

Characterization of a fold in TANGO1 evolved from SH3 domains for the export of bulky cargos

Received: 13 January 2022

Accepted: 28 March 2023

Published online: 20 April 2023

 Check for updatesOliver Arnolds^{1,2} & Raphael Stoll¹ ✉

Bulky cargos like procollagens, apolipoproteins, and mucins exceed the size of conventional COPII vesicles. During evolution a process emerged in metazoans, predominantly governed by the TANGO1 protein family, that organizes cargo at the exit sites of the endoplasmic reticulum and facilitates export by the formation of tunnel-like connections between the ER and Golgi. Hitherto, cargo-recognition appeared to be mediated by an SH3-like domain. Based on structural and dynamic data as well as interaction studies from NMR spectroscopy and microscale thermophoresis presented here, we show that the luminal cargo-recognition domain of TANGO1 adopts a new functional fold for which we suggest the term MOTH (MIA, Otoraplin, TALI/TANGO1 homology) domain. These MOTH domains, as well as an evolutionary intermediate found in invertebrates, constitute a distinct domain family that emerged from SH3 domains and acquired the ability to bind collagen.

The evolution of multicellular organisms (metazoa) brought forth the need to secrete bulky cargos to supply the extracellular matrix with the building blocks it requires¹. For this purpose, an intricate export machinery emerged that is predominantly governed by the transport and Golgi organization (TANGO) 1 protein family; these are transmembrane proteins located at the endoplasmic reticulum exit sites (ERES) found in most metazoans^{2,3}. The simultaneous emergence of more complex organisms naturally led to a greater variety of large molecule-complexes that need to be exported⁴. Whereas only TANGO1 is present in invertebrates, vertebrates express different isoforms and a homologue termed TANGO1-like (TALI)^{2,5}.

The TANGO1 protein family is responsible for organization of membranes and sorting of cargo at the ERES. The transmembrane proteins mediate the export of procollagens, apolipoproteins, and mucins, all of which exceed the size of conventional transport vesicles^{5–7}. The formation of tunnel-like conduits between ERES and the Golgi apparatus enables export, facilitated by the cytosolic part of TANGO1 and TALI².

Cargo recognition and binding in all metazoan homologues of TANGO1 is mediated by a domain annotated as SH3-like, which resides

in the lumen of the endoplasmic reticulum³. However, the process of cargo-recognition and binding remains elusive. The export of procollagen in vertebrates appears to depend on the formation of a ternary complex between procollagen, TANGO1's cargo-recognition domain, and the vertebrate-specific collagen-chaperone HSP47^{6,8,9}. Contrarily, how procollagens are localized at the ERES in invertebrates remains unclear to date as they lack HSP47^{2,10}. Furthermore, a mutation in TALI's cargo-recognition domain leads to reduced secretion of apolipoproteins in mice, indicating the dependence of apolipoprotein secretion on the cargo-recognition domain¹¹.

This broad range of different cargos evokes the following questions: Are all vastly different types of cargo recognized by one single domain type or is cargo-specificity achieved in a less promiscuous way? With the luminal SH3-like domain of TANGO1 present in most metazoans, how do invertebrates export procollagen without HSP47?^{2,3,5,12}

SH3 domains are a family of non-catalytic protein-protein interaction modules located in the cytosol that are involved in a plethora of signaling pathways^{13,14}. Typically, these domains adopt the highly conserved fold of a small β -barrel, which consists of five β -strands that

¹Biomolecular Spectroscopy and RUBiospek|NMR, Faculty of Chemistry and Biochemistry, Ruhr University of Bochum, Bochum, Germany. ²Present address: Structural Genomics Consortium, Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. ✉e-mail: raphael.stoll@rub.de

form two perpendicular antiparallel β -sheets, three distinct loop regions (termed RT, nSrc, and distal) as well as a 3_{10} -helix^{13,15}. SH3 domains predominantly mediate protein-protein interactions by recognizing a left-handed polyproline-2 (PPII) helix. Two classes of consensus sequences, +x Φ Px Φ P (class I) and Φ Px Φ Px + (class II) (with P for proline, Φ for a hydrophobic, + for a basic, and x for any residue), interact with a shallow hydrophobic region located between the nSrc and RT loop^{13,16,17}.

Our results presented here show that the characterization of the cargo-recognition domains of the TANGO1 protein family as SH3 or SH3-like domain is in fact misleading, as it suggests a similar mode of operation (a) to SH3 domains and (b) within the domain family itself. We therefore propose that SH3 domains have evolved into a new functional fold present in TANGO1 to export bulky cargos in metazoans, for which we suggest the term MOTH (MIA, Otoraplin, TALI/TANGO1 homology) domain.

Results

TANGO1's cargo-recognition domain adopts a modified SH3-like fold

In order to elucidate the molecular mechanisms that govern the recognition of cargo in the TANGO1 protein family we determined the structure of what hitherto has been termed as an SH3 domain using solution NMR spectroscopy. The sequence of the construct used corresponds to residues 21–131 of human TANGO1 (hsTANGO1(21-131)) and was based on the homology to the sequence of the MIA protein. Our high-precision structure ensemble exhibits backbone and heavy atom RMSDs over all secondary structure elements of $0.29 \pm 0.05 \text{ \AA}$ and $0.64 \pm 0.08 \text{ \AA}$, respectively. (Supplementary Table 1) This structure revealed a typical small β -barrel fold, similar to SH3 domains, consisting of five antiparallel β -strands $\beta 2$ (Y48-A52), $\beta 3$ (D70-L78), $\beta 4$ (V85-V90), $\beta 5$ (T93-P98), and $\beta 6$ (I102-E107). In addition, a 3_{10} -helix was identified between β -strands five and six, formed by residues K99 to L101. These elements, together with the three loop-regions (RT, nSrc, and distal), are also found in SH3 domains^{13,16,17}. Notably, these features are extended by elongated termini that form two additional β -strands $\beta 1$ (H35-C38) and $\beta 7$ (L113-P116), which cover the classical SH3 fold in a lid-like manner. Moreover, the four cysteines form two disulfide bonds, thereby creating an additional loop (termed disulfide loop) between β -strands one and two, as well as tethering the unstructured C-terminus to the tip of the RT loop (Fig. 1a). These supplementary structural elements have already been observed in a similar fashion for the MIA protein^{18,19} and are presumably present in other members of this domain family, i.e., TALI and Otoraplin, as judged from multiple sequence alignments (Fig. 1b)²⁰.

The canonical function of SH3 domains is abolished within the *mia* gene family

The TANGO1 protein family is encoded by genes of the *mia* family (*mia*, *mia2*, *mia3*, and *mial1*), all of which either consist of or carry a homologous domain described as an SH3 domain²⁰. The eponymous MIA protein is an extracellular homologue to the cargo-recognition domains of TANGO1 and TALI, and has already been shown not to interact with classical PPII ligands of SH3 domains¹⁹. Due to the sequential and structural similarities of the cargo-recognition domains from the TANGO1 protein family to SH3 domains, we investigated their capability to interact with class I and II PPII helices that correspond to the recognition sequences of classical SH3 domains. The titration experiments using NMR spectroscopy analysis did not display any substantial chemical shift perturbations (CSPs) of the backbone amide resonances of the domains from human TANGO1 and TALI as well as the MIA protein, indicating no interaction between the domains and peptides (Supplementary Fig. 1). Only Otoraplin exhibited CSPs at a high molar excess of a class II ligand, for which two-dimensional lineshape analysis revealed a low and physiologically probably irrelevant affinity

with a dissociation constant K_D of $1.3 \pm 0.006 \text{ mM}$ and a dissociation rate k_{off} of $3.5 \times 10^5 \pm 3.2 \times 10^4 \text{ s}^{-1}$. However, CSP mapping to the surface of the predicted Otoraplin structure by DeepMind's AlphaFold identified the interaction site to be at the disulfide and distal loop²¹. This is opposite to the side of the protein compared to the interaction site of classical SH3 domains located between the RT and nSrc loops (Supplementary Fig. 2). Moreover, in case of Otoraplin, the interaction appears to be mainly driven by electrostatic forces between the negatively charged patch, comprised of the disulfide and distal loop, and the three arginine residues at the C-terminal end of the class II peptide. This might also explain why significant shifts were not observed for the class I ligand, which contains only a single arginine residue.

In spite of the structural homology to SH3 domains, many crucial residues required for binding of PPII helix ligands are not conserved in any of the four domains encoded by the *mia* gene family^{17,22}. This abolishes the molecular basis for the interaction with PPII helix ligands, in good agreement with the results of the NMR-based titration experiments (Fig. 1b, c).

