

An Evolutionary Perspective on the Origin, Conservation and Binding Partner Acquisition of Tankyrases

Sven T. Sowa ¹, Chiara Bosetti ¹, Albert Galera-Prat ¹, Mark S. Johnson ² and Lari Lehtiö ^{1,*}

¹ Faculty for Biochemistry and Molecular Medicine & Biocenter Oulu, University of Oulu, Oulu 90220, Finland

² Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering and InFLAMES Research Flagship Center, Åbo Akademi University, Turku 20520, Finland

* Correspondence: lari.lehtio@oulu.fi

Supplementary material content

Table S1. Protein sequence identifiers from the analysis of tankyrase distribution and origin. Related to Figure 2.

Figure S1. Phylogenetic analysis of tankyrases. Related to Figure 2.

Figure S2. Predicted structure of the TNKS-like ankyrin-repeat domain from the archaeon *Acidianus manzaensis*. Related to Figure 2.

Figure S3. Predicted structure of the TNKS-like ART domain from the amoeba *Naegleria fowleri*. Related to Figure 2.

Figure S4. Conservation of TNKS ARC domain residues mapped to the structures. Related to Figure 4.

Figure S5. The ARC3 domains in human and early-diverging tankyrases are unlikely to bind TBM peptides. Related to Figure 4.

Figure S6. Size-exclusion chromatography of TNKS1 and *Aq*TNKS constructs. Related to Figure 5.

Figure S7. Distribution of possible tankyrase binding motifs in orthologs of RNF146. Related to Figure 6.

Table S1. Protein sequence identifiers from the analysis of tankyrase distribution and origin. Related to Figure 2. Accession identifiers for the NCBI protein database are shown for each species. ANK = Ankyrin-repeat domain; ART = ADP-ribosyltransferase domain. *Data from transcriptome shotgun assembly; identifier refers to the NCBI nucleotide databank. **Identifier refers to the transcript in the EnsemblGenomes database.

