

Supplementary Material

Conservation of complex knotting and slipknotting patterns in proteins

J. I. Sulkowska, E. J. Rawdon, K. C. Millett, J.N. Onuchic and A. Stasiak

Guide to the abbreviated names of protein families listed in Table 1 (in the main text) and in tables S1 and S2.

AA-permease - amino acid permeases, ART - arginine ADP-ribosyltransferase family, BCCT - betaine/carnitine/choline transporters, DUF - domains of unknown function, EMG1 - EMG1/NEP1 methyltransferases, Herpes_TK - herpes virus thymidine kinase, MetaL - Metalloenzymes, NCS1 - nucleobase-cation-symport-1 transporters, NPP- Nucleotide pyrophosphatases, PHY - Phytochrome region family, PmbA-TldD - Putative modulators of DNA gyrase, Pyr_Redox_2 - Pyridine nucleotide-disulphide oxidoreductases, RbsD/FucU Ribose/fucose transport protein family, SAM synthetase - S-adenosylmethionine synthetases, SDF - sodium:dicarboxylate symporters, SNF - Sodium: neurotransmitter symporters, SpoU_methylase - SpoU rRNA Methylase family, SPOUT-MTase - Predicted SPOUT methyltransferase protein family, SSF - sodium:solute symporters, Trm 56 - tRNA ribose 2'-O-methyltransferases, tRNA_m1G_MT - tRNA (Guanine-1)-methyltransferases, virC2 - virulence C2 protein family.

Explanation of notation used in Tables S1 and S2. In several cases, the structure determination of the proteins was not complete, resulting in an uncertainty concerning the position of some fragments. For the knotting analysis, we replaced missing fragments with a straight line if the missing peptide could be placed in the vicinity of the straight line without clashing with the rest of the chain (these cases are indicated with *a*). With *b* we indicated the cases, where the missing peptide had to follow an arc to avoid a steric clash with the rest of the determined protein structure An [n] indicates proteins whose knotted pattern are established in this study.

Table S1: Simple knotting patterns (S_3 and S_3) in slipknotted proteins. The pictograms show specific knotting patterns observed for the respective protein families (Pfam classification, [1]). Respective protein names, their PDB code and host species are indicated. Dark green and light green colors indicate right- and left-handed 3_1 knot, respectively.

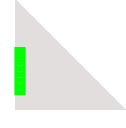
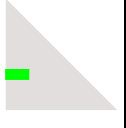
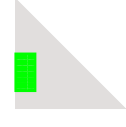
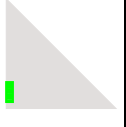
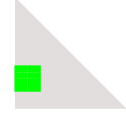
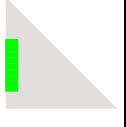

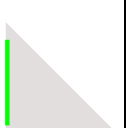
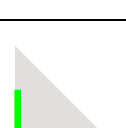


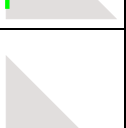
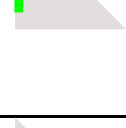
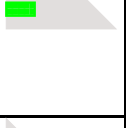
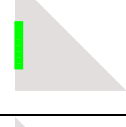
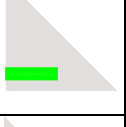
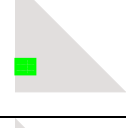
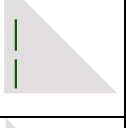
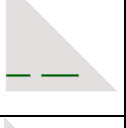

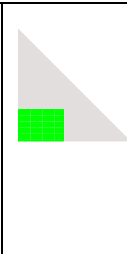

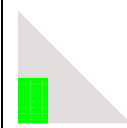
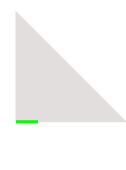
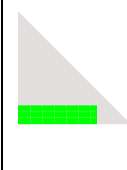
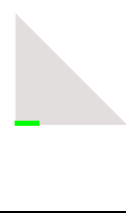
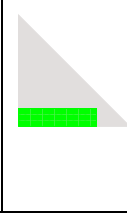
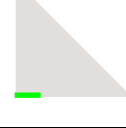
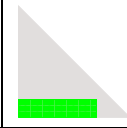

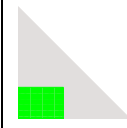

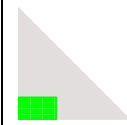
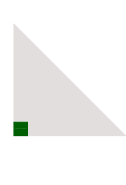
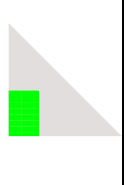