Structural differences between the cargo-recognition domain in invertebrates and vertebrates

In vertebrates, procollagens are prepared for export at ERES by a ternary complex between TANGO1's cargo-recognition domain, the vertebrate-specific chaperone HSP47, and the procollagen itself^{8,23}. In invertebrates, however, a conundrum emerges as they lack HSP47, yet the luminal cargo-recognition domain, previously annotated as an SH3, is apparently conserved on the domain level throughout metazoans^{3,24}. Yet, on a sequence level, stark differences between invertebrates and vertebrates can be observed (Fig. 2a). Both cysteines that form the highly conserved second disulfide bridge, thereby tethering the domain's C-terminus to the RT loop in vertebrates, are absent in invertebrates (Fig. 2b). Using NMR spectroscopy, assignment of the backbone resonances of the sequence (Fig. 2a) from *Drosophila melanogaster* (dmTANGO1(30-139)) corresponding to hsTANGO1(21-131) provided first structural insights based on the chemical shift index (CSI). The chemical shift of NMR resonances depends on the chemical environment of the observed nucleus and therefore encodes structural information, from which the secondary structure element composition based on the CSI can be derived²⁵. This analysis revealed a topology similar to the human domain, with two additional β -strands in the RT loop, which are also observed for canonical SH3 domains²⁶. Subsequent analysis of the pico- to nanosecond dynamics via the heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOE (hetNOE) showed substantial differences in the dynamic properties (Fig. 2c, d). Firstly, both domains exhibit decreased hetNOEs for both termini, suggesting these to be unstructured, i.e., highly dynamic. However, for dmTANGO1(30-139) this is already observed for the region directly following $\beta 7$. Conversely, the human domain only displays this behavior after the second disulfide bridge, i.e., the last cysteine (Fig. 2c, indicated in orange). Secondly, in the invertebrate domain the RT loop displays fast dynamics on the pico- to nanosecond timescale, while it is completely rigid in the human domain. This is presumably due to the disulfide bridge that connects the RT loop to the C-terminus. In contrast, the nSrc loop and residues between $\beta 6$ and $\beta 7$ displayed decreased hetNOE values for the human domain, indicating fast dynamic structural fluctuations.

Finally, dmTANGO1(30-139) was also tested for its interaction with PPII ligands even if most residues in SH3 domains critical for this interaction are absent in invertebrate domains of TANGO1. As shown for the human domain, an interaction between the domain and a PPII class II ligand could not be observed (Supplementary Fig. 3).

In vertebrates, a C-terminal helix is conserved in TANGO1's cargo-recognition domain

In addition, we compared our experimentally determined structure in solution of hsTANGO1(21-131) with the predicted structure by

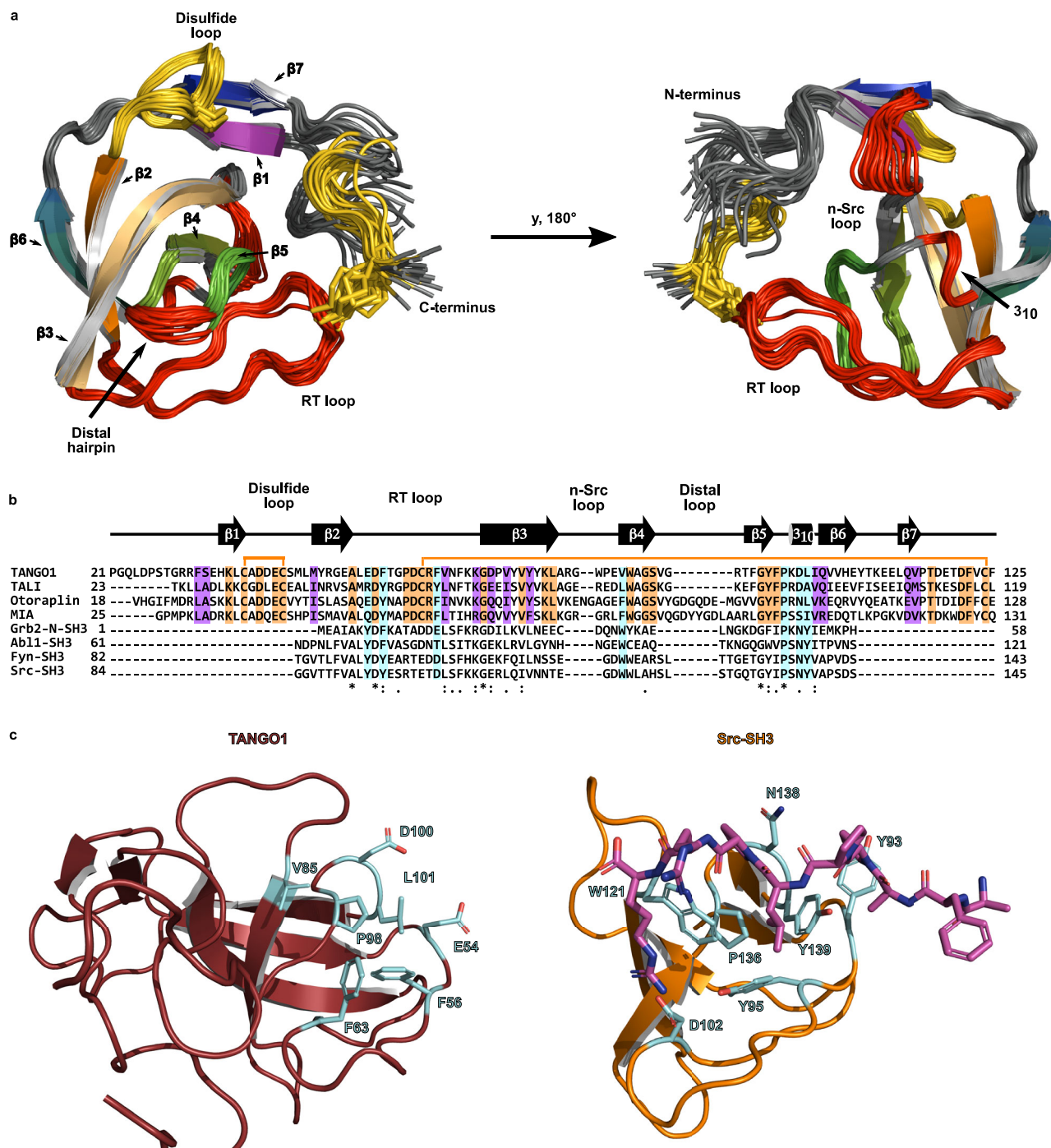


Fig. 1 | TANGO1's cargo-recognition domain adopts a modified small β -barrel fold without retaining functional residues of SH3 domains. **a** Structure ensemble of human TANGO1's cargo-recognition domain solved by solution NMR spectroscopy. Features also present in SH3 domains are highlighted in red, new features created by extended termini in yellow. Disulfide bridges are shown as sticks. (See also Supplementary Table 1) **b** Multiple sequence alignment of the members of TANGO1 protein family and SH3 domains performed with ClustalOmega. Secondary structure elements of TANGO1's domain are displayed above; disulfide bridges are represented by orange lines. Conserved and semi-conserved

residues within the family are labeled in orange and purple, respectively. Residues, conserved within the domain family as well as SH3 domains are marked by an asterisk, semi-conserved residues by a colon. Residues that are critical for binding of PPII helices are highlighted in cyan, according to previous reports^{17,22}. **c** Canonical binding site of PPII helices between the RT and nSrc loop on TANGO1's domain (ruby) compared to the SH3 domain of human tyrosine-protein kinase Src (orange, PDB-ID: IPRM)¹⁷. Residues critical for binding are shown in cyan. The ligand peptide is displayed in magenta. (See also Supplementary Figs. 1 and 2).

AlphaFold. Surprisingly, AlphaFold predicted residues 137–148, which were not included in the original construct for our structure determination, to form an amphipathic α -helix that is in contact with TANGO1's core via hydrophobic residues between the RT and nSrc loop. Aromatic and hydrophobic residues located within this helix that

are in contact with the interface at the RT loop appear to be conserved or at least semi-conserved in vertebrates (Fig. 3a). Comparison of our experimental structure to the one predicted by AlphaFold reveals significant differences for the aromatic network in the domain's core as well as changes to the surface area between the disulfide and

