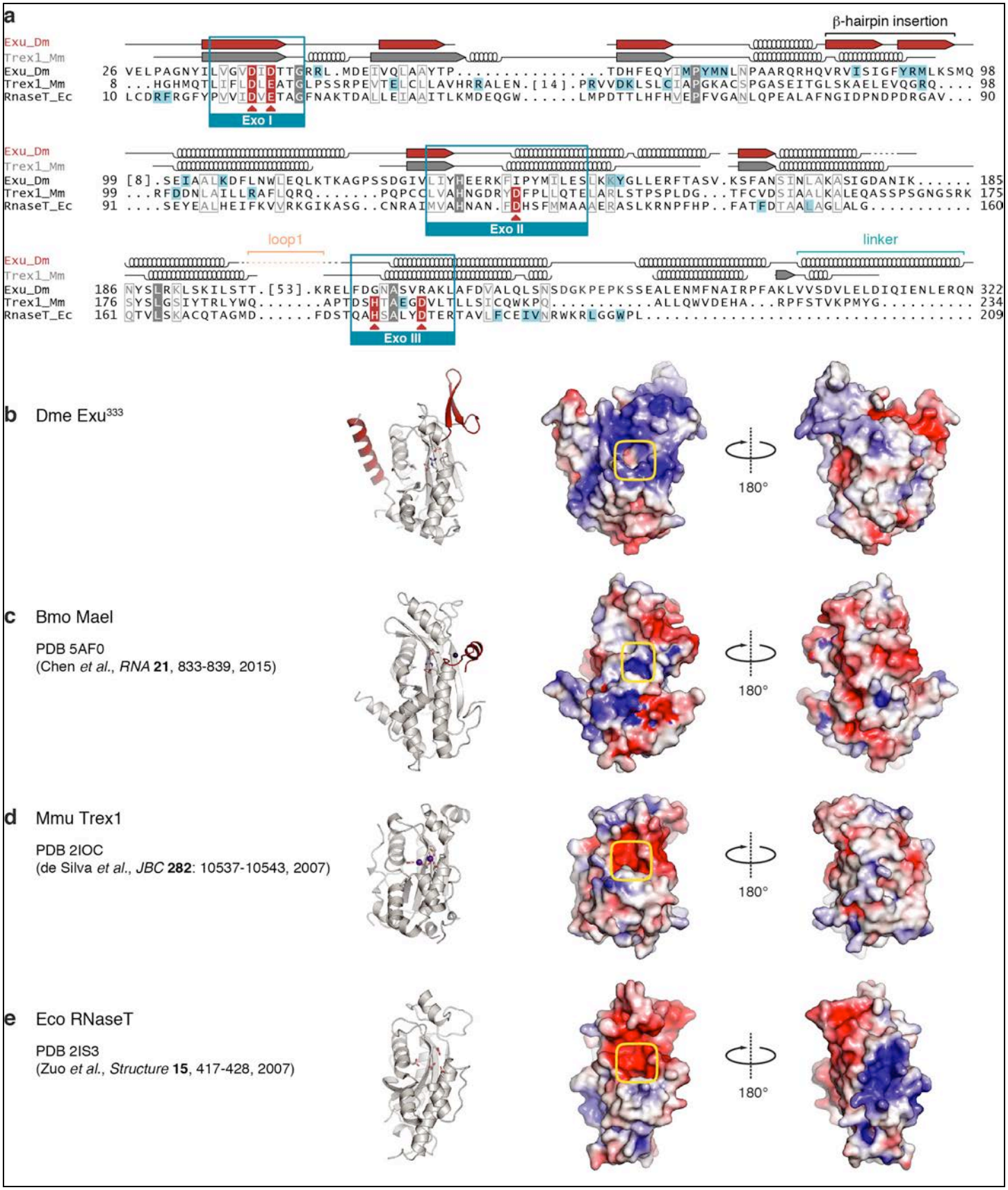
The secondary structure of Dme Exu^{333} and Exu^{406} is schematized above the alignment (β-sheets: ; α-helices: ; residues not ordered in the structure are shown as dotted lines; gaps indicate regions for which no structural information is available). Residues involved in Dme Exu dimerization: ; residues confirmed by mutagenesis: ; conserved residues are highlighted in red. Non conserved loop1 is framed by a orange box; the corresponding residues in Bmo Exu (used to replace the loop in the Exu406 construct used for crystallization), are highlighted in the same color. Catalytic residues in the exonuclease domain: ; mutated residues: ; conserved catalytic residues are highlighted in blue. Residues cross-linked to RNA: ; Arg339: . Black arrows indicate the boundaries of the Exu^{333}, Exu^{406} and Exu^{410} constructs. The phosphorylation sites identified in Dme Exu (Riechmann and Ephrussi, *Development*. **131**, 5897-5907, 2004) are marked by asterisks. The β-hairpin insertion is shaded in grey; the linker between the Exo-like and the SAM-like domains in light teal; the helices of the SAM-like domain in light red. The alignment was generated with MUSCLE (Edgar, *Nucleic Acids Res.* **32**, 1792-1797, 2004), visualized with ESPript (Robert and Gouet, *Nucleic Acids Res.* **42**, W320-W324, 2014) and edited in Adobe Illustrator.



Supplementary Figure 2

Structural homologs of Exu SAM-like domain.

a) Cartoon representation of Exu[406] (left) and of Exu SAM-like domain structure (aa 321-397; right). **b**) A structural similarity search in PDBeFold (Krissinel and Henrick, *Acta Cryst.* **D60**, 2256-2268, 2004) with Exu SAM-like domain outputs the sterile alpha motif (SAM), with Z-score = 4.4, and the C-terminal domain of RNA polimerase alpha, with Z-score = 4.2. **c**) Yeast Vts1p is a homolog of the *Drosophila* protein Smaug, and also contains a SAM domain which is involved in RNA binding (see also **Supplementary Fig. 5**). All structures in **b**) and **c**) belong to the SAM domain-like fold (SCOP 47768). Structural alignment was done with cealign in Pymol (v 1.7.6.0).