Phylogenetic group	Species	Common name	Accession identifier
Archaea	<i>Acidianus manzaensis</i>	N/A	ANK: WP_148690368.1
Percolozoa	<i>Naegleria fowleri</i>	Brain-eating amoeba	TNKS-like ART: XP_044559609.1
Choanoflagellata	<i>Salpingoeca koevrii</i>	N/A	TNKS*: GGOX01000296.1
	<i>Salpingoeca macrocollata</i>	N/A	TNKS*: GGOT01036632.1
Porifera (Sponges)	<i>Amphimedon queenslandica</i>	N/A	TNKS: XP_019848937.1
Placozoa	<i>Trichoplax adhaerens</i>	N/A	No TNKS ortholog found.
Ctenophora (Comb jellies)	<i>Mnemiopsis leidyi</i>	Sea walnut	TNKS*: ML097525a-RA
Cnidaria (Cnidarians)	<i>Nematostella vectensis</i>	N/A	TNKS: XP_032220531.1
	<i>Hydra vulgaris</i>	N/A	TNKS: XP_047128184.1
Mollusca (Molluscs)	<i>Octopus bimaculoides</i>	California two-spot octopus	TNKS: XP_014782906.1
	<i>Crassostrea gigas</i>	Pacific oyster	TNKS: XP_019923312.2
Nematoda (Roundworms)	<i>Brugia malayi</i>	N/A	TNKS: CDP94254.1
	<i>Trichinella spiralis</i>	Pork worm	TNKS: KRY42746.1
	<i>Caenorhabditis elegans</i>	N/A	No TNKS ortholog found.
Arthropoda (Arthropods)	<i>Drosophila melanogaster</i>	Fruit fly	TNKS: NP_651410.1
	<i>Daphnia pulex</i>	Water flea	TNKS: XP_046457952.1
	<i>Ixodes scapularis</i>	Deer tick	TNKS: XP_029842287.2
Echinodermata	<i>Asterias rubens</i>	Common starfish	TNKS: XP_033632864.1
	<i>Acanthaster planci</i>	Crown-of-thorns starfish	TNKS: XP_022094330.1
Cephalochordata (Lancelets)	<i>Branchiostoma belcheri</i>	Belcher's lancelet	TNKS: XP_019641281.1
	<i>Branchiostoma lanceolatum</i>	European lancelet	TNKS: CAH1257770.1
Tunicata (Tunicates)	<i>Ciona intestinalis</i>	Vase tunicate	TNKS: XP_002121662.3
	<i>Oikopleura dioica</i>	N/A	TNKS: CBY09100.1
Agnatha (Jawless fish)	<i>Petromyzon marinus</i>	Sea lamprey	TNKS: XP_032806710.1
	<i>Eptatretus burgeri</i>	Inshore hagfish	TNKS*: ENSEBUT00000023950.1
Chondrichthyes (Cartilaginous fish)	<i>Callorhynchus milii</i>	Australian ghostshark	TNKS1: XP_042200715.1 TNKS2: XP_007894887.1