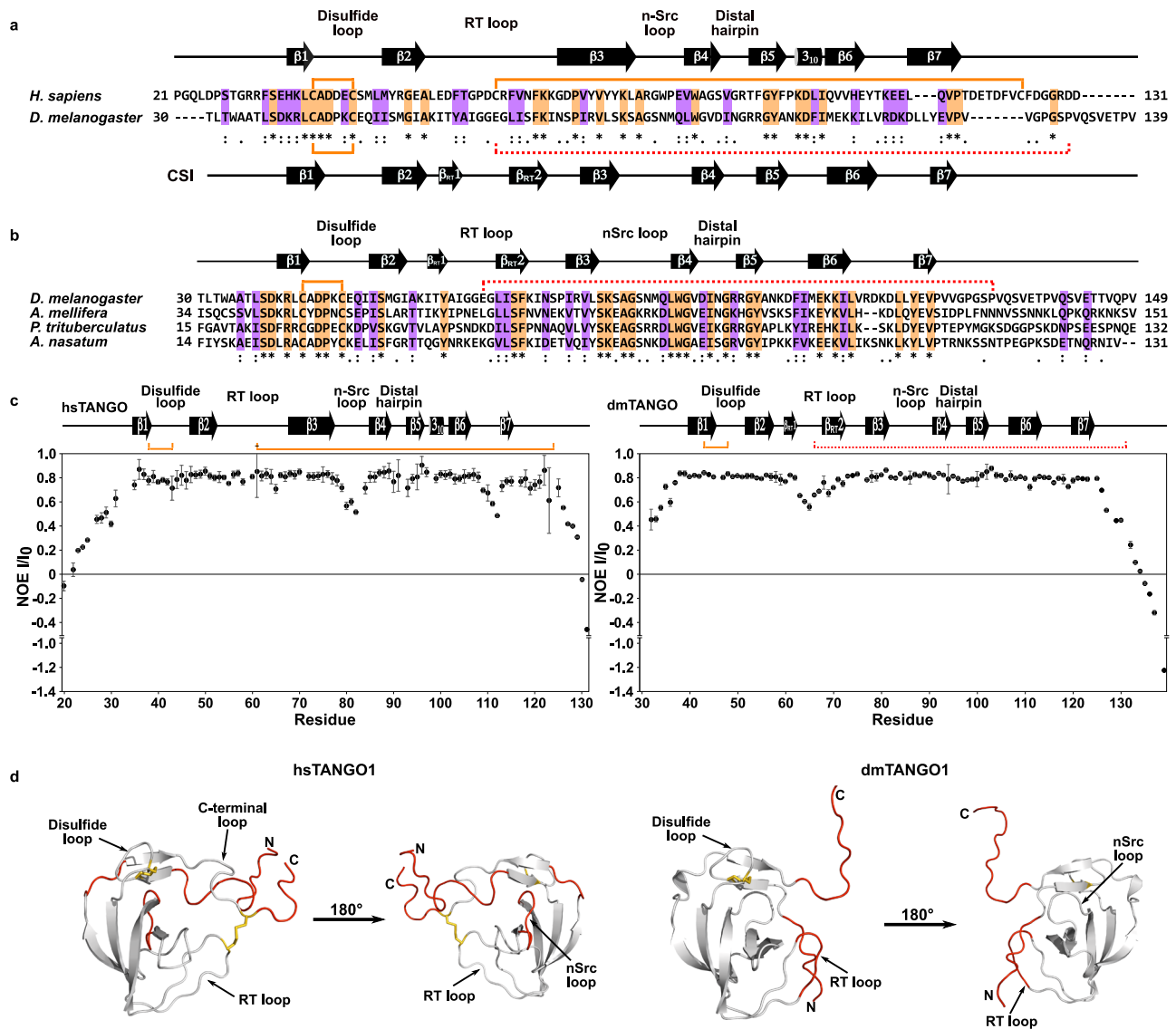


Fig. 2 | The cargo-recognition domain is conserved in invertebrates and distinctly different from the vertebrate domain. **a** Sequence alignment of TANGO1's cargo-recognition domain from *Homo sapiens* and *Drosophila melanogaster* computed with ClustalOmega. Conserved and semi-conserved residues are displayed by asterisks and colons also highlighted in orange and purple, respectively. Disulfide bridges are indicated by orange lines, secondary structure elements above and below, according to the determined structure or the chemical shift index (CSI). The missing second disulfide bridge in invertebrates is indicated by dotted red line (See also Supplementary Fig. 3). **b** Multiple sequence alignment of several

invertebrate organisms. **c** $^{15}\text{N}/\text{[}^1\text{H]}$ hetNOE of TANGO1's cargo-recognition domain from *Homo sapiens* and *Drosophila melanogaster*. Data are presented as mean values \pm standard deviation calculated from all measurements ($n = 3$). Source data are provided as a Source Data file. **d** Regions that display motions on the pico- to nanosecond timescale projected to the structures of the domains from human and a predicted structure of *D. melanogaster* (AF-Q9VMA7-F1 [<https://alphafold.ebi.ac.uk/entry/Q9VMA7>], accessed via <https://alphafold.ebi.ac.uk/>) in red. Disulfide bridges are shown as yellow sticks.

C-terminal loop (Fig. 3e, f). AlphaFold also predicts such a helix for mouse and zebrafish (Fig. 3b), albeit with varying degrees of confidence. Based on this prediction, a synthetic peptide that spans from residues 132–151 of human TANGO1 was titrated to the corresponding cargo-recognition domain (comprising residues 21–131) using solution NMR spectroscopy. Subsequent CSP analysis revealed significant chemical shift differences of the amide resonances surrounding the RT and nSrc loop as well as the 3_{10} -helix, in good agreement with the predicted structure (Fig. 3c, d). Residues displaying shift differences exceeding twice the standard deviation and with a relative surface accessibility of 30% or more were used for two-dimensional lineshape analysis, yielding a dissociation constant of $320.8 \pm 4.1 \mu\text{M}$ with a dissociation rate k_{off} of $1.1 \times 10^3 \pm 29.1 \text{s}^{-1}$ (Supplementary Fig. 4c). This C-terminal helix appears to be conserved for TANGO1 throughout

vertebrates, which, however, is not the case for TALI (Fig. 4a, b). Instead, residues located C-terminally to the second disulfide bridge in TALI seem to be unstructured and not conserved.

Type IV collagen binding is conserved in TANGO1 protein family members

Among the substantially smaller subset of collagens found in invertebrates compared to vertebrates, type IV collagen is one of the fundamental molecules facilitating cell-matrix adhesion^{27,28}. In order to address the cargo-recognition of the TANGO1 protein family in invertebrates, we performed titration series of dmTANGO1(30–139) and collagen IV using microscale thermophoresis (MST). For collagen IV at a constant concentration of 25 nM and increasing concentrations of dmTANGO1(30–139), significant changes in thermophoresis were

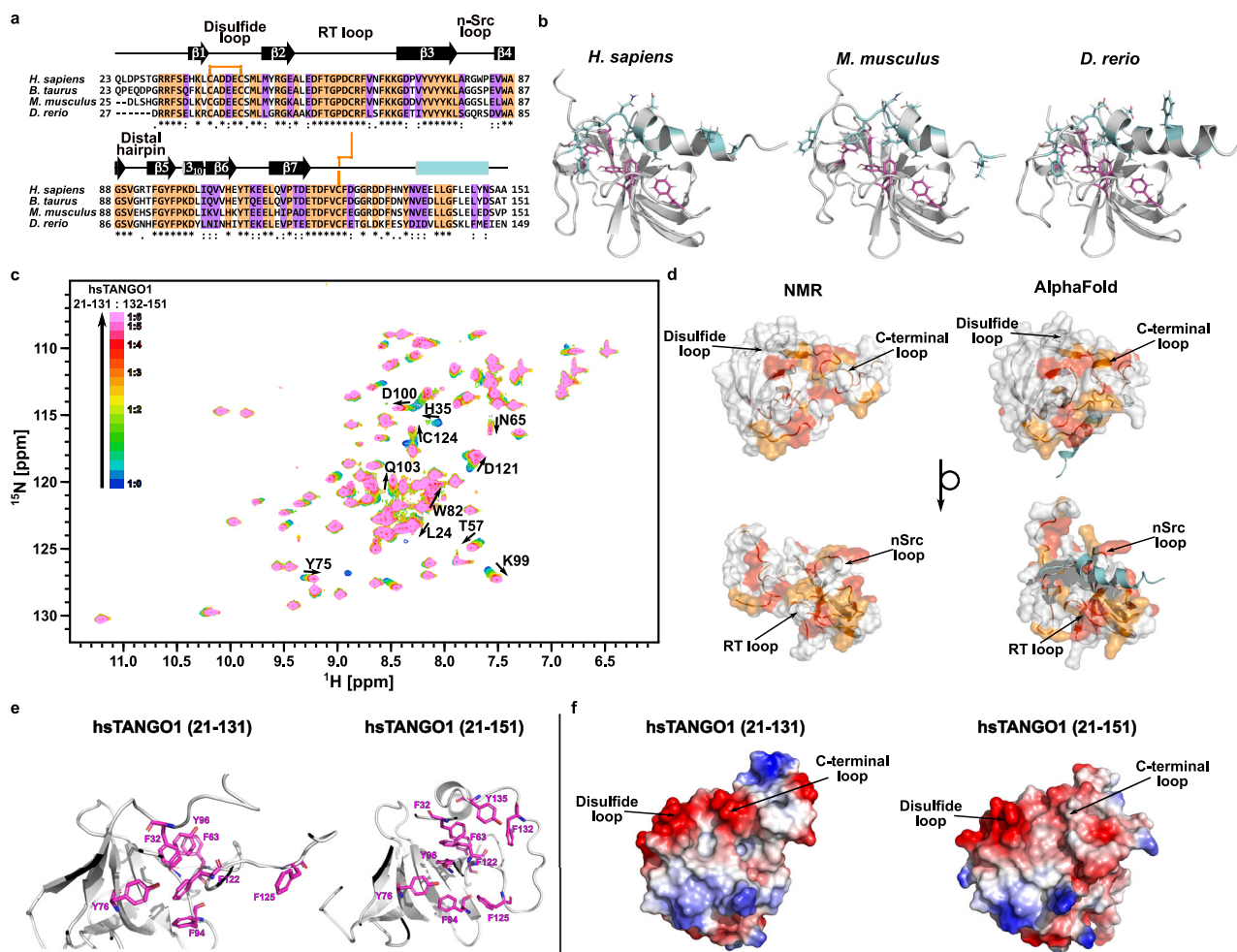


Fig. 3 | The predicted C-terminal helix is conserved and binds to TANGO1's cargo-recognition domain. **a** Multiple sequence alignment of the cargo-recognition domain of TANGO1 from different vertebrate organisms. Conserved and semi-conserved residues are highlighted in orange and purple, respectively. Conserved disulfide bridges are indicated by orange lines. **b** TANGO1's cargo-recognition domain displays a C-terminal helix that is conserved throughout vertebrate organisms, e.g., human, mouse, and zebrafish (AF-Q5JRA6-F1 [<https://alphafold.ebi.ac.uk/entry/Q5JRA6>], AF-Q8BI84-F1 [<https://alphafold.ebi.ac.uk/entry/Q8BI84>], and AF-F1R5N2-F1 [<https://alphafold.ebi.ac.uk/entry/F1R5N2>], respectively. Accessed via <https://alphafold.ebi.ac.uk/>). Conserved aromatic residues within the domain are shown as magenta, conserved residues corresponding to the region of the synthetic peptide as cyan sticks. **c** 2D ^1H ^{15}N HSQC titration spectra. Labeled residues were used for subsequent determination of K_D and k_{off}

values. (See also Supplementary Fig. 4) **d** Residues from CSP analysis exceeding single and double standard deviations based on the average shift differences for all residues projected onto the surface structure of TANGO1's cargo-recognition domain. Experimental NMR structure on the left, AlphaFold prediction including C-terminal helix (cyan) on the right. **e** Aromatic residues (shown in magenta) forming the hydrophobic core of hsTANGO1(21-131) and hsTANGO1(21-151) based on the here determined structure and the prediction by AlphaFold. Changes between both constructs suggest a structural relevance of the C-terminal α -helix. **f** Connolly surface and electrostatic potential (positive displayed in blue, negative in red) of hsTANGO1(21-131) and hsTANGO1(21-151) generated by PyMOL. Differences observed in **(e)** lead to variations in surface charges of the C-terminal loop and the surface area close the disulfide and C-terminal loop.

observed (Fig. 5a), indicating a direct interaction between both molecules with a dissociation constant K_D of $6.9 \pm 3.2 \mu\text{M}$.

