The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs

Mahesh Lingaraju¹, Dennis Johnsen²,⁷, Andreas Schlundt³,⁴,⁵,⁷, Lukas M. Langer ³, Jérôme Basquin¹, Michael Sattler ³,⁴, Torben Heick Jensen², Sebastian Falk ¹,⁶ & Elena Conti¹

The nuclear exosome and its essential co-factor, the RNA helicase MTR4, play crucial roles in several RNA degradation pathways. Besides unwinding RNA substrates for exosome-mediated degradation, MTR4 associates with RNA-binding proteins that function as adaptors in different RNA processing and decay pathways. Here, we identify and characterize the interactions of human MTR4 with a ribosome processing adaptor, NVL, and with ZCCHC8, an adaptor involved in the decay of small nuclear RNAs. We show that the unstructured regions of NVL and ZCCHC8 contain short linear motifs that bind the MTR4 arch domain in a mutually exclusive manner. These short sequences diverged from the arch-interacting motif (AIM) of yeast rRNA processing factors. Our results suggest that nuclear exosome adaptors have evolved canonical and non-canonical AIM sequences to target human MTR4 and demonstrate the versatility and specificity with which the MTR4 arch domain can recruit a repertoire of different RNA-binding proteins.

¹Department of Structural Cell Biology, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. ²Department of Molecular Biology and Genetics, Aarhus University, C.F. Møllers Alle 3, 8000 Aarhus C, Denmark. ³Center for Integrated Protein Science Munich (CIPSM) at Department of Chemistry, Technical University of Munich (TUM), 85747 Garching, Germany. ⁴Institute of Structural Biology, Helmholtz-Zentrum München, 85764 Neuherberg, Germany. ⁵Present address: Institute for Molecular Biosciences and Center for Biomolecular Magnetic Resonance (BMRZ) at Johann Wolfgang Goethe-University, Frankfurt am Main 60438, Germany. ⁶Present address: Max F. Perutz Laboratories, Department of Structural and Computational Biology, University of Vienna, Campus Vienna Biocenter 5, 1030 Vienna, Austria. ⁷These authors contributed equally: Dennis Johnsen, Andreas Schlundt. Correspondence and requests for materials should be addressed to S.F. (email: sebastian.falk@univie.ac.at) or to E.C. (email: conti@biochem.mpg.de)
Eukaryotic cells generate a multitude of RNA species that require timely maturation and decay to maintain a healthy transcriptome. A central machinery in nuclear RNA processing, quality control and decay pathways is a conserved 3′→5′ exoribonuclease complex known as the RNA exosome (reviewed in12,16). Most mechanistic studies to date have analyzed the RNA exosome from S. cerevisiae, the species in which it was originally identified 20 years ago1. The yeast exosome consists of a 10-subunit complex (Exo10), the activity of which depends on a single processive exoribonuclease (Rrp44, also known as Dis3)5,6. The Exo10 core is present in both nuclear and cytoplasmic compartments, but its cofactors and regulators have distinct subcellular localizations (reviewed in12,16). In the nucleus, Exo10 is bound to the distributive ribonuclease Rrp6 and its associated protein Rrp47 as well as to the Mpp6 protein7,8. Together, Rrp6–Rrp47 and Mpp6 recruit the RNA helicase Mtr4 to the exosome10–12. Orthologues of all these 14 proteins exist in human cells, and engage in similar interactions to form the corresponding human nuclear exosome complex12,13.

In both yeast and humans, the nuclear helicase Mtr4/MTR4 is central to exosome function14,15. First, it functions as an enzyme unwinding activity, and to present the unwound RNA substrate to the exosome core12,16. Furthermore, it functions as a binding platform for RNA-binding adaptors, providing the primary interactions to transcripts subjected to exosomal degradation in both RNA processing and decay pathways17,18. For example, S. cerevisiae Mtr4 binds Nop53, a ribosome biogenesis factor that recruits the exosome for a late step in rRNA processing, namely the trimming of ITS2 (Internal Transcribed Spacer 2). Two other factors, Trf4 and Air2, bind Mtr4 to form the so-called TRAMP complex19–21, which allows the exosome to target aberrant tRNA22, rRNAs and small nuclear and nucleolar RNAs (sn/snoRNAs) for decay23. Higher eukaryotes not only have orthologues of Nop53 and TRAMP, but also have an increased number of nuclear exosome adaptors. In human cells, MTR4 has been reported to interact with the early ribosome biogenesis factors WDR74 and NVL, which take part in the processing of ITS1 (Internal Transcribed Spacer 1)24,25. Human MTR4 also binds to two large Zinc-finger proteins, ZCCHC8 and ZFC3H1. ZCCHC8 interacts with MTR4 and the RNA-binding protein RRM7 to form the trimERIC Nuclear EXosome Targeting (NEXT) complex17,26, which targets enhancer RNAs (eRNAs), promoter upstream transcripts (PROMPTs) and intronic RNAs for exosome-mediated decay27. ZFC3H1 instead directs MTR4 and the nuclear exosome to polyadenylated nuclear RNAs by connecting to the nuclear poly(A) binding protein PABN128–30. Furthermore, MTR4 binds NRDE-2, a negative regulator that prevents the nuclear exosome from targeting RNAs that should be exported to the cytoplasm31.