	<i>Amblyraja radiata</i>	Thorny skate	TNKS1: XP_032873799.1 TNKS2: XP_032889463.1
	<i>Carcharodon carcharias</i>	Great white shark	TNKS1: XP_041041558.1 TNKS2: XP_041066021.1
Osteichthyes (Bony fish)	<i>Salmo salar</i>	Atlantic salmon	TNKS1: XP_014017013.1 TNKS2: XP_014034742.2
	<i>Protopterus annectens</i>	West African lungfish	TNKS1: XP_043916872.1 TNKS2: XP_043912589.1
	<i>Latimeria chalumnae</i>	Coelacanth	TNKS1: XP_006007641.1 TNKS2: XP_006006371.1
Amphibia (Amphibians)	<i>Xenopus laevis</i>	African clawed frog	TNKS1: XP_018099068.1 TNKS2: XP_018082988.1
	<i>Nanorana parkeri</i>	High Himalaya Frog	TNKS1: XP_018421084.1 TNKS2: XP_018427335.1
Reptilia (Reptiles)	<i>Crocodylus porosus</i>	Saltwater crocodile	TNKS1: XP_019389515.1 TNKS2: XP_019411065.1
	<i>Chelonia mydas</i>	Green sea turtle	TNKS1: XP_037752450.1 TNKS2: XP_007059469.2
Aves (Birds)	<i>Gallus gallus</i>	Red junglefowl	TNKS1: NP_989671.2 TNKS2: NP_989672.2
	<i>Tyto alba</i>	Barn owl	TNKS1: XP_042664360.1 TNKS2: XP_032840041.2
Mammalia (Mammals)	<i>Ornithorhynchus anatinus</i>	Platypus	TNKS1: XP_001508887.3 TNKS2: XP_028915823.1
	<i>Sarcophilus harrisii</i>	Tasmanian devil	TNKS1: XP_031799023.1 TNKS2: XP_031811897.1
	<i>Bos taurus</i>	Cattle	TNKS1: NP_001193089.1 TNKS2: XP_024841840.1
	<i>Mus musculus</i>	House mouse	TNKS1: NP_780300.2

			TNKS2: NP_001157107.1
	<i>Homo sapiens</i>	Modern human	TNKS1: NP_003738.2
			TNKS2: NP_079511.1

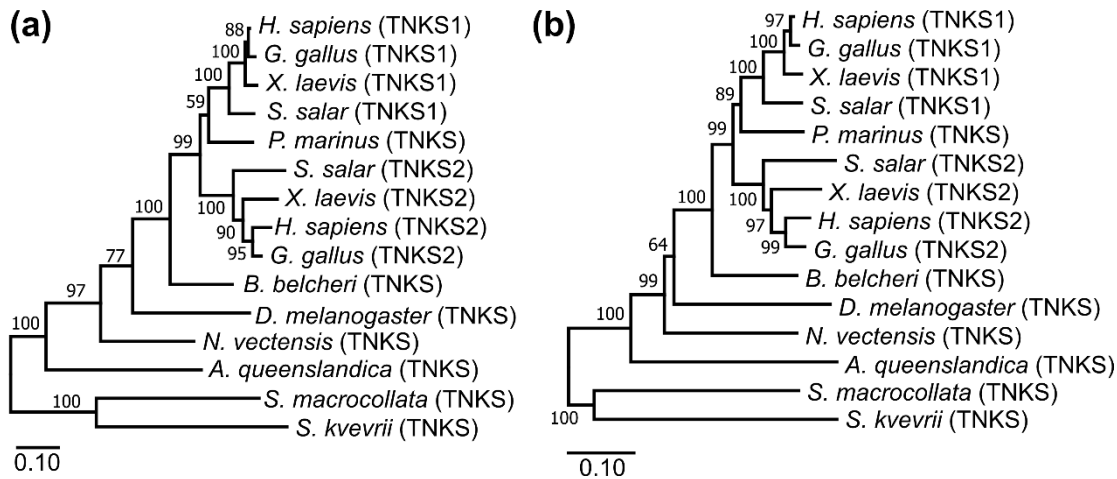


Figure S1. Phylogenetic analysis of tankyrases. Related to Figure 2. Construction of the phylogenetic trees was done with Maximum Likelihood **(a)** or Neighbor Joining **(b)**. Percentages for bootstrap support values are shown for each node. Bootstrap tests were done with 1000 bootstrap replicates. The analysis was performed with truncated sequences starting from the ARC3 domain to match the likely incomplete sequence from *S. macrocollata*.