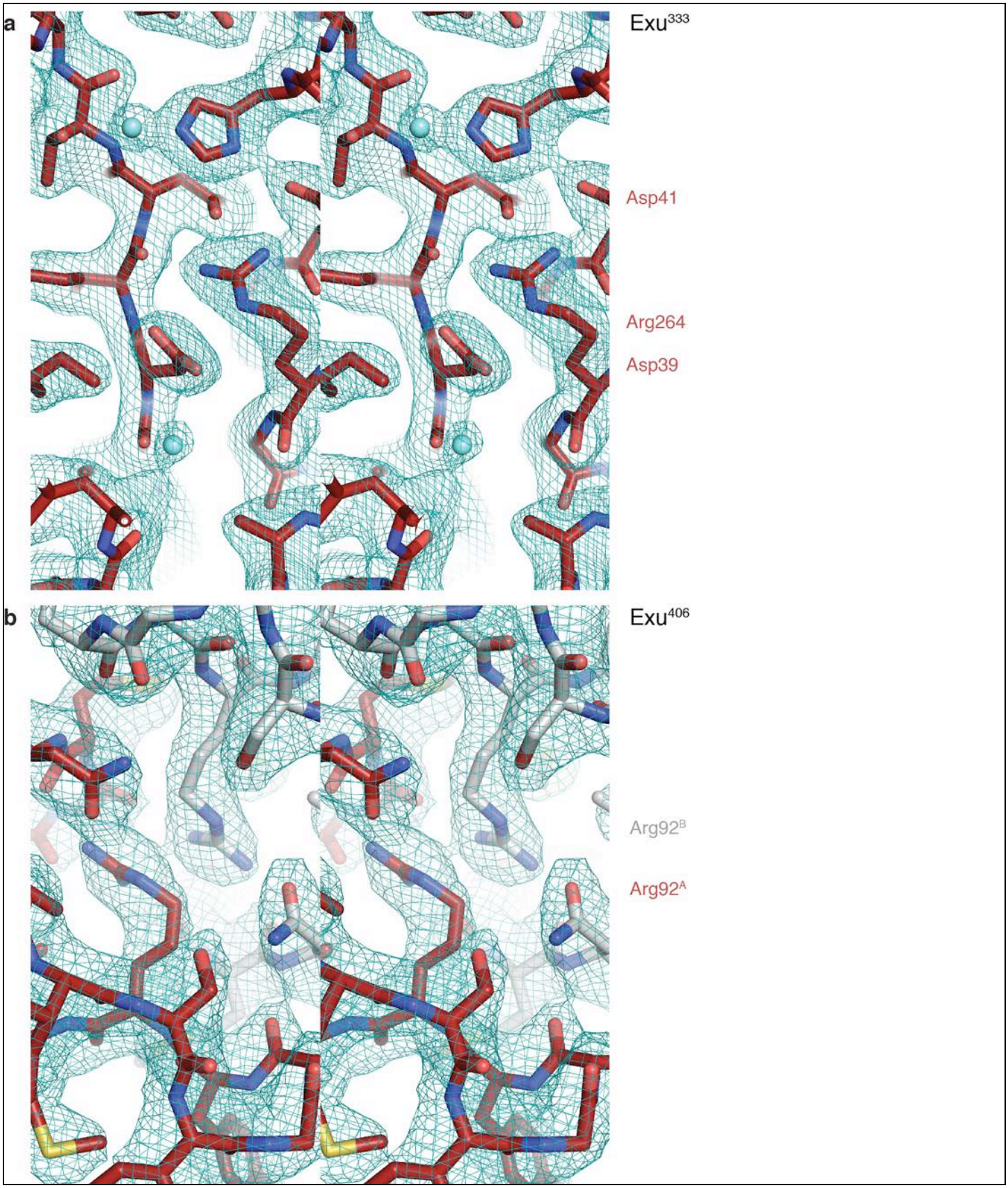

a
β-hairpin insertion
Exu_Dm
Trex1_Mm
Exu_Dm 26 VELPAGNYILVGVDIDTTGRRL.MDEIVQLAAYTP.....TDHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQ 98
Trex1_Mm 8 ...HGHMQTLIFLDLEATGLPSSRPEVTELCLLAVHRRALEN.[14].PRVVDKLSLCIAPGKACSPGASEITGLSKAELEVQGRQ.... 98
RnaseT_Ec 10 LCDRFRGFYPVVIDVETAGFNAKTDALLEIAAITLKMDEQGW.....LMPDTTLHFHVEPFVGANLQPEALAFNGIDPNDPDRGAV... 90
Exo I
Exu_Dm 99 [8].SEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASV.KSFANSINLAKASIGDANIK..... 185
Trex1_Mm 99 ...RFDDNLAILLRAFLQRQ.....PQPCCLVAHNGDRYDFPLLQTELARLSTPSPLDG...TFCVDSIAALKALEQASSPSGNGSRK 175
RnaseT_Ec 91 ....SEYEALHEIFKVVRKGIKASG..CNRAIMVAHNAN.FDHSFMMAAAERASLKRNPFHP..FATFDTAALAGLALG..... 160
Exo II
loop1
linker
Exu_Dm 186 NYSLRKLSKILSTT.[53].KRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFAKLVVSDVLELDIQIENLERQN 322
Trex1_Mm 176 SYSLGSIYTRLYWQ.....APTDSHTAEGDVLTLLSICQWKPQ.....ALLQWVDEHA...RPFSTVKPMYG..... 234
RnaseT_Ec 161 QTVLSKACQTAGMD.....FDSTQAHSALYDTERTAVLFCEIVNRWKRLGGWPL..... 209
Exo III
b Dme Exu333
180°
c Bmo Mael
PDB 5AF0
(Chen et al., RNA 21, 833-839, 2015)
180°
d Mmu Trex1
PDB 2IOC
(de Silva et al., JBC 282: 10537-10543, 2007)
180°
e Eco RNaseT
PDB 2IS3
(Zuo et al., Structure 15, 417-428, 2007)
180°