How does the nuclear exosome helicase mediate binding to such a diverse and functionally distinct set of proteins? Structural studies have shown that yeast Mtr4 contains an unstructured N-terminal region, a DEX helicase core and an ‘arch’ domain with a globular Kyprides, Ouzounis, Woese (KOW) domain14,32. All Mtr4 domains are involved in protein–protein interactions: the N-terminal region binds Rrp6–Rrp4711, the helicase core binds Mpp6 as well as Trf4–Air233,34 and the KOW domain binds a short sequence known as ‘arch-interacting motif’ that is present in Nop53, Utp18, and Air235,36. In the case of human MTR4, both the arch domain and the DEX helicase core bind NRDE-231. Structural data have also elucidated how the helicase core of human MTR4 binds a region of ZCCHC837. This protein is however expected to harbor another MTR4-binding region37, which remains to be identified. Also unclear is how other human MTR4-binding proteins are recognized. With the exception of the expected AIM sequence in human NOP53 orthologue (also known as GLTSCR2) and in NRDE-231, there is no consensus motif that can be identified with confidence at the sequence level in other MTR4-binding factors. Here, we used a combination of biochemical studies, X-ray crystallography and nuclear magnetic resonance (NMR) experiments to obtain mechanistic insights into how human MTR4 interacts with two metazoan RNA exosome adaptors: the RNA processing factor NVL and the RNA decay factor ZCCHC8.

Results

The unstructured region of human NVL interacts with MTR4. The nuclear VCP-like (NVL) protein, also known as NVL2, is a ribosome biogenesis factor of the AAA-ATPase family that has been reported to interact with human MTR438. NVL is a multi-domain protein characterized by an N-terminal nucleolin-bind- ing domain (residues 10–74)39 and a linker region (residues 76–266) followed by two globular domains, characteristic of AAA-ATPases that are responsible for catalytic activity (Fig. 1a, Supplementary Fig. 1a). Bioinformatic analyses suggested that the linker region is mostly unstructured (residues 76–239) (Supple- mental Fig. 1b). In other AAA-ATPases, unstructured regions upstream of the catalytic domains often mediate protein-protein interactions40–42. To test if the portion upstream of the NVL catalytic domains mediates the interaction with MTR4, we expressed and purified the N-terminal region of human NVL tagged with thioredoxin (Trx) as the prey protein (Trx–NVL1–266) and the structured portion of MTR4 tagged with glutathione-S-transferase (GST) as the bait protein (GST–MTR4–ΔN). In these assays, Trx–NVL1–266 specifically co-precipitated with GST–MTR4–ΔN (Fig. 1b, lane 11, compare with GST control in lane 9). Interestingly, pull-down assays with the recombinant yeast orthologues showed that the N-terminal unstructured region of Rix7 does not interact with yeast Mtr4 (Supplementary Fig. 1c), indicating that the interaction between human NVL and MTR4 is not conserved in S. cerevisiae.

We narrowed down the MTR4-interacting region of human NVL based on bioinformatic analysis. Sequence alignments showed that the N-terminal unstructured region of NVL contains an insertion that is present in the human protein and other chordates but not in the yeast orthologue (Supplementary Fig. 1a). Upon testing whether this insertion is responsible for the interaction with human MTR4, we found that a construct encompassing the conserved portion of the human NVL insertion spanning residues 167–216 (Supplementary Fig. 1a) was indeed able to co-precipitate with GST–MTR4–ΔN in pull-down assays (Fig. 1b, lane 14, Trx–NVL167–216). Thus, the MTR4-binding determinants of NVL reside in an unstructured region that is present in the human orthologue and more generally in chordate NVL proteins, but not in lower eukaryotes.

NVL targets the MTR4 KOW domain. To identify where the MTR4-binding determinants for NVL reside, we performed GST pull down assays with MTR4 constructs harboring the DEXH core of the helicase (GST–MTR4–ΔNarch) and the KOW domain (GST–MTR4–KOW). Neither Trx–NVL1–266 nor Trx–NVL167–216 co-precipitated with GST–MTR4–ΔNarch (Fig. 1b, lanes 12 and 15). In contrast, the KOW domain characteristic of the MTR4 arch (GST–MTR4–KOW) was able to co-precipitate both Trx–NVL1–266 and Trx–NVL167–216 (Fig. 1b, lanes 13 and 16). We quantified the strength of the interaction by biophysical approaches. Using microscale thermophoresis (MST), we measured a dissociation constant of ~1 μM between a YFP-tagged version of NVL167–216 and MTR4–KOW (Fig. 1c). A similar dissociation constant was obtained when testing the interaction of NVL167–216
and MTR4-ΔN using isothermal titration calorimetry (ITC) (Supplementary Fig. 1d). The fact that the KOW domain alone binds NVL167–216 as strongly as MTR4-ΔN indicated that the NVL167–216 binding region resides within the KOW domain.

We proceeded to characterize the NVL167–216 binding site of MTR4 with structural approaches. Using NMR spectroscopy, we confirmed that the MTR4-KOW domain has a sequential arrangement of secondary structure elements consistent with the fold observed in a recent crystal structure of an MTR4-NRDE-2 complex31, namely a five-stranded β-sheet flanked by a long C-terminal helix and containing smaller helical segments within loops (Supplementary Fig. 2). We then carried out titration experiments using a 15N-labelled MTR4-KOW protein and adding increasing amounts of unlabeled NVL167–216. In line with a dissociation constant of ~1 μM, we observed an intermediate exchange regime for most of the peaks during the titration in HSQC (heteronuclear single quantum coherence) spectra. Chemical shift perturbations (CSP) measured upon titration in HSQC spectra. Chemical shift perturbations (CSP) measured upon titration in HSQC (heteronuclear single quantum coherence) spectra. Chemical shift perturbations (CSP) measured upon titration in HSQC (heteronuclear single quantum coherence) spectra. Chemical shift perturbations (CSP) measured upon titration in HSQC (heteronuclear single quantum coherence) spectra. Chemical shift perturbations (CSP) measured upon titration in HSQC (heteronuclear single quantum coherence) spectra.