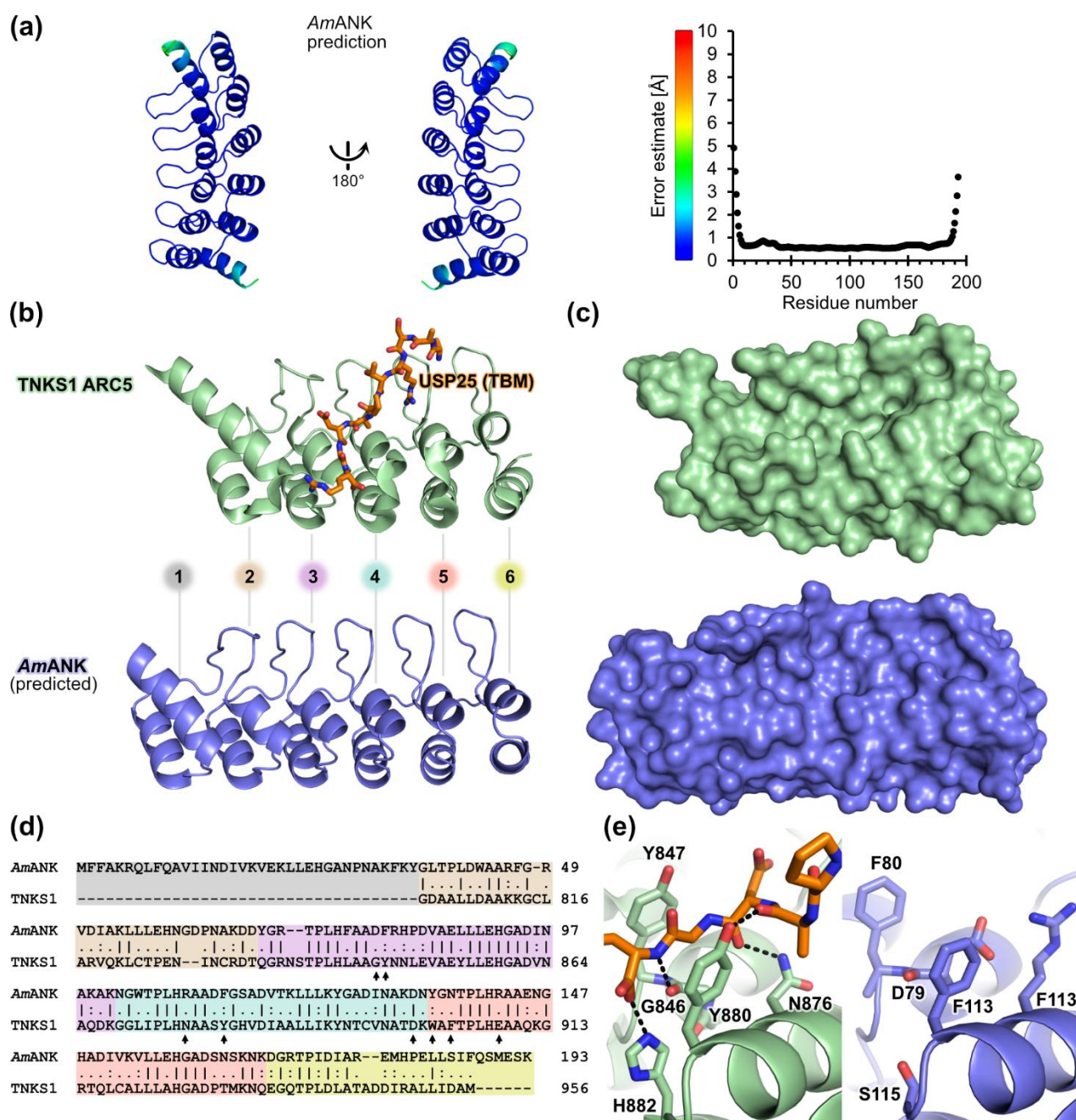


Figure S2. Predicted structure of the TNKS-like ankyrin-repeat domain from the archaeon *Acidianus manzaensis*. Related to Figure 2. (a) The structure of the *A. manzaensis* ankyrin-repeat domain (*AmANK*) was predicted by RoseTTAFold. Coloring of the model corresponds to the estimation of errors in Å. (b) Side-by-side view of the human TNKS1 ARC5 domain in complex with an USP25 TBM peptide (PDB: 5GP7) and the predicted structure of *AmANK*. ARC5 contains 5 and *AmANK* contains 6 ankyrin-repeat units. The ankyrin-repeat units are numbered. (c) Surface representations of TNKS1 ARC5 and *AmANK*. While both structures are similar, the peptide-binding pocket in ARC5 appears more well-defined. (d) Pairwise sequence alignment of *AmANK* and the ARC5 domain from TNKS1. The ankyrin-repeat units are highlighted in different colors. Important residues for TBM binding are indicated by arrows for TNKS1. (e) Close-up view of the aromatic glycine-sandwich sub-site in ARC5 binding the USP25 TBM peptide (left)

and corresponding view of *AmANK* (right). While the two tyrosines are replaced with phenylalanines in *AmANK*, other residues facilitating this interaction differ and would likely not permit this mode of peptide-binding.