Supplementary Figure 3

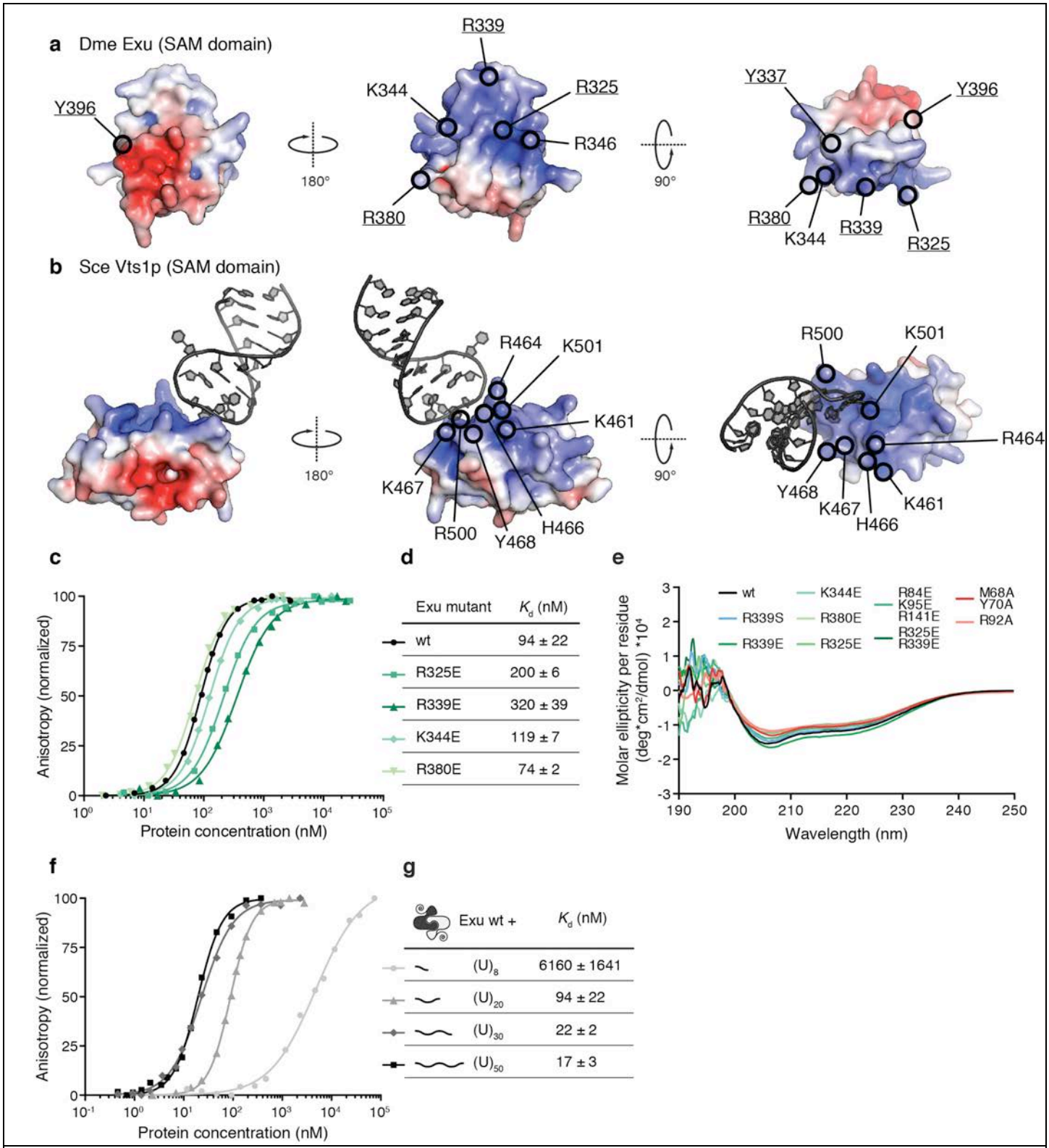
Comparison of DEDD exo- and pseudo-nucleases.

a) Structure-based sequence alignment of the EXO and EXO-like domains of *Drosophila* Exu (Exu_Dm) and of structurally similar 3'-5' DEDD exonucleases, mouse Trex1 (Trex1_Mm) and *E. coli* RNaseT (RNaseT_Ec). Secondary structure elements are shown above the sequences, in red for Exu and in dark gray for Trex1. Conserved residues are highlighted in dark gray. Blue boxes indicate EXO signature motifs, while signature catalytic residues are marked in red. Brackets indicate protein-specific insertions that are hidden for clarity. The positions of Exu β-hairpin insertion, loop1 and linker are indicated. For each protein, residues involved in homodimerization are highlighted in light blue.

b-e) Cartoon (left) and surface (middle and right) representation of the indicated protein structures. Exu^{333} (**b**) and Trex1 (**d**) are in the same orientation as in **Fig. 2a, b**. Residues in the (pseudo-)catalytic site are shown as sticks; ions (Zn^{2+} in (**c**) and Mn^{2+} in (**d**)) are represented as spheres. Protein-specific features are highlighted in red: linker helix and β-hairpin insertion in Exu (**b**); Zn^{2+} coordinating extension in Maelstrom (Mael) (**c**). Surface is colored according to the electrostatic charge, with the (pseudo)-catalytic site boxed in yellow. Structural alignment was done with cealign in Pymol (v 1.7.6.0). Scale of electrostatic charge distribution is the same for all domains (-5 to +5 kT/e).


a
Exu333
Asp41
Arg264
Asp39
b
Exu406
Arg92B
Arg92A

Supplementary Figure 4
Quality of the electrondensity.
Stereo view of the electrondensity of the 2Fo-DFc maps for Exu^{333} (**a**) and Exu^{406} (**b**) structures after refinement. The structures are shown in similar views as in **Fig. 2e** and **Fig. 3b**, respectively. Monomer A is colored red, monomer B gray, the electrondensity visualized as a teal mesh contoured at 1σ. Water molecules are represented as light blue spheres.