Identification of W-AIM: a tryptophan arch-interacting motif of NVL. Given that NVL167–216 binds the KOW domain of MTR4 similarly to Nop53, we expected the presence of a similar AIM motif. A stretch of amino acids within this segment (NVL residues 185–190) appeared to show significant sequence similarity to the Nop53 AIM sequence (Fig. 3a). SURPRISINGLY IN THIS CONTEXT, THE F186A AND D189R MUTATIONS IN THE NOP53-LIKE STRETCH DID NOT AFFECT NVL167–216 BINDING TO MTR4 IN A PULL-DOWN ASSAY WITH RECOMBINANT PROTEINS (Fig. 3b, compare lanes 5 and 6), indicating that this segment is not a bona-fide AIM. To identify the arch-interacting motif in NVL167–216, we analyzed the sequences from different species in terms of evolutionary conservation. We noticed that vertebrates feature a conserved hydrophobic segment GWFDKTP (residues 172–179, Fig. 3a; Supplementary Fig. 1a). Mutations in this segment (W173A,
I175E) either abolished or impaired NVL167–216 binding to MTR4ΔN in pull-down assays in vitro (Fig. 3b, lanes 7 and 8), suggesting that this stretch functions as an arch-interacting motif, which we coin ‘tryptophan-centered arch - interacting motif’ (W-AIM).

We co-crystallized the NVL W-AIM peptide in complex with a construct of human MTR4 encompassing both the DXE core and the arch domain (residues 70–1042) (Table 1). The overall structure of the helicase is similar to that of the human NRDE-2-MTR4 complex31, but with a notably different conformation of the arch domain. In the NVL-bound crystal structure, the arch domain of MTR4 is an open conformation, with the KOW domain clearly separated from the helicase core (Fig. 3c and Supplementary Fig. 4a). The NVL W-AIM polypeptide chain binds in an extended conformation, similar to that reported for NRDE-2, Nop5336 and Air234 (Supplementary Fig. 4b, 4c). The hydrophobic side chains of Trp173NVL, Phe174NVL and Ile175NVL contact Val766MTR4, Tyr745MTR4 and Val742MTR4 respectively, while Asp176NVL interacts with Arg743MTR4. Consistent with the structural data, mutation of Asp176NVL (D176A mutant) impairs the binding of NVL167–216 to MTR4ΔN in a GST pull down experiment (Supplementary Fig. 5b).

Finally, we tested the effect of the MTR4-interacting residues of NVL in human cells. To this end, we carried out co-immunoprecipitation (co-IP) assays in HeLa cells stably expressing MTR4 with an eGFP ‘localization and affinity purification’ (LAP) tag at the N-terminus and transiently expressing FLAG-tagged full-length NVL constructs (wild type, or with the W173A/I175A mutation or with deletion of the entire hydrophobic segment). Western blotting analysis confirmed that wild type NVL bound MTR4 and that disruption of the 172–180 segment, either by mutation or deletion, impaired the NVL-MTR4 interaction (Fig. 3e, compare lanes 8, 9, 10). Taken together, we conclude that NVL interacts with the KOW domain of MTR4 using a short linear W-AIM sequence. The NVL W-AIM is more hydrophobic than the Nop53 AIM35, rationalizing why it binds MTR4 with an order of magnitude higher affinity than the yeast Nop53-Mtr436 and human Nop53-MTR4 interactions (Supplementary Fig. 3d).

The unstructured region of ZCCHC8 interacts with the MTR4 KOW domain. Identification of the W-AIM sequence in NVL.