To investigate, whether the ability to bind collagen is retained in the vertebrate TANGO1 protein family, all four members were titrated to collagen type IV using MST. Interestingly, significant changes in thermophoresis were observed for hsTANGO1(21-151), TALI(23-123), and Otoraplin, again, indicating an interaction between collagen IV and these proteins (Fig. 5b, c) in the μM -range (Supplementary Table 2) Notably, this was not observed for MIA (Fig. 5c).

Discussion

Based on sequence homology, the luminal cargo-recognition domain of TANGO1 has previously been annotated as an SH3 domain. Notably, the UniProt database assigned the SH3 domain to residues 45–107 (UniProt entry Q5JRA6) but neglected terminal extensions that are conserved throughout the *mia* gene family. Indeed, the

structure presented here revealed a small β -barrel fold at the domain's core, which is, importantly, complemented by terminal elongations. These create two additional β -strands and two disulfide bridges that tether these new features to the classical SH3 fold (Fig. 1a). Whereas previous reports already showed similar structural features for the MIA protein, a sequence alignment and structures predicted by AlphaFold for TALI's cargo-recognition domain and Otoraplin strongly suggest that this is indeed the case for all members of this domain family (Fig. 4c)^{18,19}. Here, we report a hitherto undiscovered α -helix C-terminal of TANGO1's cargo-binding domain that appears to be conserved in vertebrates (Fig. 3a, b), based on structures predicted by AlphaFold. NMR-based titration experiments with a peptide corresponding to the residues forming the C-terminal α -helix in TANGO1's cargo-binding domain indicate binding of the helix at the predicted interaction site. Due to its conservation throughout different vertebrate organisms, we propose this motif to

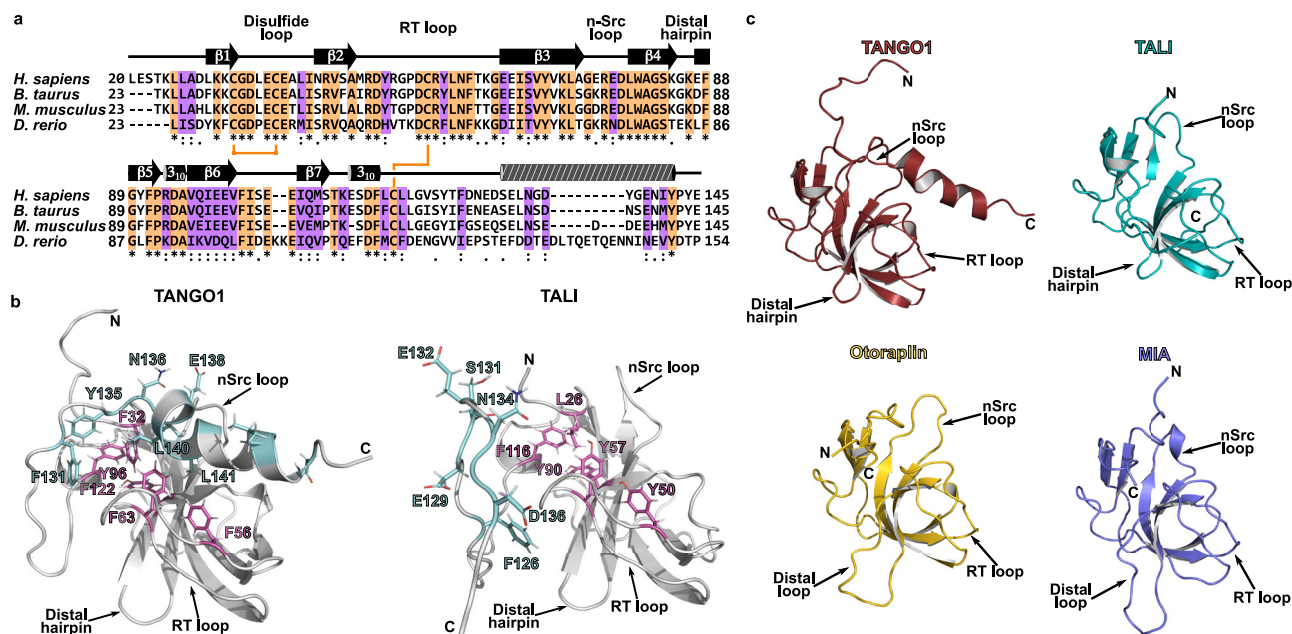


Fig. 4 | A conserved C-terminal helix is unique for vertebrate TANGO1.

a Multiple sequence alignment of the cargo-recognition domain of TALI from different vertebrate organisms. Conserved and semi-conserved residues are highlighted in orange and purple, respectively. Conserved disulfide bridges are indicated by orange lines. C-terminal helix found in TANGO1 is indicated by striped cylinder. **b** Structures of human TANGO1's and TALI's cargo-recognition domain predicted by AlphaFold. Conserved aromatic residues within the domain are shown

as magenta, conserved and semi-conserved residues corresponding to the region of the helix present in TANGO1 as cyan sticks. **c** Domain family encoded by the *mia* gene family. Structures for TANGO1, TALI, and Otoraplin are AlphaFold predictions (AF-Q5JRA6-F1 [<https://alphafold.ebi.ac.uk/entry/Q5JRA6>], AF-Q96PC5-F1 [<https://alphafold.ebi.ac.uk/entry/Q96PC5>], and AF-Q9NRC9-F1 [<https://alphafold.ebi.ac.uk/entry/Q9NRC9>], respectively. Accessed via <https://alphafold.ebi.ac.uk/>). The RCSB-PDB database entry for MIA is 1IIJ [<https://www.rcsb.org/structure/1IIJ>]¹⁸.

be of functional significance, as no other member of the TANGO1 protein family contains this helix (Fig. 4c).

Changes in thermophoresis of type IV collagen upon addition of hsTANGO1(21-151), TALI(23-123), or Otoraplin suggest a newly acquired ability of the *mia* gene family members to bind collagen IV with μM affinity (Supplementary Table 2). These results for human TANGO1 contrast previous reports showing that TANGO1's cargo-binding domain is not able to bind collagen directly⁸. However, a shorter construct of TANGO1's cargo-binding domain was used in these studies, lacking the C-terminal helix. Comparison of the structure determined here with the predicted AlphaFold model revealed changes in the aromatic core and surface of the domain in the area close to the disulfide loop in the presence of the C-terminal α -helix (Fig. 3e, f). This indicates that this helix is necessary for the domain's functional integrity, which is in good agreement with findings of Saito et al. who identified collagen VII as a binding partner for full length TANGO1⁶.

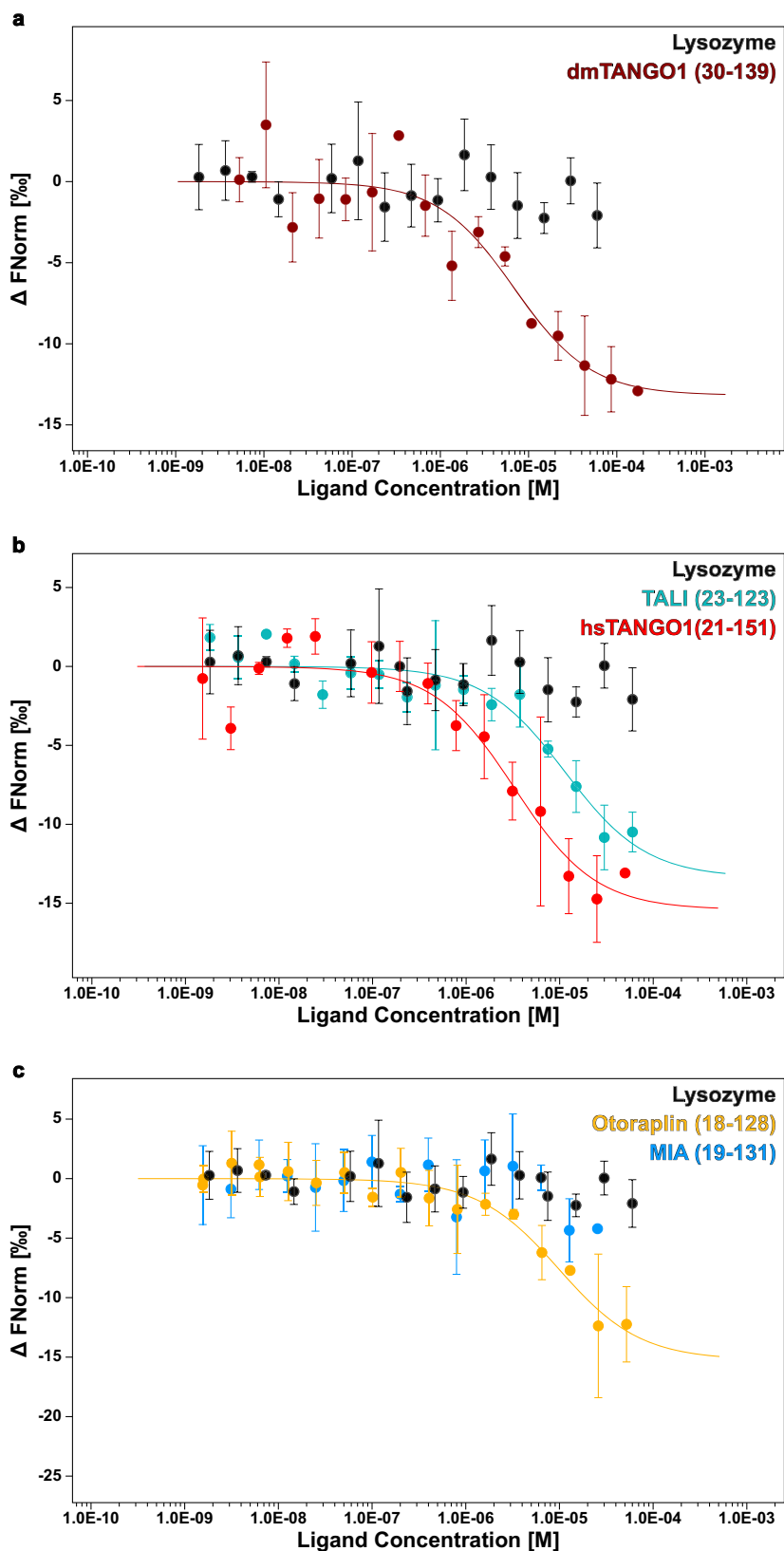
Furthermore, we report that none of these domains retained the ability to interact with PPII helix motifs in a manner that SH3 domains do due to changes in amino acids located at the RT loop, critical for binding PPII-motifs (Fig. 1b)^{13,16,17}. Only Otoraplin displayed very weak affinity towards a class II ligand of classical SH3 domains, but no significant interaction could be observed for a class I PPII motif. Because of this, it seems unlikely to be a specific interaction, presumably mediated by electrostatic interactions between the acidic disulfide loop and arginine side chains of the peptide (Supplementary Fig. 2).