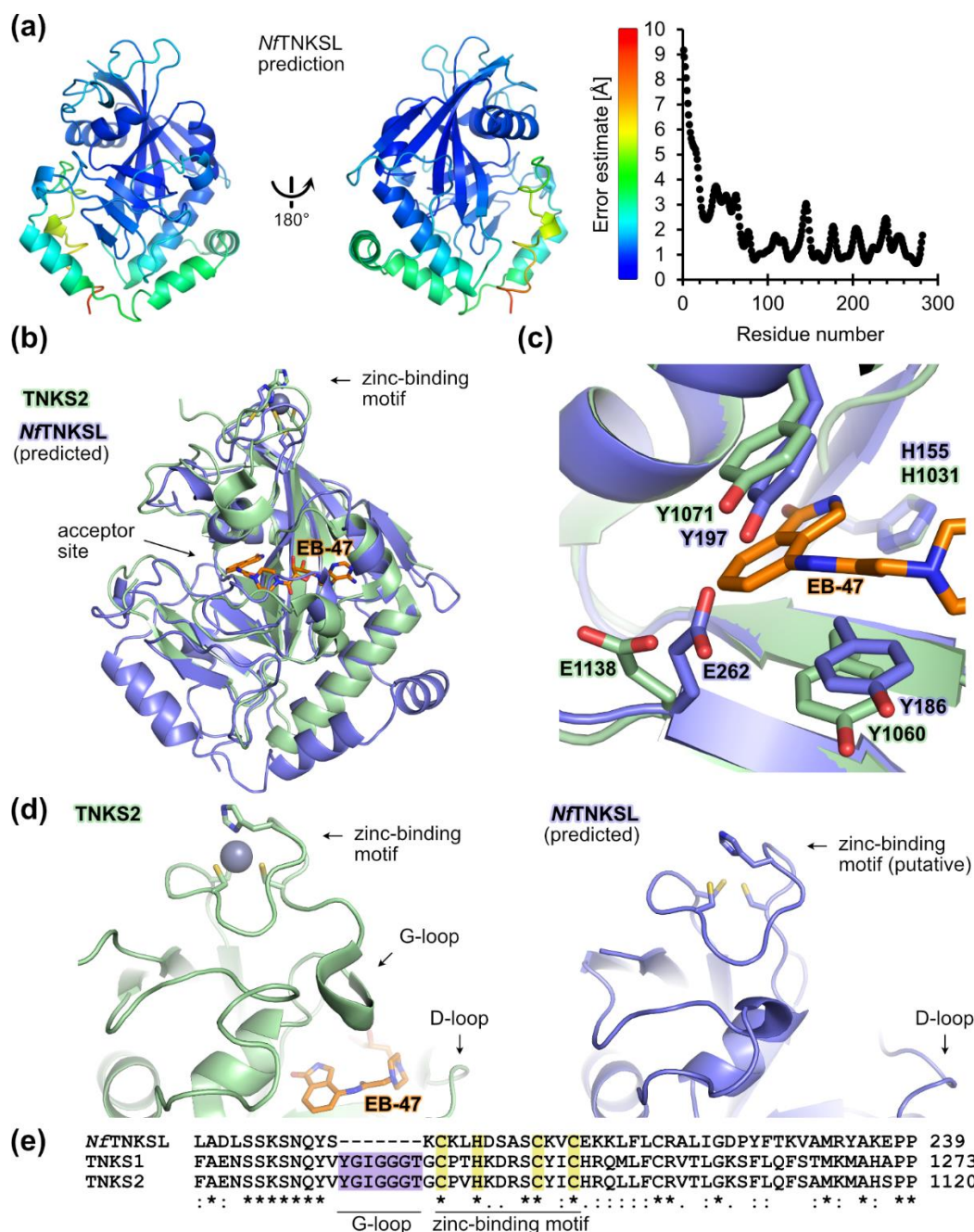


Figure S3. Predicted structure of the TNKS-like ART domain from the amoeba *Naegleria fowleri*. Related to Figure 2. (a) The structure of the *N. fowleri* TNKS-like (*NfTNKSL*) ART domain was predicted by RoseTTAFold. Coloring of the model corresponds to the estimation of errors in Å. (b) Superposition of the predicted *NfTNKSL* structure and the crystal structure of *TNKS2* ART domain in complex with the NAD⁺-mimic EB-47 (PDB: 4BJ9). (c) Close-up view of the nicotinamide sub-pocket for the superimposed structures. (d) Side-by-side comparison of the structures from *TNKS2* and *NfTNKSL*. The unique zinc-binding motif in *TNKS*s was modelled in very similar conformation in the predicted *NfTNKSL* structure, while the Glycine-rich loop (G-loop) present in *TNKS*s is absent in *NfTNKSL*. (e) Multiple sequence alignment of ART domains from human *TNKS1*, human *TNKS2* and *NfTNKSL*. The G-loop region is

colored in violet and not present in *Nf*TNKSL; the ligand residues from the zinc-binding motif are highlighted in yellow.