Fig. 2 Analysis of MTR4 KOW-NVL complex by NMR and site directed mutagenesis. a Overlay of 1H,15N-HSQC of either MTR4 KOW alone (black) or in complex with a six-fold excess of NVL2 (red). Selected residues experiencing large chemical shift perturbations (CSPs) are labeled. b Plot of CSPs per residue of the MTR4 KOW sequence. The red line marks the threshold of significant CSPs, which are mapped on the model of MTR4 KOW (PDB 6IEH) in panel c. Gaps indicate either prolines or residues that could not be assigned. c MTR4 KOW domain (PDB 6IEH) with labeled secondary structural elements as assigned by NMR and residues showing significant CSPs (displayed as red/green spheres). d Protein co-precipitations by pull down assays. GST-tagged MTR4ΔN (WT or mutants) were incubated with Trx-NVL167–216 before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes 1–4) and 30% of the eluates (lanes 5–8) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue.
motivated us to examine whether other known MTR4 interactors also contain a similar tryptophan-centered motif. One such interactor is ZCCHC8, the scaffolding subunit of the NEXT complex. This modular protein contains a predicted N-terminal coiled-coil domain (residues 40–80), a Zinc-knuckle domain (residues 222–246), a proline-rich domain (residues 287–334) that interacts with RBM7, and a C-terminal domain (CTD) (residues 659–707) (Fig. 4a) that interacts with the DExH core of MTR4 and activates it. The N-terminal portion of ZCCHC8 is expected to contain an additional MTR4-binding site. Within the N-terminal portion, we focused our attention on the linker between the coiled-coil and the Zinc-knuckle domains, as it appeared to contain an NVL-like tryptophan-containing sequence (Supplementary Fig. 6a). We expressed a large portion of this linker (residues 91–211) tagged to maltose binding protein (MBP-ZCCHC891–211) and purified the resulting protein.
Fig. 3 The vertebrate specific W-AIM in NVL is crucial for binding to MTR4 KOW. a Sequence alignment of vertebrate specific insertion regions of representative NVL sequences, Homo sapiens (Hs), Gallus gallus (Gg), Xenopus tropicalis (Xt), highlighting the W-AIM (GWFDIKTP) (red box), and the Nop53-like region (LxFxD) (purple). The sequences were obtained from the UniProt database and aligned using the T-coffee server69. b Protein co-precipitations by pull down assays. GST-tagged MTR4ΔN was incubated with either wild type Trx-NVL167–216 or its indicated mutant before co-precipitation with glutathione sepharose beads. A total of 3% of the input (top) and 30% of the eluates (bottom) were analyzed on 15% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. c Overall structure of MTR4ΔN70–1042. NVL167–216 complex, with the DExH core of MTR4 colored in gray and arch colored in light blue. NVL is colored in orange. d Zoom-in view of the interactions between MTR4 KOW (light blue) domain and NVL (orange). Domains are colored as in Fig. 3a and viewed are 90° rotation along the horizontal and vertical axes with respect to the view in Fig. 3a. Residues discussed in the text are highlighted and labeled. e Cellular co-IP assay. FLAG-tagged wild type NVL construct, or its indicated mutant variant, were transiently expressed in cells stably expressing MTR4-LAP. After precipitation of MTR4-LAP, a total of 0.5% of the input (left) and 8.0% of the eluates (right) were analyzed on 4–12% SDS-PAGE gel followed by western blotting analysis. The primary antibody used is indicated below the panel within the N-terminal region of ZCCHC8. We first interrogated the NVL-like tryptophan-containing patch (GWEIPK, residues 197–202, Fig. 5a). Surprisingly, mutation of Trp198 and Lys202 (W198A/K202E) did not alter the interaction of ZCCHC81–211 with MTR4-ΔN in GST pull down assays (Fig. 5b, lane 7), indicating that this patch of ZCCHC8 is not a bona-fide W-AIM sequence.

To identify the arch-interacting motif in ZCCHC891–211, we took a similar bioinformatics approach as described above for NVL. When analyzing the evolutionary conservation of ZCCHC891–211, we identified a conserved patch centered at the FCLDKLG segment (residues 178–184, Fig. 5a). Mutation of Phe178 and Asp181 (F178A/D181R mutant) in this segment impaired the interaction with MTR4-ΔN (Fig. 5b, lane 6). This cysteine-centered arch-interacting motif (C-AIM) is loosely related to the Nop53 AIM and the NVL W-AIM segments, and is thus expected to bind to the same site of the KOW domain. Consistently, the Arg743 reverse-charge substitution (R743E) in MTR4 also impaired the binding of ZCCHC891–211 (Fig. 5c, lane 8).

ZCCHC8 also contains a non-canonical arch-interacting motif: I-AIM. While attempting to narrow down the KOW binding region of ZCCHC8 further, we observed that N-terminal truncation of ZCCHC891–211 resulted in a near loss of MTR4 binding (Supplementary Fig. 6b). Based on these results, we reasoned that ZCCHC891–211 might harbor an additional MTR4-binding motif that would be predicted to bind to an adjacent surface on the KOW domain (and thus show no resemblance to the previously identified arch-interacting motifs). Using bioinformatics approaches, we identified another conserved segment upstream of C-AIM (IEEF, residues 112–115). Indeed, mutation of Ile112 and Phe115 (I112R/F115R mutant) severely weakened the interaction of MTR4-ΔN - ZCCHC891–211 interaction in GST pull-down assays (Fig. 5b, lane 8). Thus, this segment (which we refer to as an isoleucine-centered arch-interacting motif, or I-AIM) also contributes to MTR4 binding. Next, we mapped the possible MTR4-binding site of I-AIM. The β-barrel of the MTR4 KOW domain is structurally related to that of Tudor domains (Supplementary Fig. 7a), small globular folds that generally present a hydrophobic pocket for binding methylated arginines and lysines44. Although there is little overall sequence similarity, MTR4-KOW features hydrophobic residues at the equivalent position of the substrate–binding residues of Tudor domains (Supplementary Fig. 7b)44,45. In particular, MTR4 Phe677 is evolutionarily conserved (Supplementary Fig. 7b) and is located on a surface adjacent to the Arg743 site where the Nop53 like AIM, W-AIM and C-AIM sequences are recognized (Supplementary Fig. 7c). In line with this site being used for the additional I-AIM sequence of ZCCHC891–211, mutation of Phe677 (F677E) disrupted binding of ZCCHC891–211 (Fig. 5c, lane 7) and did not affect the binding construct for pull-down assays with the versions of GST-tagged MTR4 described above (Fig. 1a). MBP-ZCCHC891–211 co-precipitated with GST-MTR4-ΔN but not with GST-MTR4-ΔNArch (Fig. 4b, lanes 6 and 7). Similar to the results we had obtained for NVL167–216, the pull-down assays showed that MBP-ZCCHC891–211 interacts with GST-MTR4-KOW (Fig. 4b, lane 8). Next, we determined the affinity of the MTR4-KOW-ZCCHC891–211 interaction using microscale thermophoresis (MST) with a fluorescent-tagged version of ZCCHC891–211 (that we had engineered by fusing a YFP to the C-terminus). In this quantitative assay, we measured a dissociation constant (Kd) of ~0.3 µM (Fig. 4c), indicating a higher affinity than that of the NVL167-216 fragment. To corroborate these results, we pre-formed a GST-MTR4-ΔN - Trx-NVL167–216 complex and incubated it with increasing amounts of MBP-ZCCHC891–211 before subjecting the mixtures to GST pull-down assays. The competition assay showed that ZCCHC891–211 could displace NVL167–216 from the pre-formed complex (Fig. 4d, lanes 5–8), suggesting that they interact with the exosome helicase in a mutually exclusive manner.