The interaction of SH3 domains with PPII ligands is classically driven by conserved aromatic and hydrophobic residues, most of which are not conserved in all members of the *mia* gene family (Fig. 1b)^{13,16,17}. Hence, we propose that the stable fold of the small β -barrel has been adapted and modified for new physiological tasks in non-cytosolic space. During this evolutionary process, the canonical SH3 function has not been retained, which has been observed for Sm-like domains in a similar fashion²⁹. This motivates us to suggest a new name for this domain family in order to distinguish it from SH3

domains: the MOTH (MIA, Otoraplin, TALI/TANGO1 homology) domain.

Nonetheless, the structural and dynamical differences between the TANGO1-domains of different phyla together with the emergence of four distinct MOTH-domains in vertebrates with diverse expression patterns, mirror the development of a more complex catalogue of bulky cargo in vertebrates. Notably, proteins encoded by the *mia* gene family are involved in several processes involving the export or binding of bulky cargo, and are expressed in a diverse set of tissues. For instance, the extracellular MIA is found in cartilage and displays weak affinity to fibronectin III modules^{30,31}. Similarly, Otoraplin is also found in cartilage, but specifically in the cochlea of murine embryos, without a known interaction partner, prior to this study³²⁻³⁴. TALI, however, is mostly detected in hepatocytes and the small intestine and was shown to be involved in apolipoprotein secretion⁵. Furthermore, silencing TALI's MOTH domain led to decreased levels of cholesterol and triglycerides¹¹. In contrast to TALI, TANGO1 is found ubiquitously, facilitates binding of collagen via HSP47 as well as organization of ERES and export of bulky cargo with several other cytosolic proteins^{6,8,35}. Due to this physiological complexity, questions about the cargo-specificity have been raised.

Here, we demonstrate that the MOTH domains of TALI and TANGO1 are capable of directly binding type IV collagen. This suggests a multiple functionalities of these MOTH domains, as they have been linked previously to many different interactions and export processes, possibly indicating further interaction partners for both domains^{8,11}. In addition, the conserved α -helix located at the C-terminal end of TANGO1's MOTH domain poses a significant structural difference between TALI and TANGO1, and is potentially involved in further interactions or modulation thereof. Notably, this α -helix appears to be non-essential for the interaction between HSP47 and TANGO1's MOTH domain, as previous reports on the interaction with HSP47 used a shortened construct of the MOTH domain that did not contain the residues forming this C-terminal helix⁸. Further investigation of



HSP47's interaction with the full MOTH domain of TANGO1 poses an interesting prospect for future studies.

Otoraplin and MIA, the extracellular MOTH domains, displayed differences in their ability to bind type IV collagen, suggesting functional variabilities. This correlates with the observation that their expression is detected at different times during murine embryogenesis

in cartilaginous tissue and that Otoraplin's expression is mostly limited to the mesenchyme surrounding the otic epithelium^{30,32-34}.

As we show here, an evolutionary intermediate between SH3 domains and the vertebrate MOTH domains can already be observed in invertebrates, such as *D. melanogaster*, which has also already lost the ability to interact with PPII helix ligands (Supplementary Fig. 3).

Fig. 5 | TANGO1 family proteins bind type IV collagen. In all experiments, RED-labeled type IV collagen was kept at a constant concentration of 25 nM while the concentration of the non-labeled binding partner was varied. An MST-on time of 1.5 s was used for analysis and dissociation constants K_D were determined from $n = 3$ ($n = 4$ for lysozyme) independent measurements. Source data are provided as a Source Data file. Data are presented as mean values \pm standard deviation. Lysozyme concentration was varied between 0.0018–60 μ M. No binding was observed for lysozyme and MIA. **a** dmTANGO1(30-139) (maroon) was varied between

0.0053–173 μ M and a K_D of $6.9 \pm 3.2 \mu$ M was derived for the interaction. **b** Intracellular cargo-binding domains of human TANGO1 (red) and TALI (cyan) were varied between 0.0015–50 μ M and 0.0018–60 μ M, respectively. K_D values of $3.3 \pm 1.2 \mu$ M and $11.4 \pm 4.8 \mu$ M were determined. **c** Concentration of extracellular proteins Otoraplin (yellow) and MIA (blue) were varied between 0.0016–52 μ M and 0.0016–51 μ M. As no binding was observed for MIA, a dissociation constant of $9.9 \pm 2.9 \mu$ M could only be derived for Otoraplin.

Whereas the extended termini and first disulfide bridge have already emerged to adopt a similar topology present in the MOTH domains as indicated by the CSI (Fig. 2a), the second disulfide bridge is notably missing. Dynamic data from heteronuclear NOE NMR experiments show that this leaves the residues C-terminal of the last β -strand ($\beta 7$) completely unstructured, as the C-terminus is not tethered to the RT loop.

Furthermore, different regions of hsTANGO1(21-131) and dmTANGO1(30-139) display pronounced dynamic properties (Fig. 2c, d). dmTANGO1(30-139) appears to be rather rigid, with only the RT loop and C-terminus exhibiting movements on the pico- to nanosecond timescale, which is probably facilitated by the missing disulfide bridge. Conversely, these regions are rather rigidified in hsTANGO1, while the nSrc loop and the unstructured part between $\beta 6$ and $\beta 7$ were found to be flexible. Despite these differences, changes in thermophoresis of type IV collagen upon addition of dmTANGO1(30-139) indicate a direct interaction between the invertebrate domain and collagen IV with μ M affinity. The comparable affinity of the vertebrate MOTH domains and conserved structural features like the $\beta 1$ and $\beta 7$ as well as the acidic and conserved disulfide loop suggests that these elements are important for binding collagen IV. The direct binding of dmTANGO1 to type IV collagen implies that invertebrates may not need a collagen-binding protein analogous to the vertebrate-specific HSP47. Noteworthy, the ability to bind type IV collagen with a μ M affinity appears to be conserved between the invertebrate domain and most MOTH domains, except for MIA.

In conclusion, MOTH domains constitute a distinct domain family that emerged from SH3 domains and acquired the ability to bind collagen. Our results also shed light on the foundation of the cargo-recognition of bulky molecules and may ultimately aid drug development for diseases like fibrosis in which regulation of cargo export is impaired.

Methods

Constructs

MOTH domains of human TALI (23-123), Otoraplin (18-128), and MIA (19-131), and TANGO1 (30-139) from *Drosophila melanogaster* (dmTANGO1(30-139)) were expressed from codon-optimized sequences in a modified pQE40 expression vector³⁶ in M15 pRep4 *E. coli* strain. Human TANGO1 (21-131) (hsTANGO1 (21-131)) and TANGO1 (21-151) (hsTANGO1 (21-151)) were expressed from codon-optimized sequences in a modified pET19b vector (provided by Matthias Lübben, PhD of the Department of Biophysics at the Ruhr University of Bochum) in BL21(DE3)RIL *E. coli* strain. The sequences of the used constructs were based on their homology to the sequence of the MIA protein.

Protein expression

Cells were typically grown in custom minimal medium (0.3 mM CaCl₂, 1 mM MgCl₂, 3 ml/l 100 \times BME vitamins, 50.0 mg/l EDTA, 8.3 mg/l FeCl₃ \times 6 H₂O, 0.84 mg/l ZnCl₂, 0.13 mg/l CuCl₂ \times 2 H₂O, 0.1 mg/l CoCl₂ \times 6 H₂O, 0.1 mg/l boric acid, 13.5 μ g/l MnCl₂ \times 4 H₂O, 10 g/l ¹²C-D-glucose, 5 ml/l ¹²C-glycerol, 2 g ¹⁴N-NH₄Cl, 42.3 mM Na₂HPO₄ \times 2 H₂O, and 22 mM KH₂PO₄; pH adjusted to 7.4). For expression, cells were incubated in isotopically-enriched medium. To this end, ¹³C₆-D-glucose (4 g/l) and ¹⁵N-NH₄Cl (2 g/l) were used to substitute their respective

isotopes. Crucially, no glycerol was added to media, if ¹³C-enrichment was required.