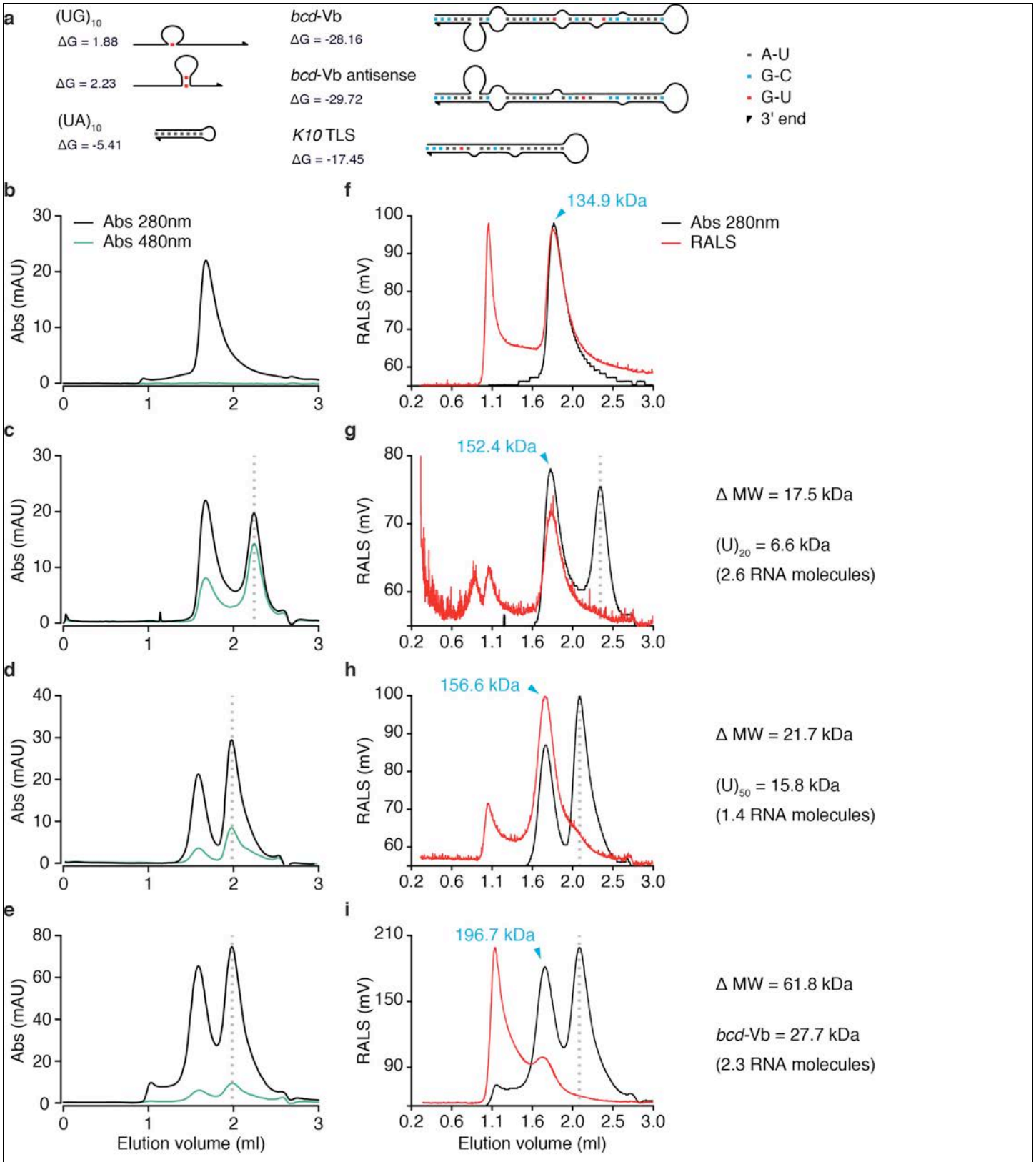


Supplementary Figure 5

Exu and Vts1p SAM domains use different surfaces for RNA binding.

a-b) Surface rendering of electrostatic charges of Exu SAM-like (**a**) and yeast Vts1p (**b**; PDB 2B6G, chain A) domain structure (see also **Supplementary Fig. 2**). **a**) Exu residues cross-linking with RNA (**Fig. 5**) are underlined. **b**) Vts1p residues shown to interact directly with the RNA (Aviv *et al.*, *NSMB* **13**, 168-176, 2006; Johnson and Donaldson, *NSMB* **13**, 177-178, 2006; Oberstrass *et al.*, *NSMB* **13**, 160-167, 2006) are indicated. Structural alignment was done with cealign in Pymol (v 1.7.6.0). Scale of electrostatic charge distribution is the same for all domains (-5 to +5 kT/e).

c-d, f-g) A constant amount of 5'-fluorescein labelled oligo was incubated with increasing concentrations of recombinantly purified Exu wt or mutant. The fluorescence anisotropy data were fitted to the Hill equation to obtain the dissociation constant (K_d); mean K_d and standard deviation from three independent experiments are reported in tables (**d, g**). **c, f**) Data from a representative fluorescence anisotropy measurement, with the best fit plotted as a solid line. **c, d**) FA measurements of SAM-like domain mutants with $(U)_{20}$. **f, g**) Affinity of Exu for oligo(U) RNAs of increasing length. **e**) Circular Dichroism (CD) spectra of Exu wt and the indicated mutants.