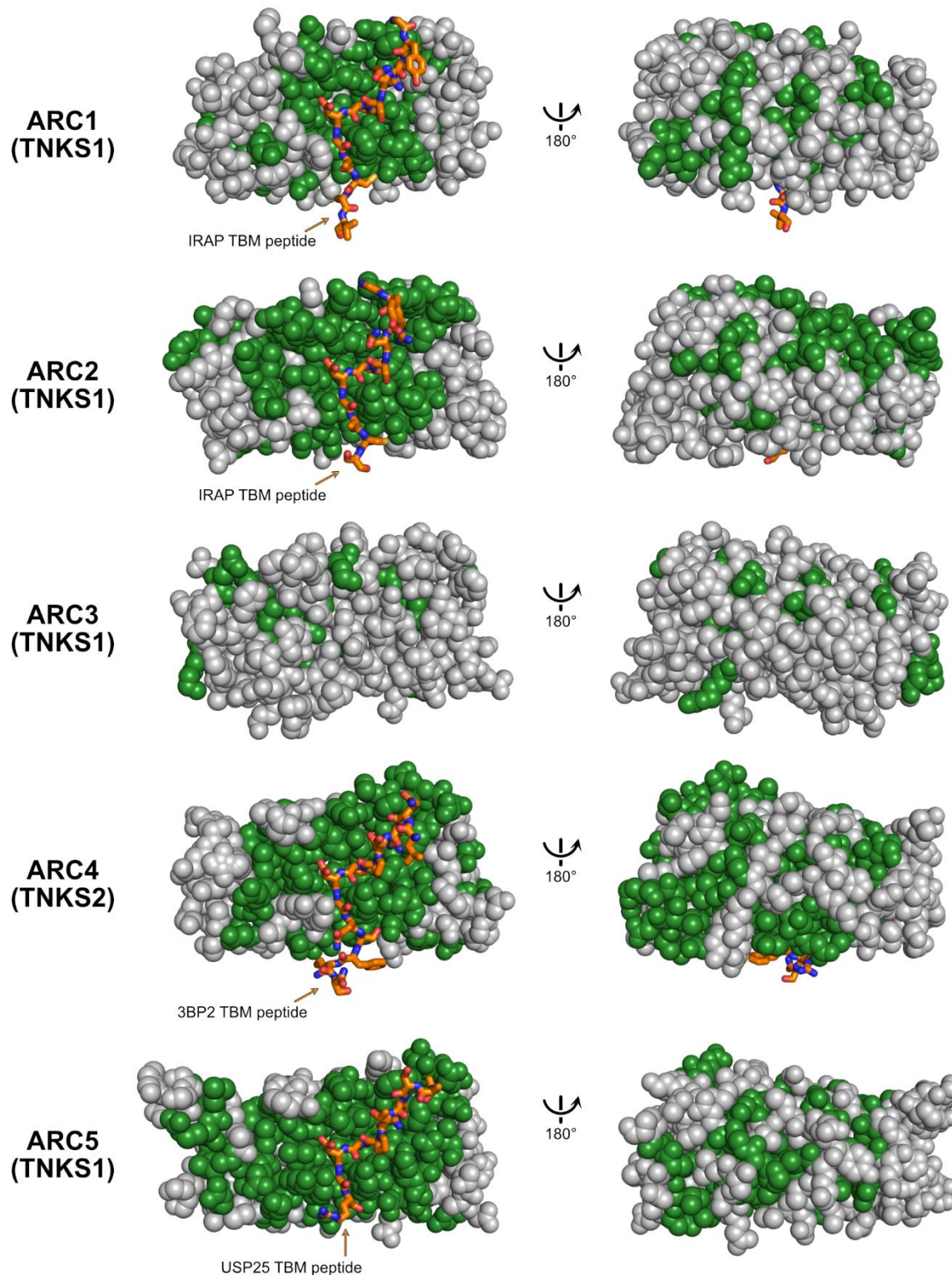


Figure S4. Conservation of TNKS ARC domain residues mapped to the structures. Related to Figure 4. A multiple sequence alignment of TNKSs from 11 metazoan species (as exemplified in Figure 4b) was used to map all identical residues to the structures of TNKS ARC domains, shown in green: TNKS1 ARC1 (PDB:

5JHQ), TNKS1 ARC2 (PDB: 5JHQ), TNKS1 ARC3 (PDB: 5JHQ), TNKS2 ARC4 (PDB: 3TWR), TNKS1 ARC5 (PDB: 5GP7).