Chemically-competent cells were transformed with respective plasmid DNA and then transferred to 200 ml of minimal medium, which was incubated overnight at 37 °C. 2 l of minimal medium were inoculated from the pre-culture to an OD_{600nm} of 0.1. The culture was incubated at 37 °C and to an OD_{600nm} of 0.8–1.0. Next, the cells were harvested by centrifugation at 37 °C and 3000 \times g for 10 min. The resulting pellets were re-suspended in 500 ml isotopically-enriched medium. Expression was induced directly after re-suspension by 1 mM IPTG. The culture was incubated for 21 h at 30 °C. Cells were harvested by centrifugation at 4 °C and 3000 \times g for 10 min, and resulting pellets were re-suspended in 50 mM Tris/HCl, 1 mM EDTA, pH 8.

For MST experiments, BL21(DE3)RIL *E. coli* cells were cultured in LB medium (Luria/Miller; Carl Roth; Cat.-No.: X968.4) analogously to expression in minimal medium. Typically, expression in 2 l LB medium was induced with 1 mM IPTG at OD_{600nm} of 0.6–0.8.

Protein purification from inclusion bodies

Cells were mechanically lysed by micro-fluidization. The resulting homogenates were cleared by centrifugation at 10,000 \times g and 20 °C for 30 min. Pelleted inclusion bodies were re-suspended in 50 mM Tris/HCl, 1 mM EDTA, 1% Triton X-100, pH 8 by vortexing vigorously for 5 min. The suspension was cleared again by centrifugation at 7500 \times g and 20 °C for 10 min and the supernatant discarded. This was repeated until the supernatant was clear. Afterwards, the pellet was re-suspended in 50 mM Tris/HCl, 1 mM EDTA, pH 8 by vortexing and the suspension again cleared by centrifugation. This process was repeated until no more detergent was observed. Inclusion bodies were solubilized at room temperature in 15 ml of 6 M guanidinium chloride, 12.5 mM NaHCO₃, 87.5 mM Na₂CO₃, 0.2 M DTT, pH 10. Following this, the pH of the solution was adjusted to 3 and then cleared by centrifugation at 10,000 \times g and 20 °C for 30 min. The cleared opaque supernatant was filtered using a Filtropur S 0.45 μ m filter (Sarstedt). The buffer of the filtered solution was then exchanged with 3 M guanidinium chloride, 4.7 mM sodium citrate dihydrate, 45.7 mM citric acid, pH 3. Afterwards, the DTT-free solution was dropped very slowly under stirring to refolding buffer (1 M arginine hydrochloride, 50 mM Tris/HCl, 1 mM EDTA, pH 8 and different ratios of oxidized and reduced glutathione, depending on the protein).

Solubilized inclusion bodies were diluted 1:200 in refolding buffer with 0.5 mM of oxidized and 2.5 mM reduced glutathione and incubated at room temperature for 3 days. Subsequently, 2.5 volumes of 50 mM Tris/HCl, 1 mM EDTA, pH 8 were added to the solution and filtered through folded paper filters. Then, the volume was reduced to a volume of 400 ml using the ÄktaFlux system with a 3 kDa MWCO cartridge (GE Healthcare). Afterwards the system was used for buffer exchange with 50 mM HEPES, pH 8, and 1 mM EDTA. The N-terminal His-tag was subsequently cleaved off with 0.6 mg of TEV-protease by incubating for 17 h at 20 °C. TEV-protease and unprocessed hsTANGO1(21-131) were removed by Protino-Kit (Macherey-Nagel). Flow-through and wash fractions were collected, and monomeric protein was purified by size-exclusion chromatography using a HiLoad™ 26/600 Superdex™ 75 pg column (Merck) equilibrated with 25 mM HEPES, 150 mM NaCl and pH 7.4.

Redox ratios during refolding of oxidized:reduced glutathione were 0.5 mM:2.5 mM for dmTANGO1(30-139) and hsTANGO1(21-151), 0.5 mM:0.5 mM for TALI(23-123) and Otoraplin, and 0.5 mM:5.0 mM for MIA. Otoraplin and MIA were incubated with a 1:200 dilution at room temperature for 3 days, dmTANGO1(30-139) with 1:200 at 8 °C for 3 days, TALI(23-123) with 1:40 at 8 °C for 1.5 days, and hsTANGO1(21-151) with 1:200 at 8 °C for 3 days. Purification of TANGO1(21-151)'s, Otoraplin's, and TALI(23-123)'s MOTH domain as well as dmTANGO1(30-139)'s cargo-recognition domain was carried out as described for hsTANGO1(21-131). PBS buffer at pH 7.4 was used for buffer exchange with the *ÄktaFlux* system as well as for equilibration in size-exclusion chromatography.

After refolding MIA's MOTH domain, 1.4 M ammonium sulfate was added and then applied to a hydrophobic interaction chromatography column with a Toyopearl Butyl-650S-substituted matrix (Tosoh Bioscience) using a peristaltic pump. After washing with 50 ml of 50 mM Tris/HCl, pH 7.4, and 1.4 M (NH₄)₂SO₄, the protein was eluted by a step-gradient with a decreasing concentration of ammonium sulfate by 0.2 M per 50 ml step to a final concentration of 0 M. Eluted fractions were analyzed by SDS-PAGE, fractions containing protein were pooled together, and combined fractions were dialyzed twice with a 3 kDa MWCO membrane in 5 l of PBS buffer, pH 7.4. After concentrating, size-exclusion chromatography equilibrated with PBS buffer at pH 7.4 was used to obtain monomeric protein (adapted from ref. 36).

Protein concentrations were determined for hsTANGO1(21-131), hsTANGO1(21-151), dmTANGO1(30-139), Otoraplin, and MIA via absorbance at 280 nm and the Lambert-Beer law using molar extinction coefficients predicted by the Expasy's ProtParam webtool^{37,38}. Concentration of TALI(23-123)'s MOTH domain was determined using the Pierce™ BCA Protein Assay Kit (Thermo Scientific; Cat.-No.: 23250). All monomeric protein solutions were further concentrated, aliquoted, snap-frozen in liquid nitrogen, and stored at -80 °C.

Solution NMR spectroscopy

All spectra (see Supplementary Table 3) were recorded on Bruker DRX 600, AVANCE NEO 600, and AVANCE III HD 700 spectrometers at 298 K. Typically, samples of 1 mM [U-¹⁵N-¹³C]-enriched protein were measured for three- or four-dimensional spectra. For titration analysis, [U-¹⁵N]-enriched protein samples of 0.2 mM were prepared, which is described in detail below. Interscan delay was typically set to 1 s. Mixing time for NOESY experiments was set to 120 ms. Spectra were referenced to the methyl signal of DSS, processed with Topspin 3.6.1 or 4.1.4, and subsequently assigned and analyzed using CcpNmr Analysis 2.4.2³⁹. The MOTH domain of hsTANGO1(21-131) was typically measured in 25 mM HEPES, pH 7.4, 150 mM NaCl, 1% CHAPS, 10% D₂O, 0.02% (w/v) NaN₃, and DSS. All other MOTH domains and dmTANGO1(30-139) were measured in PBS buffer, pH 7.4, 10% D₂O, 0.02% (w/v) NaN₃, and DSS.

For the assignment of backbone and side chain resonances of hsTANGO1(21-131), three-dimensional HNCO, HNcaCO, HNCA, CBCaCO, HNCACB, hCCcoNH, HcccoNH, and ¹H¹⁵N-H-NOESY as well as four-dimensional ¹H¹⁵N-¹H¹³C-NOESY spectra were recorded. For three-dimensional HCCH-COSY, HCCH-TOCSY, and ¹H¹³C-H-NOESY as well as four-dimensional ¹H¹³C-H¹³C-NOESY that do not require an amide proton for detection, the sample was lyophilized and re-solvated in 100 % D₂O. For the assignment of the backbone resonances of TANGO1's cargo-recognition domain from *D. melanogaster*, HNCO, HNcaCO, HNCA, HNcoCACB, and HNCACB spectra were recorded. Backbone resonances from *D. melanogaster* were analyzed using the webtool CSI 3.0 to extract structural information based on the chemical shift index⁴⁰.

For titration experiments, two-dimensional ¹H¹⁵N HSQC spectra were recorded on samples containing only protein as reference and subsequently after adding a small amount of peptide from a high-

concentrated stock solution. All synthetic peptides were resolved in the buffer used for the respective protein, with the pH adjusted to 7.4.

For the interaction of a class I PPII helix peptide, spectra of hsTANGO1(21-131) (215 μM), TALI(23-123) (196 μM), Otoraplin (165 μM), MIA (182 μM), and dmTANGO1(30-139) (160 μM) were each recorded with a 10-fold molar excess of from a stock solution of 16.5 mM p85α(91-104). The interaction with a class II PPII helix peptide (SOS1 (1149-1158)) was investigated by a 10-fold molar excess of peptide to protein for hsTANGO1(21-131) (193 μM), TALI(23-123) (200 μM), and Otoraplin (168 μM), whereas a 24-fold excess was used for MIA (200 μM) from a 20 mM stock solution. To determine the dissociation constant of SOS1 (1149-1158) and Otoraplin, a series of titration spectra with increasing amounts of peptide were recorded with molar ratios of protein to peptide of 1:0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 10.0, 15.0, 20.0, 25.0, 30.0, 40.0, and 50.0.