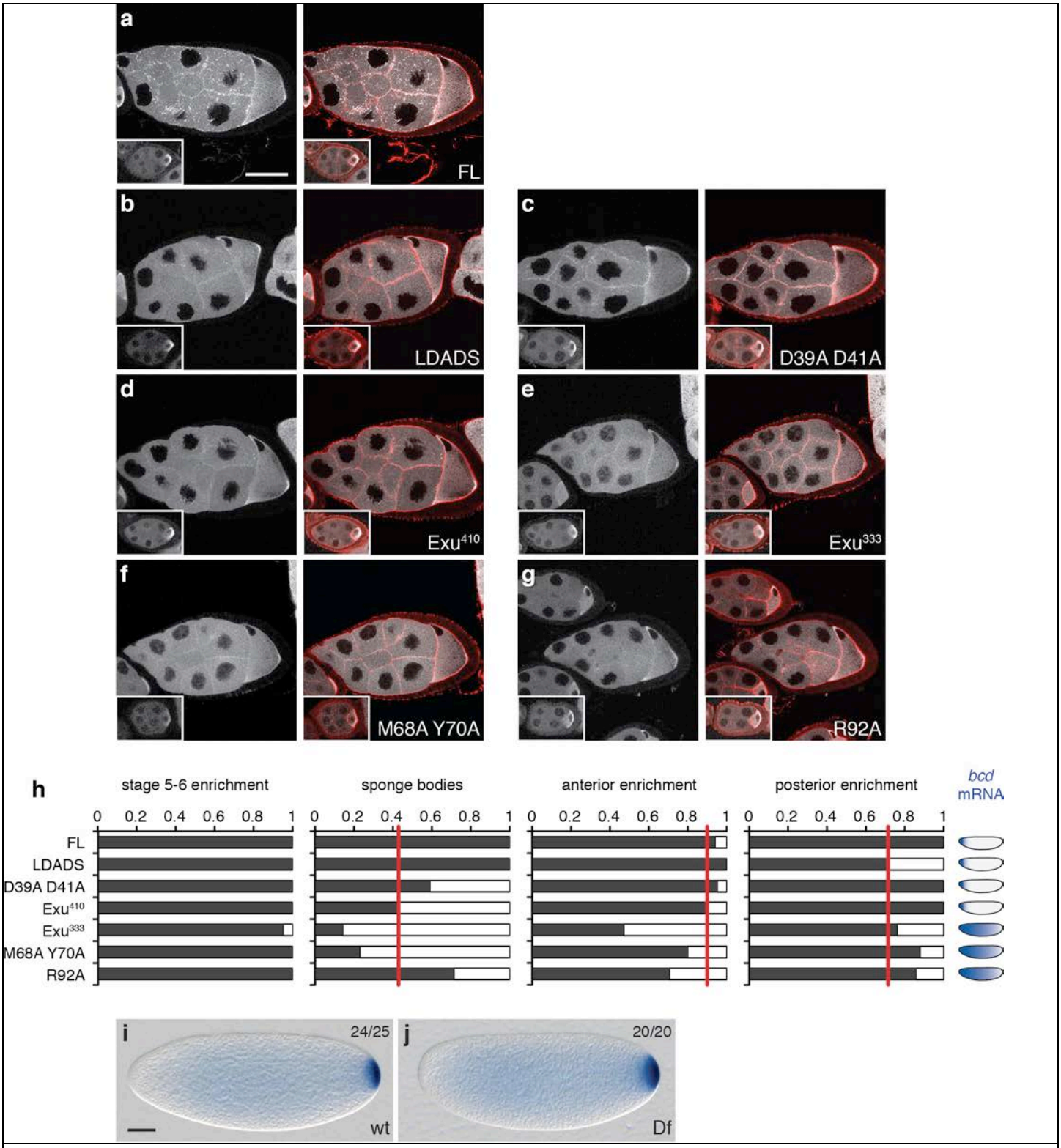


Supplementary Figure 6

Exu can bind two RNA molecules.

a) Oligonucleotides used in this study predicted to have a secondary structure, using the mFold server (Zuker, *Nucleic Acids Res.* **31**, 3406-3415, 2003). Numbers indicate the ΔG (in kcal/mol) at 22°C.

b-e) SEC profiles of purified Exu alone (**b**) or pre-incubated with fluorescein-conjugated $(U)_{20}$ (**c**), $(U)_{50}$ (**d**) or *bcd*-Vb (**e**). The elution volume of the free RNA is marked by a dotted line. **f-i**) Static light scattering profiles of the samples in (**b-e**). For each plot, the calculated molecular weight (MW) at the peak is indicated in blue; the difference in molecular weight (Δ MW) between the Exu-RNA complex and Exu alone is indicated on the right, together with the MW of the corresponding fluorescein-conjugated oligonucleotide. RALS = Right Angle Light Scattering.



Supplementary Figure 7

Localization of Venus-tagged Exu transgenes.

a-g) Fluorescence microscopy images of fixed *Drosophila* egg chambers at stage 5-6 (small box) or stage 9. The localization of Venus-

tagged Exu wt (**a**) and mutants (**b-g**) is shown in gray (left image); the merged images (right) show Venus-Exu in gray and Rhodamine-Phalloidin in red. **h**) For each genotype, at least 20 egg chambers were scored for the presence (black) or absence (white) of the following characteristics: enrichment of Venus-tagged Exu in the oocyte at early stages (1st panel from left); enrichment in sponge bodies in the nurse cells at stage 9 (2nd panel); enrichment at the anterior pole of the oocyte at stage 9 (3rd panel); enrichment at the posterior pole of the oocyte at stage 9 (4th panel). The rightmost column schematizes the localization of *bcd* mRNA in early embryos of the corresponding genotype (as in **Fig. 7**). Red lines mark the lowest percentage of egg chambers having the indicated characteristic amongst the Exu constructs which rescue *bcd* mRNA localization.

i-j) *oskar* (*osk*) *in situ* hybridization of *Drosophila* early embryos (0-2 h). Numbers at the top right corner indicate the number of embryos displaying the illustrated phenotype vs the total number of embryos examined. The genotype of each embryo is reported at the bottom of the image: **i**) wt; **j**) *Df(2R)exu1*/*exu*^{VL} (Df). *osk* localization is not impaired in embryos lacking Exu. Scale bars: 50 μm.

position	peptide	RNA	cross-linked aa
L47-R78	LMDEIVQLAAYTPTDHFEQYIMPYMNLNPAAR	U	L47-M48
L47-R78	LM(Oxidation)DEIVQLAAYTPTDHFEQYIM(Oxidation)PYM(Oxidation)NLNPAAR	U-H2O	P75-A76
H81-R92	HQVRVISIGFYR	UU	R84
V85-R92	VISIGFYR	U-H2O	F90
V85-K95	VISIGFYRMLK	U	Y91
M93-K101	MLKSMQTYK	U	K95
S96-K101	SMQTYK	U	Y100
S96-K101	SMQTYK	UU	M97
S105-K113	SKSEIAALK	U	-
A126-R141	AGPSSDGIVLIYHEER	U	Y137
A126-K142	AGPSSDGIVLIYHEERK	U	R141
K154-R160	KYGLLER	U	K154
S167-K176	SFANSINLAK	U	S167-F168
A177-K185	ASIGDANIK	U	-
A177-R190	ASIGDANIKNYSLR	U	K185
A177-R190	ASIGDANIKNYSLR	UU -H2O	N183-I184
I195-R264	ILSLDADSLFDGNASVR	U -H2O	F257
Q321-K333	QNSFRPVFLNYFK	U	R325
T334-R339	TTLYHR	U -H2O	L336
T334-R339	TTLYHR	U	Y337
T334-R341	TTLYHRVR	U	R339
I347-K366	IVLAENGFDLNTLSAIWAEK	U -H2O	-
N367-R380	NIEGLDIALQSIGR	U -H2O	I373
N367-K382	NIEGLDIALQSIGRLK	UU	G379-R380
S383-K400	SKDKAELLELLDSYFDPK	U	Y396
T402-K409	TTVKPVVK	U	T402-T403
G410-R419	GNSNNNNNYR	U	Y418
D432-K451	DARPSSSPSASTEFGAGGDK	U -HPO3	F445
D432-K451	DARPSSSPSASTEFGAGGDK	U -H2O	P435-S436
S454-K464	SVSSLPDSTTK	U -H2O	-
T465-R472	TPSPNKPR	UU -HPO3	-
Q481-K491	QSLGATPNGLK	U -H2O	L490

Supplementary Table 1. UV-cross-linking and MS results.