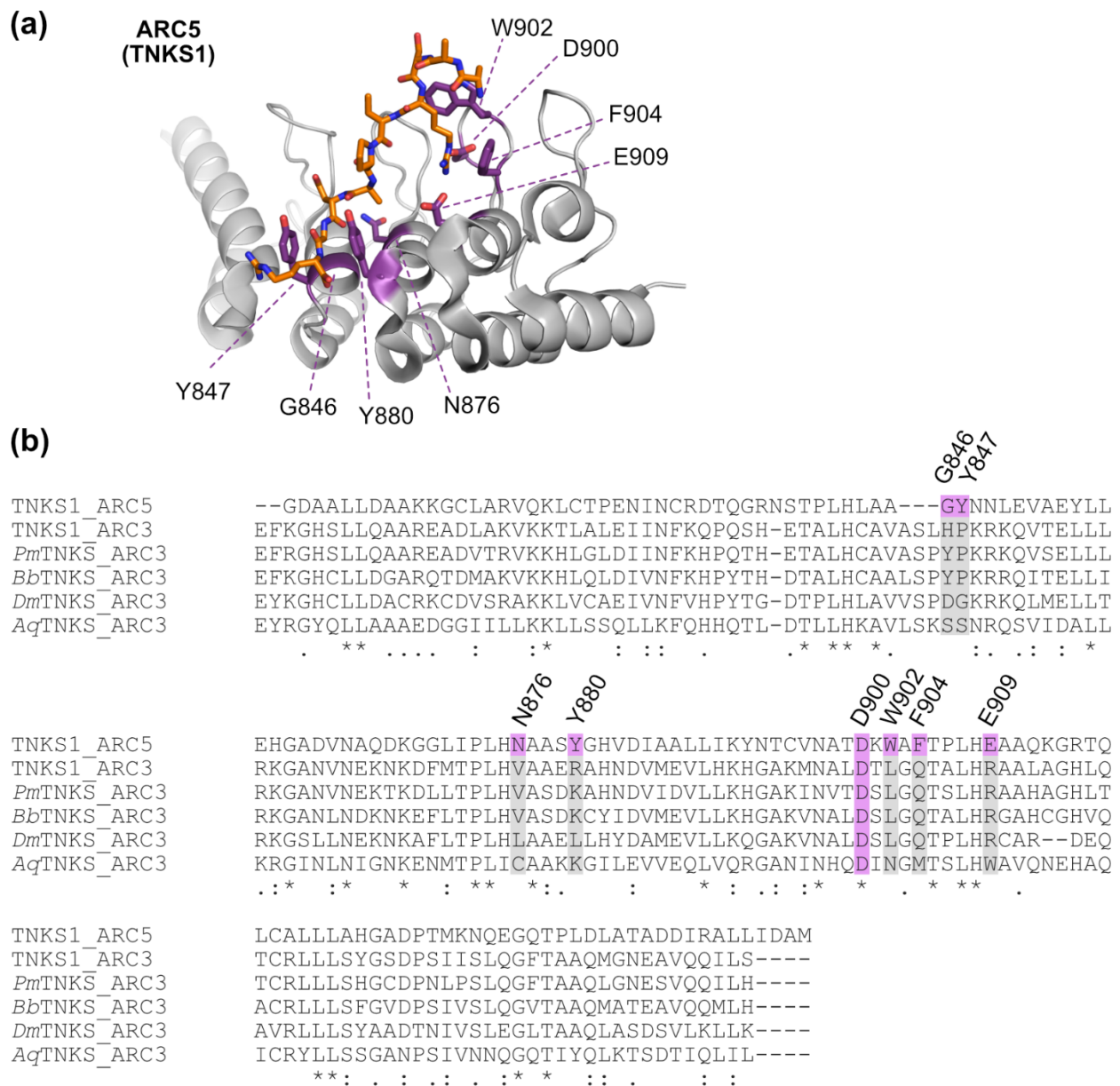


Figure S5. The ARC3 domains in human and early-diverging tankyrases are unlikely to bind TBM peptides. Related to Figure 4. (a) Residues important for binding TBMs by TNKS ARC domains are colored purple and mapped to the structure of ARC5 from human TNKS1 (PDB: 5GP7). The TBM peptide of USP25 is shown in orange. (b) Multiple sequence alignment of ARC5 from human TNKS1 and ARC3 sequences from human TNKS1, *Petromyzon marinus* TNKS (*Pm*TNKS), *Branchiostoma belcheri* TNKS (*Bb*TNKS), *Drosophila melanogaster* TNKS (*Dm*TNKS) and *Amphimedon queenslandica* TNKS (*Am*TNKS). Residues important for binding TBMs are highlighted in purple for TNKS1 ARC5. Non-identical residues in these positions are marked in gray.