For hsTANGO1(21-131)'s interaction with the synthetic peptide corresponding to residues 132–151, a sample of 208 μM hsTANGO1(21-131) in 25 mM HEPES, pH 7.4, 150 mM NaCl, 10% D₂O, 0.02% (w/v) NaN₃, and DSS was measured as a reference. The peptide was added corresponding to molar ratios of 1:0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 4.0, 5.0, and 6.0 from a stock solution of 24.1 mM.

Heteronuclear ¹⁵N{¹H} NOE data were recorded as pseudo-three-dimensional spectra of 1 mM [U-¹⁵N]-enriched protein samples as triplicates with an interscan delay of 5 s. Signals were picked in CcpNMR Analysis 2.4.2.

Structure calculation

Distance restraints were extracted from the initial peak lists of all NOESY spectra after complete side chain assignment. Dihedral restraints based on predicted Φ- and Ψ-torsion angles by TALOS+, disulfide bridges between cysteines 38 and 43 as well as 61 and 124, and the *cis*-conformation of proline 83 were set as additional restraints⁴¹. Structures were calculated using a two-step-approach in ARIA 2.3.1⁴². In both steps, the algorithm for torsion angle dynamics was applied for the simulated annealing protocol of the molecular dynamics simulation. Folded conformations were computed from an unstructured, extended strand. The total energy of a calculated structure was used as a criterion to sort the resulting coordinate files. In the first step, ARIA 2.3.1 was utilized to complete the assignment of interresidual proton-proton-contacts using nine iterations with decreasing violation thresholds and number of calculated structures (see Supplementary Table 4). For the assignment process itself, network anchoring and for the distance restraint, a potential a log-harmonic shape was used, which included Bayesian weighting of the distant restraints in order to increase the quality of assignments. After completion of a full calculation protocol, all distance violations >0.5 Å were systematically checked and reassigned, if necessary. Dihedral restraint violations between 5° and 12° were excluded from the calculation, because the uncertainty of the Φ- and Ψ-torsion angles predicted by TALOS+ was reported as 12.6° and 12.3°, respectively⁴¹. In the second step, a structural ensemble was calculated from a fully assigned peak list and subsequently refined in explicit water⁴³. Because the log-harmonic potential was not compatible with the refinement in explicit solvent, the more traditional flat-bottom potential shape was used in this second approach. To this end, a structural ensemble was calculated from an extended strand as a starting point in ARIA 2.3.1 as well, but only a single iteration was applied. An initial structure ensemble from previously calculated structures was used as a reference for the assignment step of ambiguous restraints during the ARIA protocol. 200 structures with the lowest total energy were refined in explicit water. From these, 20 structures without any NOE or dihedral violations were chosen for the final ensemble based on the lowest values in three energy terms with decreasing priority (i.e., total energy, NOE energy, and van-der-Waals energy). Finally, structure quality analysis of

this ensemble was carried out by PROCHECK-NMR^{44,45}. Assignment of secondary structure elements displayed in this paper is based on the analysis of the STRIDE webserver^{46,47}.

Chemical shift perturbation analysis

In order to determine the binding site and dissociation constant, only residues with a shift difference exceeding twice the standard deviation (SD) and displaying a relative surface accessibility of at least 30% according to PyMOL were used for further analysis⁴⁸. Determination of the dissociation constant via TITAN version 1.6 required processing of the spectra using nmrPipe version 9.8^{49,50}. Two-dimensional ¹H-¹⁵N HSQC spectra were processed with an exponential window function for apodization with 4 and 10 Hz of exponential line broadening in the proton and nitrogen dimension, respectively. Signals of residues meeting the aforementioned criteria were fitted to a simple two-state binding model with subsequent bootstrap error analysis.

Microscale thermophoresis

Collagen type IV from human placenta (Sigma-Aldrich; Cat.-No.: C7521) was dissolved in PBS pH 7.2 to a concentration of 2 mg/ml by pipetting and subsequent incubation at 37 °C for 1 h. Concentration was confirmed by absorption at 280 nm using Lambert-Beer law and averaged extinction coefficients for all listed chains from the provider predicted by ExPASy's ProtParam webtool based on Uniprot sequences (P08572, P29400, P53420, Q01955, and Q14031). Collagen type IV was labeled with the Protein Labeling Kit RED-NHS 2nd Generation Kit from NanoTemper Technologies (Cat.-No.: MO-L011). The degree of labeling (DOL) was determined using UV/VIS spectrophotometry at 650 and 280 nm, resulting in a DOL of 0.74. Aliquots of 10 µl of 2.6 µM were snap frozen and stored at -80 °C. Storage buffer of MOTH domains and dmTANGO1(30-139) was exchanged to ligand buffer (10 mM HEPES, pH7.2, 150 mM NaCl, 0.005% Tween-20) using PierceTM protein concentrators (10 K MWCO; ThermoFisher Scientific, Cat.-No.: 88513). Solutions were concentrated to -1/5 of the starting volume, then diluted with ligand buffer. This process was repeated four additional times. Lyophilized lysozyme (Sigma-Aldrich; Cat.-No.: L4919-1G) was dissolved in this buffer. Stock dilutions were diluted to 346 µM (dmTANGO1(30-139)), 120 µM (lysozyme, TALI(23-123)), 104 µM (Otoraplin), 102 µM (MIA), and 100 µM (hsTANGO1(21-151)). Stock solutions were spun for 10 min at 20,000 × g directly prior to use. For MST measurements a dilution series of sixteen sequential 1:1 dilutions with ligand buffer starting with the respective stock solution. Labeled collagen type IV was diluted to 50 nM with 10 mM HEPES, pH7.2, 150 mM NaCl, 0.005% Tween-20, and 30 µM BSA. Finally, the prepared ligand solutions were mixed 1:1 with diluted collagen type IV to final concentrations of 25 nM collagen type IV, 10 mM HEPES, pH7.2, 150 mM NaCl, 0.005% Tween-20, 15 µM BSA, and the respective ligand concentration. These were incubated for 10 min at room temperature and then spun down for 10 min at 20,000 × g. Final solutions were loaded into Standard Monolith Capillaries (NanoTemper Technologies; Cat.-No.: MO-K022). Measurements were carried out with a Monolith NT.115 from NanoTemper Technologies at 25 °C, using 60% LED power and high MST power. MST was recorded for 21 s. Measurements were recorded as biological triplicates (lysozyme (negative control) as quadruplicate).

Multiple sequence alignment

Multiple sequence alignments were generated with ClustalOmega from the EBI tools webservice^{51,52}. Amino acid sequences were taken from UniProt database (<https://www.uniprot.org/>) using entries for vertebrate TANGO1 from *Homo sapiens* (Q5JRA6), *Bos taurus* (Q0VC16), *Mus musculus* (Q8BI84), and *Danio rerio* (FIR5N2). Invertebrate sequences for TANGO1 were used from *Drosophila melanogaster* (Q9VMA7), *Portunus trituberculatus* (A0A5B7CJZ6), and *Armadillidium nasatum* (A0A5N5SML6). Entries for TALI were used from *Homo*

sapiens (Q96PC5), *Bos taurus* (A0A3Q1LM15), *Mus musculus* (Q91ZV0), and *Danio rerio* (A5PLB3). Human sequences for Otoraplin and MIA correspond to database entries Q9NRC9 and Q16674, respectively. Exemplary sequences for SH3 domains were used from human proto-oncogene tyrosine-protein kinase Src (P12931), growth factor receptor-bound protein 2 (P62993), tyrosine-protein kinase ABL1 (P00519), and tyrosine-protein kinase Fyn (P06241). The TANGO1 sequence from *Apis mellifera* was taken from NCBI's gene database (<https://www.ncbi.nlm.nih.gov/>) entry LOC412103.

Quantification and statistical analysis

Standard deviations (SD) for all CSP analyses (Fig. 3 and Supplementary Figs. 2 and 4) were calculated based on the average shift differences observed for all signals of the 2D ¹H-¹⁵N HSQC spectra with Microsoft Excel for Mac (v16.43). Heteronuclear ¹⁵N{¹H} NOEs (Fig. 2) were quantified by calculating the ratio of peak heights of the saturated spectra and the non-saturated reference spectra. Displayed are averaged values ($n = 3$), error bars indicate standard deviation of these averaged values calculated by CcpNMR Analysis 2.4.2. Dissociation constants K_D and rates k_{off} were determined by an iterative fitting procedure to a two-state binding model with subsequent bootstrap error analysis implemented in TITAN version 1.6 (Supplementary Fig. 4)⁵⁰.

For analysis of the MST data, data of at least three independently pipetted measurements ($n = 3$; $n = 4$ for lysozyme) were analyzed (MO.Affinity Analysis software version 2.3, NanoTemper Technologies) using the signal from an MST-on time of 1.5 s. Capillaries displaying aggregation or adsorption were excluded. For K_D determination, the target concentration of 25 nM for collagen type IV was fixed. Displayed are averaged values, and error bars indicate standard deviation.

Reporting summary

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

Data availability

The NMR assignments for TANGO1(30-139) from *Drosophila melanogaster* and human MOTH domain of TANGO1(21-131) are deposited in the Biological Magnetic Resonance Bank (<https://bmr.io/>) under accession codes BMRB 51871 and BMRB 34708, respectively. The atomic coordinates for the structure in solution of human TANGO1(21-131) have been deposited with the Protein Data Bank (<https://www.rcsb.org/>) under accession code 7R3M. Source data for graphs shown in this study are provided with this paper. Source data are provided with this paper.