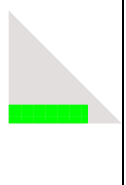
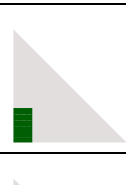
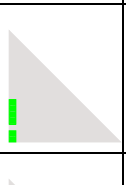
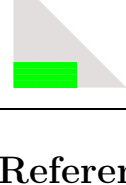
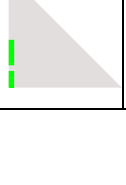
Motif	Family	Protein/PDB	Source	Motif	Family	Protein/PDB	Source
	Alk phosphatase	AP/2x98 [n]	Halobacteriu		Alk sulfatase	PUF/3lxq	Vibrio
	Alk phosphatase	AP/1ew2 AP/1shn	Human Shrimp		DUF	DUF1874/2j85 Protein 114/2x4i ^a [n]	Rudivirus Rudivirus
	Alk phosphatase	AP/1alk	E.coli		NPP	PhnA/1ei6[n] -/2gso[n]	Pseudomonas Xanthomonas
	Alk phosphatase	TAP/2iuc ^a	Antarctic bacterium		Herpes _TK	-/1osn ^a [n] HSV1-T/1p7c EHV4- TK/1p6x	Varicellovirus Simplexvirus Varicellovirus
	Alk phosphatase	SCAP/3a52 [n] -/3e2d [n]	Shewanella Vibrio		MetaL	PMG/1ejj[n] PMG/3igy[n] PMG/2ify[n]	Geobacillus Leishmania Bacillus anthracis
	Alk sulfatase	PALK/2vqr[n] pehA/2w8s[n] YidJ/2qzu[n] YidJ/3b5q[n] Putative arylsulfatase/3ed4	Agrobacterium Burkholderia Bacteroides Bacteroides E.coli		RbsD / FucU	-/2wcu[n] FuCu/3mvk[n] -/3e7n[n] Atu2016/2ob5[n] RbsD/1ogc[n]	Mouse Bifidobacteriu Salmonella Agrobacterium Bacillus subtilis
	Alk sulfatase	AS/1hdh[n]	Pseudomonas		PmbA- TldD	PmbA/3qtd ^a [n] PmbA/1vpb[n] PmbA/1vl4 ^a [n]	Pseudomonas Bacteroides Thermotoga
	Alk sulfatase	LTA/2w8d[n] LTA/2w5q[n]	Bacillus Staphylococcus		ADP- ribo	ART/1gy0[n]	Rat
	Alk sulfatase	ASA/1auk ^a ASB/1fsu ^a [n]	Human Human		Pyr_ redox_2	AIF/1gv4[n]	Mouse
					SDF	GltPh/1xfh	Pyrococcus horikoshii

Table S2: Simple knotting patterns ($K3_1$ and $K3_13_1$) in knotted proteins. The pictograms show specific knotting patterns observed for the respective protein families (Pfam classification, [1]). Respective protein names, their PDB code and host species are indicated. Dark green and light green colors indicate right- and left- hand 3_1 knot, respectively.

Motif	Family	Protein/PDB	Source	Motif	Family	Protein/PDB	Source
	Carbonic anhydrase	CA I/1hcb CA II/1lug CA II/1v9e CA II/1y7w CA III/1z93 CA III/1fj CA IV/2znc CA IV/1znc	Human Human Bovine D. salina Human Rat Human Mouse		SpoU_methylase	tRNAm/1zjr TRNAm/1v2x tRNAm/3e5y putm/3ic6 -/3onp tRNAm/3n4j[n] tRNAm/3l8u[n] TAR/2ha8	Aquifex Thermus Pseudomallei Neisseria Bordetella Yersinia Streptococcus Human
	Carbonic anhydrase	CA V/1keq	Mouse		SpoU_methylase	-/3ilk ^a	Haemophilus
	Carbonic anhydrase	CA VI/3fe4[n] CA VI/1kop	Human N. gonorrhoeae		SpoU_methylase	RNA2o/1ipa -/1gz0 RRNAm/1x7o RNAm/2i6d rRNA/3gyq YibK/1j85	Thermus E.coli Streptomyces Porphyromonas Streptomyces Haemophilus
	Carbonic anhydrase	CA VIII/2w2j[n] CA IX/3iai[n] CA XII/1jd0 CA XIII/3czv[n] CA XIV/1rj6	Human Human Human Human Mouse		Methyltrans_RNA	rmsE/1nxz rmsE/1vhk ^a -/1v6z PUF/3kw2 ^a [n] PUF/2cx8 PUF/1z85 ^b rsmE/2egv	Haemophilus Bacillus Thermus Porphyromonas Thermus Thermotoga Aquifex
	Carbonic anhydrase	PTPRZ /3jxf[n]	Human		EMG1 methyltransferase	NEP1/2v3k ^a [n] NEP1/3o7b[n] NEP1/3bbd	Saccharomyces Archaeoglobus Methanocaldococcus
	Carbonic anhydrase	PTPRZ /3jxh[n] PTPRZ /3jxg[n]	Human Mouse		Trm56	Trm56/2o3a ^a Trm56/2yy8	Archaeoglobus Pyrococcus
	virC2	-/2rh3	Agrobacterium		SPOUT_MTase	YbeA/1ns5 ^a YbeA/1o6d YbeA/1vh0 YbeA/1to0 ^a	E.coli Thermotoga Staphylococcus Bacillus

Motif	Family	Protein/PDB	Source	Motif	Family	Protein/PDB	Source
	Ribbon-helix-helix	PUF/2efv	Methanocaldococcus		tRNA _m1G _MT	TrmD/3ief TrmD/1uaj ^a TrmD/3ky7 ^a TrmD/1p9p ^a TrmD/3knu ^a	Bartonella Haemophilus Staphylococcus E.coli Phagocytophilum
	U2 snRNP	PHF5/2k0a	Saccharomyces		DUF	2168/3dcm[n] 358/2qvv ^a [n] 358/2qmm 358/3aia[n] 171/1k3r	Thermotoga Vibrio cholerae Archaeoglobus Methanocaldococcus Methanothermobacter
	designed protein	-/3mlg	Helicobacter		SAM synthetase	-/2obv	Human
	OTCace	AOTCase/1yh1 AOTCase/1js1	Xanthomonas Bacteroides		SAM synthetase	-/1fug -/1qm4 ^a -/3rv2 ^a [n] -/3iml ^a	E.coli Rat Mycobacterium pseudomallei group

References

- [1] Finn RD, et al (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.* 34: D247-D251.