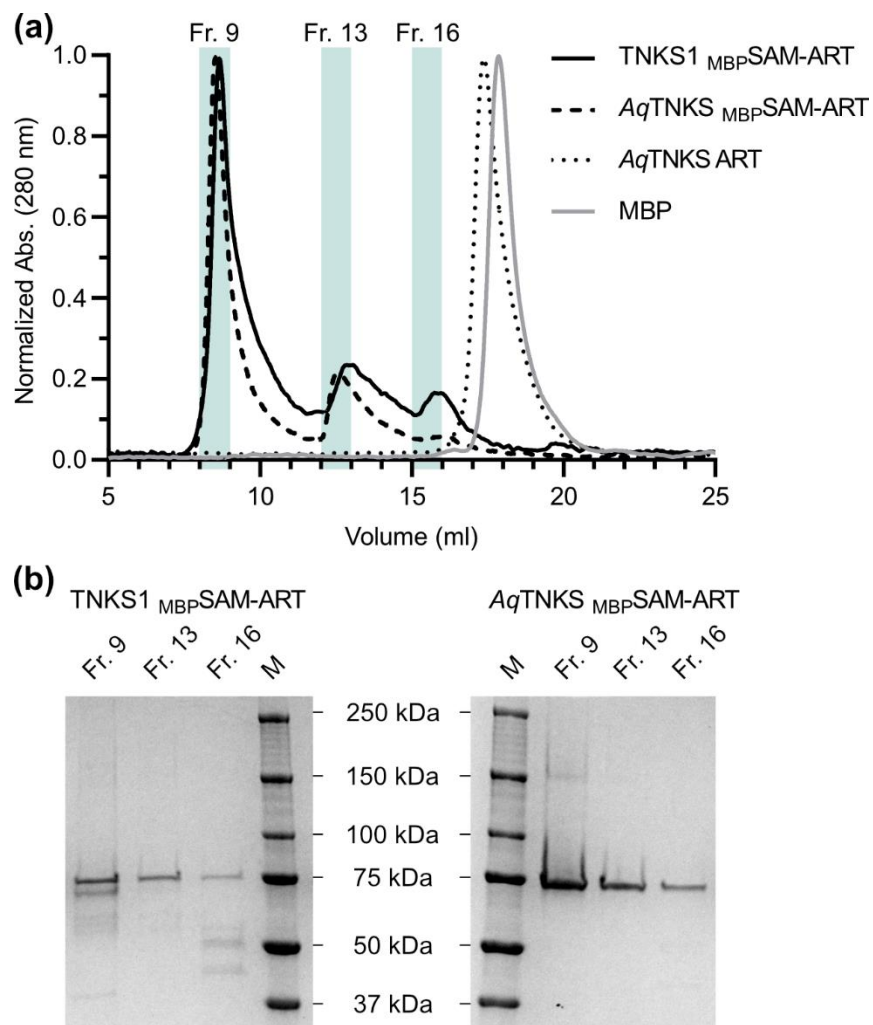


Figure S6. Size-exclusion chromatography of TNKS1 and AqTNKS constructs. Related to Figure 5. (a) Size-exclusion chromatogram of TNKS constructs or MBP. In separate runs, 0.5 mg of each protein was loaded to a Superdex 200 Increase 10/300 GL column at a flowrate of 0.5 ml/min at 4°C. The absorbance at 280 nm was monitored and fractions (Fr.) of 1 ml were taken. **(b)** For the SAM-ART constructs of TNKS1 or AqTNKS, fractions 9, 13 and 16 were analyzed by SDS-PAGE. Precision Plus Protein All Blue (BioRad) was used as protein weight marker (M).

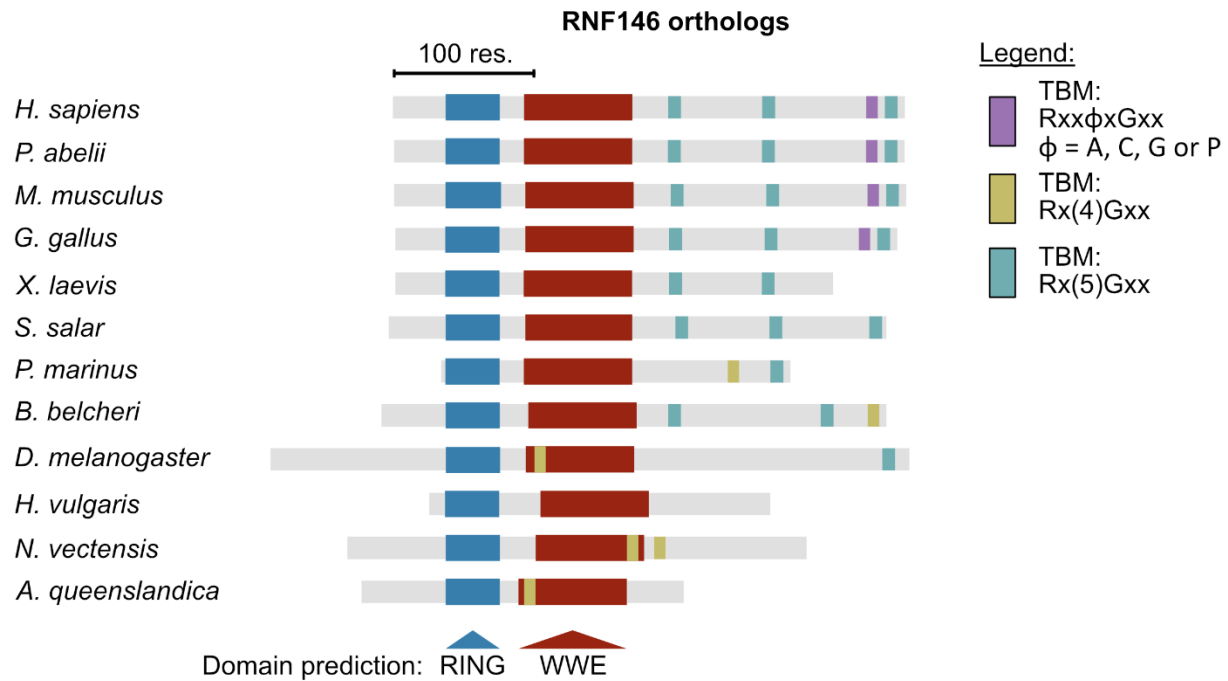


Figure S7. Distribution of possible tankyrase binding motifs in orthologs of RNF146. Related to Figure 6. Orthologs of RNF146 were examined for presence and location of possible TBM sequences. Three configurations Rxx[ACGP]xGxx, Rx(4)Gxx and Rx(5)Gxx were considered as TBM. The prediction of the domain positions for RING finger (blue) and WWE (red) domains was done using InterPro. Parts of the proteins without domain predictions are likely disordered regions and are shown in gray. A scale corresponding to 100 amino acid residues is shown.