References

- Özbek, S., Balasubramanian, P. G., Chiquet-Ehrismann, R., Tucker, R. P. & Adams, J. C. The evolution of extracellular matrix. *Mol. Biol. Cell* **21**, 4300–4305 (2010).
- Raote, I. & Malhotra, V. Tunnels for protein export from the endoplasmic reticulum. *Annu. Rev. Biochem.* **90**, 605–630 (2021).
- Feng, Z., Yang, K. & Pastor-pareja, J. C. Tales of the ER-Golgi Frontier: *Drosophila* -Centric Considerations on Tango1 Function. *Front. Cell Dev. Biol.* **8**, 1–9 (2021).
- Huxley-Jones, J. On the origins of the extracellular matrix in vertebrates. *Matrix Biol.* **26**, 2–11 (2007).
- Santos, A. J. M., Nogueira, C., Ortega-Bellido, M. & Malhotra, V. TANGO1 and Mia2/cTAGE5 (TALI) cooperate to export bulky pre-chylomicrons/VLDLs from the endoplasmic reticulum. *J. Cell Biol.* **213**, 343–354 (2016).
- Saito, K. et al. TANGO1 facilitates cargo loading at endoplasmic reticulum exit sites. *Cell* **136**, 891–902 (2009).
- Reynolds, H. M., Zhang, L., Tran, D. T. & Ten Hagen, K. G. Tango1 coordinates the formation of endoplasmic reticulum/ Golgi

- docking sites to mediate secretory granule formation. *J. Biol. Chem.* **294**, 19498–19510 (2019).
8. Ishikawa, Y., Ito, S., Nagata, K., Sakai, L. Y. & Bächinger, H. P. Intracellular mechanisms of molecular recognition and sorting for transport of large extracellular matrix molecules. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1609571113> (2016).
 9. Wilson, D. G. et al. Global defects in collagen secretion in a Mia3/TANGO1 knockout mouse. *J. Cell Biol.* **193**, 935–951 (2011).
 10. Ito, S. & Nagata, K. Roles of the endoplasmic reticulum-resident, collagen-specific molecular chaperone Hsp47 in vertebrate cells and human disease. *J. Biol. Chem.* **294**, 2133–2141 (2019).
 11. Pitman, J. L., Bonnet, D. J., Curtiss, L. K. & Gekakis, N. Reduced cholesterol and triglycerides in mice with a mutation in Mia2, a liver protein that localizes to ER exit sites. *J. Lipid Res.* **52**, 1775–1786 (2011).
 12. Raote, I., Saxena, S., Campelo, F. & Malhotra, V. TANGO1 marshals the early secretory pathway for cargo export. *Biochim. Biophys. Acta - Biomembr.* **1863**, 183700 (2021).
 13. Mayer, B. J. SH3 domains: complexity in moderation. *J. Cell Sci.* **114**, 1253–1263 (2001).
 14. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* **252**, 668–674 (1991).
 15. Kurochkina, N. & Guha, U. SH3 domains: modules of protein-protein interactions. *Biophys. Rev.* **5**, 29–39 (2013).
 16. Yu, H. et al. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945 (1994).
 17. Feng, S., Chen, J. K., Yu, H., Simon, J. A. & Schreiber, S. L. Two binding orientations for peptides to the Src SH3 domain: Development of a general model for SH3-ligand interactions. *Science* **266**, 1241–1247 (1994).
 18. Lougheed, J. C., Holton, J. M., Alber, T., Bazan, J. F. & Handel, T. M. Structure of melanoma inhibitory activity protein, a member of a recently identified family of secreted proteins. *Proc. Natl. Acad. Sci. USA* **98**, 5515–5520 (2001).
 19. Stoll, R. et al. The extracellular human melanoma inhibitory activity (MIA) protein adopts an SH3 domain-like fold. *EMBO J.* **20**, 340–349 (2001).
 20. Arnolds, O. et al. NMR-based Drug development and improvement against malignant melanoma – implications for the MIA protein family. *Curr. Med. Chem.* **24**, 1788–1796 (2017).
 21. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
 22. Simon, J. A. & Schreiber, S. L. Grb2 SH3 binding to peptides from Sos: evaluation of a general model for SH3-ligand interactions. *Chem. Biol.* **2**, 53–60 (1995).
 23. Raote, I., Bellido, M. O., Pirozzi, M., Zhang, C. & Melville, D. TANGO1 assembles into rings around COP II coats at ER exit sites. *J. Cell Biol.* <https://doi.org/10.1083/jcb.201608080> (2017).
 24. Kumar, A., Bhandari, A., Sarde, S. J. & Goswami, C. Ancestry & molecular evolutionary analyses of heat shock protein 47 kDa (HSP47/SERPINH1). *Sci. Rep.* **7**, 1–11 (2017).
 25. Wishart, D. S. & Sykes, B. D. The 13C Chemical-Shift Index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4**, 171–180 (1994).
 26. Politou, A. S., Millevoi, S., Gautel, M., Kolmerer, B. & Pastore, A. SH3 in muscles: solution structure of the SH3 domain from nebulin. *J. Mol. Biol.* **276**, 189–202 (1998).
 27. Brown, N. H. Extracellular matrix in development: insights from mechanisms conserved between invertebrates and vertebrates. *Cold Spring Harb. Perspect. Biol.* **3**, 1–14 (2011).
 28. Hynes, R. O. & Zhao, Q. The evolution of cell adhesion. *J. Cell Biol.* **150**, 89–95 (2000).
 29. Youkharibache, P. et al. The small β -barrel domain: a survey-based structural analysis. *Structure* **27**, 6–26 (2019).
 30. Bosserhoff, A. K. et al. Mouse CD-RAP/MIA gene: structure, chromosomal localization, and expression in cartilage and chondrosarcoma. *Dev. Dyn.* **208**, 516–525 (1997).
 31. Yip, K. T. et al. Human melanoma inhibitory protein binds to the FN12-14 Hep II domain of fibronectin. *Biointerphases* **12**, 02D415 (2017).
 32. Cohen-Salmon, M. et al. Fdp, a new fibrocyte-derived protein related to MIA/CD-RAP, has an in vitro effect on the early differentiation of the inner ear mesenchyme. *J. Biol. Chem.* **275**, 40036–40041 (2000).
 33. Rendtorff, N. D., Frödin, M., Attié-Bitach, T., Vekemans, M. & Tommerup, N. Identification and characterization of an inner ear-expressed human melanoma inhibitory activity (MIA)-like gene (MIAL) with a frequent polymorphism that abolishes translation. *Genomics* **71**, 40–52 (2001).
 34. Robertson, N. G. et al. A novel conserved cochlear gene, OTOR: identification, expression analysis, and chromosomal mapping. *Genomics* **66**, 242–248 (2000).
 35. Bosserhoff, A. K., Moser, M., Schölmerich, J., Buettner, R. & Hellerbrand, C. Specific expression and regulation of the new melanoma inhibitory activity-related gene MIA2 in hepatocytes. *J. Biol. Chem.* **278**, 15225–15231 (2003).
 36. Stoll, R. et al. Letter to the Editor: Sequence-specific 1H, 13C, and 15N assignment of the human melanoma inhibitory activity (MIA) protein. *J. Biomol. NMR* **17**, 87 (2000).
 37. Mäntele, W. & Deniz, E. UV-VIS absorption spectroscopy: Lambert-Beer reloaded. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **173**, 965–968 (2017).
 38. Gasteiger, E. et al. in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–608 (Humana Press, 2005).
 39. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Genet.* **59**, 687–696 (2005).
 40. Hafsa, N. E., Arndt, D. & Wishart, D. S. CSI 3.0: A web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Res.* **43**, W370–W377 (2015).
 41. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
 42. Rieping, W., Bardiaux, B., Bernard, A., Malliavin, T. E. & Nilges, M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–382 (2007).
 43. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins Struct. Funct. Genet.* **50**, 496–506 (2003).
 44. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
 45. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
 46. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins: Struct., Funct. Genet.* **23**, 566–579 (1995).
 47. Heinig, M. & Frishman, D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, 500–502 (2004).
 48. Delano, W. L. The PyMOL Molecular Graphics System. De-Lano Scientific, San Carlos, CA, USA. <http://www.pymol.org> (2002).
 49. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
 50. Waudby, C. A., Ramos, A., Cabrita, L. D. & Christodoulou, J. Two-dimensional NMR lineshape analysis. *Sci. Rep.* <https://doi.org/10.1038/srep24826> (2016).

51. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
52. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

Acknowledgements

We thank Stefanie Pütz for expert technical assistance and assistance in MST measurements, Dr. Xueyin Zhong for stimulating discussions as well as Dr. Ping Zhang for valuable guidance in interpreting MST data. O.A. thanks the RUB Research School^{PLUS} for financial support. We acknowledge support by the Open Access Publication Funds of the Ruhr-Universität Bochum. R.S. gratefully acknowledges support from the DFG (grant nos. INST 213/757-1 FUGG, INST 213/843-1 FUGG, and INST 213/1043-1 FUGG). The Structural Genomics Consortium is a registered charity (no: 1097737) that receives funds from Bayer AG, Boehringer Ingelheim, Bristol Myers Squibb, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer, and Takeda. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 875510. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and Ontario Institute for Cancer Research, Royal Institution for the Advancement of Learning McGill University, Kungliga Tekniska Högskolan, Diamond Light Source Limited. Disclaimer: This communication reflects the views of the authors and the JU is not liable for any use that may be made of the information contained herein.

Author contributions

O.A. and R.S. conceived the research, prepared the figures, and wrote the manuscript. O.A. performed all experiments. O.A. and R.S. recorded and analyzed the NMR data.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-37705-4>.

Correspondence and requests for materials should be addressed to Raphael Stoll.

Peer review information *Nature Communications* thanks Vivek Malhotra, Hartmut Oschkinat, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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/>.

© The Author(s) 2023