Identification of C-AIM: a cysteine arch-interacting motif of ZCCHC8. We proceeded to identify the MTR4-binding site

### Table 1 Data collection and refinement statistics

|                     | Mtr4ΔN70–1042 - NVL167–216 |
|---------------------|-----------------------------|
| **Data collection** |                             |
| Space group         | P61                         |
| a, b, c (Å)         | 184.37, 184.37, 90.53       |
| α, β, γ (°)         | 90.0, 90.0, 120.0           |
| Resolution (Å)      | 92.18–3.07                  |
| R_sys or R_merge    | 15.7 (527.6)a               |
| Completeness (%)    | 99.9 (99.5)                 |
| Redundancy          | 16.5                        |
| CC1/2               | 99.9 (62.7)                 |
| **Refinement**      |                             |
| Resolution (Å)      | 60.35–3.07 (3.22–3.07)      |
| No. reflections     | 32,919                      |
| R_work/R_free       | 22.3 / 25.7                 |
| No. atoms           | 7074                        |
| Protein             | 58                          |
| Ligand/ion          |                             |
| Water               | 6                           |
| B-factors           |                             |
| Protein             | 123.5                       |
| Ligand/ion          | 138.03                      |
| Water               | 92.34                       |
| R.m.s. deviations   | 0.003                       |
| Bond lengths (Å)    | 0.478                       |
| Bond angles (°)     |                             |

*a values in parentheses are for highest-resolution shell"
We tested the importance of the I-AIM segment in human cells. HeLa LAP-MTR4 cells were transfected with FLAG-tagged full-length ZCCHC8 constructs (wild type, the I-AIM F115R double mutant or a deletion construct lacking the ZCCHC8 CTD). Western blotting analysis of the resulting co-IPs revealed that disrupting the I-motif by mutation is sufficient to impair the ZCCHC8-MTR4 interaction, whereas deletion of the CTD did not significantly affect complex formation (Fig. 5d, lanes 8, 9, 10). These mutations also did not affect the ZCCHC8-RBM7 interaction (Supplementary Fig. 6c). The CTD of ZCCHC8 has been shown to bind the DExH core of MTR4 and to enhance helicase activity\(^37\). In contrast, we found that neither the ZCCHC8 nor the NVL AIM motifs have a significant effect on the catalytic properties of MTR4, as judged by assessing both ATP hydrolysis and RNA helicase activities (Supplementary Fig. 8). These results suggest a division of labor of the N-terminal and C-terminal ZCCHC8 regions: while the C-terminal domain regulates the activity of the helicase\(^37\), the N-terminal region plays a central scaffolding role in incorporating MTR4 into the NEXT complex.

Discussion

In this study, we show that the human nuclear exosome adaptors NVL and ZCCHC8 bind the MTR4 KOW domain on a surface that is also employed by Nop53\(^36\) and NRDE-231 using distinct arch-interacting motifs\(^35\). The AIDs of Nop53/Utp18/Air2 (LFxφD(x)\(_{1-3}\)/G/P), NVL (GWFDKTP), ZCCHC8 (NFCLKLG), and NRDE-2 (SRFDTKKP) can best be considered as subfamilies of canonical AIDs. With insight, sequences in both NVL and ZCCHC8 that at first glance appeared to resemble known AIDs instead contain individual amino acids that are likely to prevent MTR4 binding. The consensus sequence of canonical bona-fide AIDs can thus be re-defined as \(\omega x Label(Fig. 2d, lane 7) or human Nop53 (Supplementary Fig. 3c, lane 7).

We tested the importance of the I-AIM segment in human cells. HeLa LAP-MTR4 cells were transfected with FLAG-tagged full-length ZCCHC8 constructs (wild type, the I-AIM 1112R/
proteins that contain arch-interacting motifs, as these motifs are difficult to identify due to the degeneracy of their consensus sequence. Furthermore, we found that ZCCHC8 harbors an additional arch-interacting motif that does not conform to the canonical AIM consensus sequences. These findings show how different exosome adaptor proteins have evolved similar mechanisms to recognize MTR4 in a specific and mutually exclusive manner, but can also modulate the affinity and thus selectivity with which they are ultimately recruited to the nuclear exosome.