The first tab summarizes all unique cross-linked regions together with the position and sequence of the cross-linked peptide, the composition of the cross-linked RNA, and the position or range of the cross-linked amino acid(s). The second tab lists all identified cross-links that include redundant information, i.e., the same peptide cross-linked to RNA of different length or with different modifications, or the same RNA cross-linked to an unmodified and modified or missed cleaved version of the same peptide. In addition, calculated values for peptide, RNA and cross-link masses are listed together with experimental mass values and mass errors.

position	peptide	RNA	cross-linked aa	peptide	RNA	cross-link	z	m/z	m/z exp	error ppm
R46-R78	RLMDEIVQLAAYTPTDHFEQYIM(Oxidation)PYM(Oxidation)NLNPAAR	U	R46-D49	3942.8742	324.0359	4266.9101	4	1067.7353	1067.7347	-0.6
R46-R78	RLMDEIVQLAAYTPTDHFEQYIM(Oxidation)PYM(Oxidation)NLNPAAR	UU	-	3942.8742	630.0612	4572.9354	4	1144.2416	1144.2389	-2.4
L47-R78	LMDEIVQLAAYTPTDHFEQYIMPYMNLNPAAR	U	L47-M48	3754.7833	324.0359	4078.8192	3	1360.6142	1360.6142	0.0
L47-R78	LM(Oxidation)DEIVQLAAYTPTDHFEQYIMPYMNLNPAAR	U	-	3770.7782	324.0359	4094.8141	4	1024.7113	1024.7082	-3.0
L47-R78	LM(Oxidation)DEIVQLAAYTPTDHFEQYIMPYMNLNPAAR	UU	-	3770.7782	630.0612	4400.8394	3	1467.9543	1467.9580	2.6
L47-R78	LMDEIVQLAAYTPTDHFEQYIM(Oxidation)PYMNLNPAAR	U	L47-M48	3770.7783	324.0359	4094.8142	4	1024.7114	1024.7095	-1.8
L47-R78	LMDEIVQLAAYTPTDHFEQYIMPYM(Oxidation)NLNPAAR	U	L47-M48	3770.7783	324.0359	4094.8142	4	1024.7114	1024.7082	-3.1
L47-R78	LM(Oxidation)DEIVQLAAYTPTDHFEQYIMPYM(Oxidation)NLNPAAR	U	L47-M48	3786.7731	324.0359	4110.8090	4	1028.7101	1028.7106	0.5
L47-R78	LMDEIVQLAAYTPTDHFEQYIM(Oxidation)PYM(Oxidation)NLNPAAR	U	L47-M48	3786.7731	324.0359	4110.8090	3	1371.2775	1371.2764	-0.8
L47-R78	LM(Oxidation)DEIVQLAAYTPTDHFEQYIM(Oxidation)PYM(Oxidation)NLNPAAR	U-H2O	P75-A76	3802.7680	306.0253	4108.7933	4	1028.2061	1028.2069	0.8
L47-R78	LM(Oxidation)DEIVQLAAYTPTDHFEQYIM(Oxidation)PYM(Oxidation)NLNPAAR	U	-	3802.7680	324.0359	4126.8039	4	1032.7088	1032.7089	0.1
H81-R92	HQVRVISIGFYR	UU	R84	1473.8204	630.0612	2103.8816	3	702.3017	702.3012	-0.6
V85-R92	VISIGFYR	U-H2O	F90	953.5334	306.0253	1259.5587	2	630.7872	630.7868	-0.6
V85-R92	VISIGFYR	U	F90	953.5334	324.0359	1277.5693	2	639.7925	639.7914	-1.6
V85-R92	VISIGFYR	UU -H3PO4	G89-F90	953.5334	532.0843	1485.6177	2	743.8166	743.8160	-0.9
V85-R92	VISIGFYR	UU -HPO3	F90	953.5334	550.0948	1503.6282	2	752.8219	752.8214	-0.7
V85-R92	VISIGFYR	UU -H2O	F90	953.5334	612.0506	1565.5840	2	783.7998	783.7990	-1.0
V85-R92	VISIGFYR	UU	F90	953.5334	630.0612	1583.5946	2	792.8051	792.8045	-0.7
V85-R92	VISIGFYR	UUU -HPO3	G89-F90	953.5334	856.1201	1809.6535	2	905.8346	905.8336	-1.1
V85-R92	VISIGFYR	UUU -H2O	F90	953.5334	918.0759	1871.6093	2	936.8125	936.8114	-1.1
V85-R92	VISIGFYR	UUU	F90	953.5334	936.0865	1889.6199	2	945.8177	945.8162	-1.6
V85-K95	VISIGFYRMLK	U	Y91	1325.7529	324.0359	1649.7888	2	825.9022	825.9018	-0.5
V85-K95	VISIGFYRMLK	UU	Y91	1325.7529	630.0612	1955.8141	2	978.9148	978.9143	-0.5
V85-K95	VISIGFYRM(Oxidation)LK	U	Y91	1341.7478	324.0359	1665.7837	3	556.2690	556.2683	-1.3
V85-K95	VISIGFYRM(Oxidation)LK	UU	Y91	1341.7478	630.0612	1971.8090	3	658.2775	658.2761	-2.1
V85-K95	VISIGFYRM(Oxidation)LK	UUU	-	1341.7478	936.0865	2277.8343	3	760.2859	760.2855	-0.5
M93-K101	MLKSMQTYK	U	K95	1128.5671	324.0359	1452.6030	2	727.3093	727.3082	-1.5
M93-K101	MLKSMQTYK	UU	K95	1128.5671	630.0612	1758.6283	2	880.3219	880.3210	-1.1
M93-K101	MLKSMQTYK	UUU	-	1128.5671	936.0865	2064.6536	2	1033.3346	1033.3329	-1.6
M93-K101	M(Oxidation)LKSMQTYK	UU	K95	1144.5620	630.0612	1774.6232	2	888.3194	888.3172	-2.5
M93-K101	MLKSM(Oxidation)QTYK	U	K95	1144.5620	324.0359	1468.5979	2	735.3068	735.3072	0.6
M93-K101	MLKSM(Oxidation)QTYK	UU	-	1144.5620	630.0612	1774.6232	2	888.3194	888.3170	-2.7
S96-K101	SMQTYK	U	S96-M97	756.3476	324.0359	1080.3835	2	541.1996	541.1993	-0.5
S96-K101	SMQTYK	U	Y100	756.3476	324.0359	1080.3835	2	541.1996	541.1987	-1.6
S96-K101	SMQTYK	UU	M97	756.3476	630.0612	1386.4088	2	694.2122	694.2115	-1.0
S96-K101	SM(Oxidation)QTYK	U-H2O	Y100	772.3425	306.0253	1078.3678	2	540.1917	540.1907	-1.9
S96-K101	SM(Oxidation)QTYK	U	Y100	772.3425	324.0359	1096.3784	2	549.1970	549.1964	-1.1
S96-K104	SMQTYKIIK	UU -H2O	-	1110.6107	612.0506	1722.6613	2	862.3385	862.3373	-1.3
S96-K104	SMQTYKIIK	UU	Y100	1110.6107	630.0612	1740.6719	2	871.3437	871.3448	1.2
S96-K104	SM(Oxidation)QTYKIIK	UU	Y100	1126.6056	630.0612	1756.6668	2	879.3412	879.3407	-0.5
S105-K113	SKSEIAALK	U	-	945.5494	324.0359	1269.5853	2	635.8005	635.7997	-1.2
S105-K113	SKSEIAALK	UU	-	945.5494	630.0612	1575.6106	2	788.8131	788.8134	0.4
T124-R141	TKAGPSSDGIVLIYHEER	U	-	1971.0061	324.0359	2295.0420	3	766.0218	766.0181	-4.8
A126-R141	AGPSSDGIVLIYHEER	U	Y137	1741.8634	324.0359	2065.8993	2	1033.9575	1033.9571	-0.3
A126-K142	AGPSSDGIVLIYHEERK	U	R141	1869.9584	324.0359	2193.9943	3	732.3392	732.3385	-1.0
A126-K142	AGPSSDGIVLIYHEERK	UU	R141	1869.9584	630.0612	2500.0196	3	834.3477	834.3475	-0.2
A126-K142	AGPSSDGIVLIYHEERK	UUU	R141	1869.9584	936.0865	2806.0449	3	936.3561	936.3558	-0.3