**Fig. 5** Analysis of the ZCCHC8-MTR4 KOW complex by site directed mutagenesis. a Sequence alignment of the region between the predicted coiled coil domain and the zinc finger of ZCCHC8 from representative metazoan species, *Homo sapiens* (Hs), *Gallus gallus* (Gg), *Xenopus laevis* (Xl), highlighting NVL-like region (red), Nop53-like AIM (C-AIM) and ZCCHC8 specific I-AIM (purple boxes). The sequences were obtained from the UniProt database and aligned using the T-coffee server. b Protein co-precipitations by pull down assays testing ZCCHC8 mutants for their MTR4 binding ability. GST-tagged MTR4 was incubated with either ZCCHC8 WT or mutants before co-precipitation with glutathione sepharose beads. A total of 3% of the input (top) and 30% of the eluates (bottom) were analyzed on 12% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. c Protein co-precipitations by pull down assays testing MTR4 mutants for their ZCCHC8 binding ability. GST-tagged MTR4 (WT or mutant variants) were incubated with MBP-ZCCHC8 before co-precipitation with glutathione sepharose beads. A total of 3% of the input (lanes, 1-4) and 30% of the eluates (5-8) were analyzed on 12% SDS-PAGE gels and visualized by staining with coomassie brilliant blue. d Cellular co-IP assay. FLAG-tagged ZCCHC8 constructs (WT/IF mutant/CTD deletion) were transiently expressed in cells stably expressing MTR4-LAP. After precipitation of MTR4 taking advantage of the LAP tag, a total of 0.5% of the input (left) and 8.0% of the eluates (right) were analyzed on 4-12% SDS-PAGE gel followed by western blotting analysis. The primary antibody used is indicated below the panel.
Methods

Protein expression and purification. Human MTR4 constructs (full-length, MTR4-ΔN (75–1042), MTR4-KOW (645–787) and MTR4-ΔNarch, where residues 598–942 were substituted by 2xGSS linker) were expressed as 6xHis-GST-tagged fusion (cleavable with 3 C protease) proteins in BL21 star (DE3) E. coli cells, grown either in terrific broth (for biochemical studies) or minimal medium supplemented with 15N labelled ammonium chloride and/or 13C labelled glucose (for NMR studies). The proteins were purified using a Ni-nitrilotriacetic acid (NTA) affinity column and a heparin column (GE healthcare) for ion exchange chromatography. When appropriate, the His-GST tag was cleaved at this point by incubation with 3 C protease, followed by removal of the tag with an additional Ni-NTA affinity step. Finally, the protein was subjected to size exclusion chromatography on a Superdex 200 column (GE healthcare) in 50 mM Hepes/NaOH pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP. For NVL, 50 nM of Trx-tagged NVL samples were dialyzed against a buffer containing in 50 mM Hepes/NaOH pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP. The reaction was then initiated by the addition of ATP, MgCl2 and a trap DNA oligo complementary to the FAM labeled RNA (GGCTGTTCTACAGAGGT) to a final concentration of 2 mM, 2 mM and 400 nM respectively. The reactions were quenched after 40 min on ice and adding quenching buffer to a final concentration of 5% (v/v) SDS, 10 mM EDTA, 10% (v/v) glycerol and 0.005% (w/v) xylene cyanol. The sample were analyzed by electrophoresis on a 15% acrylamide–Tris base, boric acid, EDTA (TBE) gel. The fluorescence was imaged using a Typhoon FLA 7000. The oligonucleotides used in the assay were obtained from Ella Biotech GmbH.

NMR spectroscopy. NMR measurements of MTR4-KOW were performed in phosphate buffered saline (10 mM PO4), 137 mM NaCl, 2.7 mM KCl) mixed with 10% (v/v) D2O. Backbone chemical shifts assignments of the KOW region were obtained from two 1H,15N-labelled samples with protein concentration of 700 µM and 500 µM, respectively. HNCA, HNCA(C) and HNCO, HN(CA)CO and 3D 15N-edited NOESY spectra46 were acquired at 298 K on Bruker Avance III spectrometers at field strengths corresponding to 600 and 800 proton Larmor fre- quencies equipped with TCI cryoprobe heads. The characterization of heteronuclear [1H,13C] NOE experiment was performed at 170 µM and a field strength of 500 MHz as described previously47. Protein binding was measured from HSQC experiments containing water-flip-back/WATERGATE48,49 sequences. Titrations with the NVL peptide were carried out at 298 K with a KOW concentration of 53 µM and in presence of 0.5% (w/v) SDS. The 15N-edited NOESY spectra were recorded and processed with Topspin3.5 and analyzed with CCPNMN Analysis 2.450 and Sparky (http://www.cgl.ucsf.edu/home/sparky). The chemical shift perturbations were calculated as CSP (ppm) = [6 (AH)2 + (AN)2]1/2 51.

Crystallization and structure determination. MTR4ΔN (70–1042) was mixed with 2 molar excess of NVL (167–216) in a buffer containing 20 mM Tris pH 7.5, 0.5 mM MgCl2, 5 mM β-mercaptoethanol and 5% (v/v) glycerol. The crystals were grown following a vapour diffusion setup. Initial crystals were obtained in the (A) 2 mM ammonium sulfate, 0.1 mM Tris pH 8.5 condition of STI screen (Molecular dimensions). The best diffracting crystals were obtained at a concentration of 10 mg/ml at 277 K in 0.1 M Tris pH 8, 1.8 M ammonium sulfate. The crystals were cryo-protected with reservoir solution supplemented with 30% glycerol prior flash-freezing in liquid nitrogen. Data were collected at 100 K at PXIII beamline of the Swiss light source (Villigen, Switzerland). Data processing and scaling was performed using XDS/DIALS46,50 within CCP4 suite software51. The crystals belong to the hexagonal spacegroup 169 (P63) containing one molecule (MTR4ΔN, NVL and KOW) in the asymmetric unit and diffract to 0.7 Å resolution. The structure was solved by molecular replacement with Phaser52 within Phenix using the co-ordinates of the DExH core (98-593 and 847-1042) in the asymmetric unit and diffracting to 3.07 Å resolution. NMR spectroscopy measurements of MTR4-KOW were performed in 50 mM MOPS pH 6.5, 150 mM NaCl, 1 mM MgCl2, 2 mM ADP and 1 mM TCEP. The crystallization trials were performed using a vapour diffusion setup. Initial crystals were obtained in the (A) 2 mM ammonium sulfate, 0.1 M Tris pH 8.5 condition of STI screen (Molecular dimensions). The best diffracting crystals were obtained at a concentration of 10 mg/ml at 277 K in 0.1 M Tris pH 8, 1.8 M ammonium sulfate. The crystals were cryo-protected with reservoir solution supplemented with 30% glycerol prior flash-freezing in liquid nitrogen. Data were collected at 100 K at PXIII beamline of the Swiss light source (Villigen, Switzerland). Data processing and scaling was performed using XDS/DIALS46,50 within CCP4 suite software51. The crystals belong to the hexagonal spacegroup 169 (P63) containing one molecule (MTR4ΔN, NVL and KOW) in the asymmetric unit and diffract to 0.7 Å resolution. The structure was solved by molecular replacement with Phaser52 within Phenix using the co-ordinates of the DExH core (98-593 and 847–1042) and KOW domain (645–787) of MTR4 (PDB 8IEH)53 as search models. The model was manually completed in COOT55 and refined using phenix.refine56. The optimal TLS refinement for TLS refinement was obtained by restrained refinement of 41% of the protein backbone dihedral angles in the final model are in Ramachandran favored region. The figures of crystallographic models were prepared using pyMOL (Schrodinger, LLC).

Cell culture and co-immunoprecipitation assays. Human HeLa Kyoto LAP, MTR4-LAP, RBM7-LAP cell lines were generated as outlined by Poser et al.58 and used for example by Lubas et al.7 (RBM7-LAP) and Meola et al.59 (MTR4-LAP). For these cell lines, we performed following standard instructions. Forty-eight hours after transfection the cells were collected and resuspended in extraction buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 0.5 mM MgCl2, 5 mM β-mercaptoethanol and 5% (v/v) glycerol) or 10% FBS. Transfection was carried out using 10 µg of plasmid containing FLAG-tagged NVL or ZCCHC8 constructs, with Lipofectamine 2000 (Thermo Fisher) following the manufacturer’s instructions. Forty-eight hours after transfection the cells were collected and resuspended in extraction buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 0.5 mM MgCl2, 5 mM β-mercaptoethanol and 5% (v/v) glycerol). The lysates were sonicated twice for 5 s at 20 W, and cell debris was removed by centrifugation at 10,000 g for 10 min. Lysates were incubated 1 h with Dynabeads M-270 Epoxy (Invitrogen) coupled to a polyclonal llama anti-GFP antibody. Beads were washed three times in extraction buffer, then incubated 20 min at 25 °C with 100 µM of fluorescently labeled peptide (green (MTR4) or red (NVL)) in the asymmetric unit and diffractions were recorded on 100 µM (green) or 200 µM (red). The mean input (0.5%) and the eluate (8.0%) were analyzed by SDS-PAGE on a 7% reducing gel stained with Coomassie blue. The majority of the protein backbones were analyzed in the final model are in Ramachandran favored region. The figures of crystallographic models were prepared using pyMOL (Schrodinger, LLC).
anti-β-actin (Dilution 1:10000; Sigma, A2228) and anti-RBM7 (Dilution 1:1000; human protein atlas, HPA013993). Anti-mouse and anti-rabbit secondary antibodies coupled to Horseradish Peroxidase (Dako) were used.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

A reporting summary for this Article is available as a Supplementary Information file. NMR backbone chemical shifts of the human MTR4 KOW domain were deposited at the BMRB under accession number 27831. The coordinates and the structure factors have been deposited in the Protein Data Bank with accession code 6GRO. The source data are provided in Supplementary Fig. 10. All data is available from the corresponding author upon reasonable request.

Received: 14 January 2019 Accepted: 7 July 2019
Published online: 29 July 2019

**References**

1. Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. RNA decay machines: the exosome. *Biochim. Biophys. Acta* **1829**, 552–560 (2013).

2. Zinder, J. C. & Lima, C. D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* **31**, 88–100 (2017).

3. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollerey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′–5′ exoribonucleases. *Cell* **91**, 457–466 (1997).

4. Makino, D. L., Halbach, F. & Conti, E. The RNA exosome and proteasome: common principles of degradation control. *Nat. Rev. Mol. Cell Biol.* **14**, 654–660 (2013).

5. Dziembowski, A., Lorentzen, E., Conti, E. & Seraphin, B. A single subunit, Ddx3, is essential for yeast exosome core activity. *Nat. Struct. Mol. Biol.* **14**, 15–22 (2007).

6. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223–1237 (2006).

7. Butler, J. S. & Mitchell, P. Rp6, Rp47 and cofactors of the nuclear exosome. *Adv. Exp. Med. Biol.* **702**, 91–104 (2010).

8. Milligan, L. L. et al. A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. *Cell Biol. Cell. Biol.* **28**, 5464–5457 (2008).

9. Schilders, G., Raimjakers, R., Raats, J. M. & Prouin, G. J. MPP6 is an exosome-associated RNA-binding protein involved in 5′S RNA maturation. *Nucleic Acids Res.* **34**, 6795–6804 (2006).

10. Falk, S., Bonneau, F., Ebert, J., Kogel, A. & Conti, E. Mpp6 incorporation in the nuclear exosome contributes to RNA channeling through the Mtr4 helicase. *Cell Rep.* **20**, 2279–2286 (2017).

11. Schuch, B. et al. The exosome-binding factors Rp6 and Rp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J.* **33**, 2829–2846 (2014).

12. Weick, E. M. et al. Helicase-dependent RNA decay illuminated by a Cryo-EM structure of a human nuclear RNA exosome-MTR4 complex. *Cell* **173**, 1663–1677 e21 (2018).

13. Gerlach, P. et al. Distinct and evolutionary conserved structural features of the human nuclear exosome complex. *Elife* **7**, e38686 (2018).

14. Jackson, R. N. et al. The crystal structure of Mtr4 reveals a novel arch domain required for RNA processing. *EMBO J.* **29**, 2205–2216 (2010).

15. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).

16. Johnson, S. J. & Jackson, R. N. Skl2-like RNA helicase structures: common themes and complex assemblies. *RNA* **10**, 33–43 (2013).

17. Lubas, M. et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Cell Mol. Biol.* **43**, 624–637 (2011).

18. Ogami, K., Chen, Y. & Manley, J. L. RNA surveillance by the nuclear RNA exosome complex. *Mol. Cell* **654**, 6803–6804 (2015).

19. Fujiwara, Y. et al. Structure and function of the N-terminal nucleolin binding domain of human valosin-containing protein-like 2 (NVL2) harboring a nuclear localization signal. *J. Biol. Chem.* **286**, 21732–21741 (2011).

20. Meyer, H., Bug, M. & Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat. Cell Biol.* **14**, 117–123 (2012).

21. Hanzehmann, P. & Schindelin, H. The structural and functional basis of the p97/valosin-containing protein (VCP)-interacting motif (VIM): mutually exclusive binding of cofactors to the N-terminal domain of p97. *J. Biol. Chem.* **286**, 38679–38690 (2011).

22. Buchberger, A., Schindelin, H. & Hanzehmann, P. Control of p97 function by cofactor binding. *FEBS Lett.* **589**, 2578–2589 (2015).

23. Ogami, K. et al. Structure of the RBM7-ZCCHC8 core of the NEXT complex reveals connections to splicing factors. *Nat. Commun.* **7**, 13573 (2016).

24. Cote, J. & Richard, S. Tudor domains bind symmetrical dimethylated arginines. *J. Biol. Chem.* **280**, 28476–28483 (2005).

25. Friberg, A., Oddone, A., Klymenko, T., Müller, J. & Sattler, M. Structure of an atypical Tudor domain in the drosophila polycomblike protein. *Protein Sci.* **19**, 1096–1116 (2010).

26. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158 (1999).

27. Farrow, N. A. et al. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).

28. Grzesiek, S. & Bax, A. The importance of not saturating H2O in protein NMR - application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593–12594 (1993).

29. Pragot, M., Saused, V. & Sklenár, V. Gradient-tailed excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661 (1992).
50. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59, 687–696 (2005).
51. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. J. Appl. Crystallogr. 43, 186–190 (2010).
52. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. Acta Crystallogr. D. Struct. Biol. 74, 85–97 (2018).
53. Potterton, L. et al. CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallogr. D. Struct. Biol. 74, 68–84 (2018).
54. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. Acta Crystallogr. D. Biol. Crystallogr. 66, 486–501 (2010).
56. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).
57. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D. Biol. Crystallogr. 62, 439–450 (2006).
58. Poser, I. et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Methods 5, 409–415 (2008).
59. Di Tommaso, P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res 39, W13–W17 (2011).

Acknowledgements
We would like to thank the Crystallization Facility of MPI Biochemistry and also Vincent D. Maciej for his contributions at the early stages of this project. This study was supported by the Max Planck Gesellschaft, the European Commission (ERC Advanced Investigator Grant 294371) and the Deutsche Forschungsgemeinschaft (SFB646, SFB1035, GRK1721, FOR1680 to E.C., and SFB1035 and GRK1721 to M.S.). This work was supported by the Cluster of Excellence EXC114 (to E.C. and M.S.). T.H.J. was supported by the ERC (grant 339953), the Lundbeck- and the Novo Nordisk-Foundations.

Author contributions
S.F and E.C. initiated the project; M.L. performed the in vitro experiments; D.J. performed the co-IP experiments under the supervision of T.H.J.; M.L. and J.B. performed the crystallography experiments, M.L. and S.F. built, refined and analyzed the structure; A.S. and M.S. collected and analyzed NMR data; L.L. performed preliminary in vitro experiments with NVL under the supervision of S.F.; M.L., S.F., and E.C. prepared the paper.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11339-x.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information: Nature Communications thanks David Tollervey, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.