K154-R160	KYGLLER	U	K154	877.5021	324.0359	1201.5380	2	601.7768	601.7759	-1.5
K154-R160	KYGLLER	UUU	-	877.5021	936.0865	1813.5886	2	907.8021	907.8008	-1.4
S167-K176	SFANSINLAK	U	S167-F168	1063.5661	324.0359	1387.6020	2	694.8088	694.8087	-0.1
A177-K185	ASIGDANIK	U	-	887.4712	324.0359	1211.5071	2	606.7614	606.7600	-2.2
A177-R190	ASIGDANIKNYSLR	U	K185	1520.7946	324.0359	1844.8305	3	615.9513	615.9507	-1.0
A177-R190	ASIGDANIKNYSLR	UU -H2O	N183-I184	1520.7946	612.0506	2132.8452	3	711.9562	711.9552	-1.4
A177-R190	ASIGDANIKNYSLR	UU	K185	1520.7946	630.0612	2150.8558	3	717.9597	717.9589	-1.1
I195-R264	ILSLDADSLFDGNASVR	U -H2O	F257	1791.9002	306.0253	2097.9255	3	700.3163	700.3157	-0.9
I195-R264	ILSLDADSLFDGNASVR	U	F257	1791.9002	324.0359	2115.9361	2	1058.9759	1058.9751	-0.7
I195-R264	ILSLDADSLFDGNASVR	UU	-	1791.9002	630.0612	2421.9614	2	1211.9885	1211.9821	-5.3
Q321-K333	QNSFRPVFLNYFK	U	R325	1658.8568	324.0359	1982.8927	3	661.9720	661.9710	-1.6
Q321-K333	QNSFRPVFLNYFK	UU	R325	1658.8568	630.0612	2288.9180	3	763.9805	763.9799	-0.7
Q321-K333	QNSFRPVFLNYFK	UUU	-	1658.8568	936.0865	2594.9433	3	865.9889	865.9880	-1.0
Q321-K333	QNSFRPVFLNYFK	UUUU	-	1658.8568	1242.1118	2900.9686	3	967.9973	967.9963	-1.1
T334-R339	TTLYHR	U -H2O	L336	789.4133	306.0253	1095.4386	2	548.7271	548.7262	-1.6
T334-R339	TTLYHR	U	Y337	789.4133	324.0359	1113.4492	2	557.7324	557.7317	-1.3
T334-R339	TTLYHR	UU -H2O	L336	789.4133	612.0506	1401.4639	2	701.7398	701.7386	-1.6
T334-R339	TTLYHR	UU	Y337	789.4133	630.0612	1419.4745	2	710.7450	710.7444	-0.9
T334-R341	TTLYHRVR	U	R339	1044.5828	324.0359	1368.6187	3	457.2140	457.2134	-1.4
T334-R341	TTLYHRVR	UU	-	1044.5828	630.0612	1674.6440	3	559.2225	559.2217	-1.3
F345-K366	FRIVLAENGFDLNTLSAIWAEK	U	-	2506.3219	324.0359	2830.3578	3	944.4604	944.4581	-2.4
I347-K366	IVLAENGFDLNTLSAIWAEK	U -H2O	-	2203.1524	306.0253	2509.1777	3	837.4004	837.3993	-1.3
I347-K366	IVLAENGFDLNTLSAIWAEK	UU	-	2203.1524	630.0612	2833.2136	3	945.4123	945.4111	-1.3
N367-R380	NIEGLDIALQSIGR	U -H2O	I373	1497.8150	306.0253	1803.8403	2	902.9280	902.9282	0.3
N367-R380	NIEGLDIALQSIGR	U	I373	1497.8150	324.0359	1821.8509	2	911.9333	911.9330	-0.3
N367-K382	NIEGLDIALQSIGRLK	U	-	1738.9940	324.0359	2063.0299	2	1032.5228	1032.5226	-0.1
N367-K382	NIEGLDIALQSIGRLK	UU	G379-R380	1738.9940	630.0612	2369.0552	2	1185.5354	1185.5340	-1.2
S383-K400	SKDKAELLELLDSYFDPK	U	Y396	2110.0833	324.0359	2434.1192	3	812.3809	812.3797	-1.4
D385-K400	DKAELLELLDSYFDPK	U -H2O	Y396	1894.9563	306.0253	2200.9816	3	734.6683	734.6675	-1.1
D385-K400	DKAELLELLDSYFDPK	U	Y396	1894.9563	324.0359	2218.9922	2	1110.5039	1110.5042	0.3
D385-K400	DKAELLELLDSYFDPK	UU	-	1894.9563	630.0612	2525.0175	3	842.6803	842.6804	0.1
D385-K401	DKAELLELLDSYFDPKK	U	-	2023.0513	324.0359	2347.0872	3	783.3702	783.3700	-0.3
D385-K401	DKAELLELLDSYFDPKK	UU	-	2023.0513	630.0612	2653.1125	3	885.3786	885.3796	1.1
A387-K401	AELLELLDSYFDPK	U -H2O	Y396	1651.8344	306.0253	1957.8597	2	979.9377	979.9383	0.7
A387-K401	AELLELLDSYFDPK	U	Y396	1651.8344	324.0359	1975.8703	2	988.9430	988.9430	0.1
A387-K401	AELLELLDSYFDPK	UU	-	1651.8344	630.0612	2281.8956	2	1141.9556	1141.9540	-1.4
A387-K402	AELLELLDSYFDPKK	U -H2O	Y396-S397	1779.9294	306.0253	2085.9547	3	696.3260	696.3253	-1.1
A387-K402	AELLELLDSYFDPKK	U	Y396	1779.9294	324.0359	2103.9653	2	1052.9905	1052.9891	-1.3
A387-K402	AELLELLDSYFDPKK	UUU	-	1779.9294	936.0865	2716.0159	3	906.3464	906.3455	-1.0
K401-K409	KTTVKPVVK	U	-	998.6487	324.0359	1322.6846	2	662.3501	662.3497	-0.6
K401-K409	KTTVKPVVK	UU	-	998.6487	630.0612	1628.7099	3	543.9111	543.9105	-1.1
K401-K409	KTTVKPVVK	UUU	-	998.6487	936.0865	1934.7352	2	968.3754	968.3756	0.2
T402-K409	TTVKPVVK	U	T402-T403	870.5538	324.0359	1194.5897	2	598.3027	598.3020	-1.1
T402-K409	TTVKPVVK	UU	-	870.5538	630.0612	1500.6150	2	751.3153	751.3148	-0.6
T402-R419	TTVKPVVKGNSNNNNNYR	UUU	-	2018.0293	936.0865	2954.1158	3	985.7131	985.7140	1.0
G410-R419	GNSNNNNNYR	U	Y418	1165.4860	324.0359	1489.5219	2	745.7688	745.7683	-0.6
Q428-K451	QSVKDARPSSSPSASTEFGAGGDK	U -H2O	-	2365.1145	306.0253	2671.1398	3	891.3877	891.3857	-2.3
Q428-K451	QSVKDARPSSSPSASTEFGAGGDK	UU -HPO3	-	2365.1145	550.0948	2915.2093	3	972.7442	972.7439	-0.4
Q428-K451	QSVKDARPSSSPSASTEFGAGGDK	UU	-	2365.1145	630.0612	2995.1757	3	999.3997	999.3980	-1.7
Q428-K451	QSVKDARPSSSPSASTEFGAGGDK	UUU -HPO3	-	2365.1145	856.1201	3221.2346	3	1074.7527	1074.7502	-2.3
Q428-K451	QSVKDARPSSSPSASTEFGAGGDK	UUU	-	2365.1145	936.0865	3301.2010	3	1101.4081	1101.4108	2.4
D432-K451	DARPSSSPSASTEFGAGGDK	U -HPO3	F445	1922.8605	244.0695	2166.9300	3	723.3178	723.3147	-4.3
D432-K451	DARPSSSPSASTEFGAGGDK	U -H2O	P435-S436	1922.8605	306.0253	2228.8858	3	743.9697	743.9690	-1.0
D432-K451	DARPSSSPSASTEFGAGGDK	U	F445	1922.8605	324.0359	2246.8964	3	749.9733	749.9729	-0.5
D432-K451	DARPSSSPSASTEFGAGGDK	UU -H3PO4	-	1922.8605	532.0843	2454.9448	3	819.3227	819.3245	2.2
D432-K451	DARPSSSPSASTEFGAGGDK	UU -HPO3	-	1922.8605	550.0948	2472.9553	3	825.3262	825.3253	-1.1
D432-K451	DARPSSSPSASTEFGAGGDK	UU -H2O	P435-S436	1922.8605	612.0506	2534.9111	3	845.9782	845.9779	-0.3
D432-K451	DARPSSSPSASTEFGAGGDK	UU	F445	1922.8605	630.0612	2552.9217	3	851.9817	851.9810	-0.8

D432-K451	DARPSSSPSASTEFGAGGDK	UUU -HPO3	-	1922.8605	856.1201	2778.9806	3	927.3347	927.3344	-0.3
D432-K451	DARPSSSPSASTEFGAGGDK	UUU -H2O	P435-S438	1922.8605	918.0759	2840.9364	3	947.9866	947.9858	-0.8
D432-K451	DARPSSSPSASTEFGAGGDK	UUU	-	1922.8605	936.0865	2858.9470	3	953.9901	953.9894	-0.8
D432-R453	DARPSSSPSASTEFGAGGDKSR	U	-	2165.9937	324.0359	2490.0296	3	831.0177	831.0165	-1.4
D432-R453	DARPSSSPSASTEFGAGGDKSR	UUU	-	2165.9937	936.0865	3102.0802	3	1035.0345	1035.0349	0.4
S454-K464	SVSSLPDSTTK	U -H2O	-	1120.5611	306.0253	1426.5864	2	714.3010	714.3003	-1.0
S454-K464	SVSSLPDSTTK	U	-	1120.5611	324.0359	1444.5970	2	723.3063	723.3042	-2.9
S454-R472	SVSSLPDSTTKTPSPNKPR	UU	-	1998.0381	630.0612	2628.0993	3	877.0409	877.0411	0.2
S454-R472	SVSSLPDSTTKTPSPNKPR	UUU	-	1998.0381	936.0865	2934.1246	3	979.0493	979.0487	-0.6
T465-R472	TPSPNKPR	UU -HPO3	-	895.4875	550.0948	1445.5823	2	723.7990	723.7983	-0.9
T465-R472	TPSPNKPR	UU	-	895.4875	630.0612	1525.5487	2	763.7821	763.7817	-0.6
T465-R472	TPSPNKPR	UUU	-	895.4875	936.0865	1831.5740	2	916.7948	916.7940	-0.9
N478-K491	NSRQSLGATPNGLK	U	-	1441.7637	324.0359	1765.7996	3	589.6077	589.6069	-1.3
N478-K491	NSRQSLGATPNGLK	UU	-	1441.7637	630.0612	2071.8249	3	691.6161	691.6144	-2.4
Q481-K491	QSLGATPNGLK	U -H2O	L490	1084.5876	306.0253	1390.6129	2	696.3143	696.3134	-1.2
Q481-K491	QSLGATPNGLK	U	-	1084.5876	324.0359	1408.6235	2	705.3196	705.3188	-1.1
Q481-K491	QSLGATPNGLK	UU	-	1084.5876	630.0612	1714.6488	2	858.3322	858.3331	1.1