The dimerization mechanism of the N-terminal domain of spider silk proteins is conserved despite extensive sequence divergence

Médoune Sarr, Kristine Kitoka, Kellie-Ann Walsh-White, Margit Kaldmäe, Rimants Metlāns, Kaspar Tārs, Alessandro Mantese, Dipen Shah, Michael Landreh, Anna Rising, Jan Johansson, Kristaps Jaudzems, Nina Kronqvist

Supporting information

Figure S1 – S8 Table S1 – S2

Figure S1: Sequence alignment of NT domains of different spidroins. The characterized NTs are shown by a red arrow. The alignment was performed using Geneious Alignment and contains the following amino acid sequences of NT: AcSp from Araneus diadematus (Genbank AWK58688.1), AcSp from Araneus ventricosus (Genbank AUH99620), AcSp from Argiope argentata (Genbank AHK09813), AcSp from Argiope aurantia (Genbank AHK09770), AcSp from Argiope trifasciata (Genbank AHK09781), AcSp from Latrodectus geometricus (Genbank AFX83566), AcSp from Latrodectus hesperus (Genbank AFX83557), AcSp from Nephila clavipes (Genbank AWK58691), AcSp from Parasteatoda tepidariorum (Genbank AWK58692), AcSp from Steatoda grossa (Genbank AWK58693), FlSp from Araneus diadematus (GenBank AWK58735), FISp from Argiope argentata (GenBank AWK58731), FlSp from Latrodectus hesperus (GenBank AWK58736), FlSp from Nephila clavipes (GenBank PRD27227), FISp from Parasteatoda tepidariorum (GenBank AWK58739), FISp from Steatoda grossa (GenBank AWK58740), MaSp from Agelenopsis aperta (GenBank ADM14324), MaSp from Argyroneta aquatica (GenBank AVH80563), MaSp from Cybaeus angustiarum (GenBank AVH80556), MaSp from Diguetia canities (GenBank ADM14315), MaSp 1 from Araneus diadematus (GenBank AWK58624), MaSp 1 from Argiope argentata

(GenBank AWK58623), MaSp 1 from Cyrtophora moluccensis (GenBank AGQ04592), MaSp 1 from Euprosthenops australis (GenBank CAJ90517), MaSp 1 from Kukulcania hibernalis (GenBank ADM14314), MaSp 1 from Latrodectus geometricus (GenBank ABY67428), MaSp 1 from Latrodectus hesperus (GenBank ABR68856), MaSp 1 from Latrodectus mactans (GenBank ADO78764), MaSp 1 from Nephila clavipes (GenBank ACF19411), MaSp 1 from Parasteatoda tepidariorum (GenBank AWK58631), MaSp 1 from Steatoda grossa (GenBank AWK58634), MaSp 2 from Araneus diadematus (GenBank AWK58651), MaSp 2 from Argiope argentata (GenBank AWK58646), MaSp 2 from Argiope bruennichi (GenBank AFN54363), MaSp 2 from Argiope trifasciata (GenBank AAZ15371), MaSp 2 from Latrodectus geometricus (GenBank ABY67417), MaSp 2 from Latrodectus hesperus (GenBank ABR68858), MaSp 2 from Nephila clavipes (GenBank AWK58654), MaSp 2 from Nephila inaurata madagascariensis (GenBank AAZ15322), MiSp from Araneus diadematus (GenBank AWK58672), MiSp from Araneus ventricosus (GenBank AFV31615), MiSp from Argiope argentata (GenBank AWK58671), MiSp from Latrodectus geometricus (GenBank ARA91186), MiSp from Latrodectus hesperus (GenBank ARA91182), MiSp from Latrodectus tredecimguttatus (GenBank ARA91189), MiSp from Metepeira grandiosa (GenBank ADM14328), MiSp from Nephila clavipes (GenBank AWK58680), MiSp from Parasteatoda tepidariorum (GenBank ARA91217), MiSp from Steatoda grossa (GenBank AWK58683), MiSp from Uloborus diversus (GenBank ADM14326), PiSp from Araneus diadematus (GenBank AWK58658), PiSp from Argiope argentata (GenBank AQR58363), PiSp from Argyroneta aquatica (GenBank AVH80558), PiSp from Latrodectus hesperus (GenBank AWK58659), PiSp from Nephila clavipes (GenBank PRD25616), PiSp from Steatoda grossa (GenBank AWK58661), TuSp from Agelenopsis aperta (GenBank ADM14330), TuSp from Araneus ventricosus (GenBank ASO67373), TuSp from Argiope argentata (GenBank ATW75951), TuSp from Argiope bruennichi (GenBank BAE86855), TuSp from Latrodectus hesperus (GenBank ABD24296), TuSp from Nephila antipodiana (GenBank ACI23395), TuSp from Nephila clavata (GenBank BAE54451), TuSp from Nephila clavipes (GenBank PRD35275), TuSp from Steatoda grossa (GenBank AWK58644).

1	1	0 2 ₀ 3 ₀ 4	10 50 60	70	80	90 100	110 120	130 140 150
AcSp [A.diadematus]	FSRSP	ANPAKASSLMNCLITKIASSNVLPQQEKE	D LESIMDTLMSAINGASAKGKTSAA	QLQAINMAVASSLA	EIVVAEDVGN	NQASMAVKTQALSGALE(OCFQAVMGTVDRKFIK	INDLISMFARQAATEANEI
AcSp [A.ventricosus]	FSTSP	ANPAKASSLMNCLLTKIASSNVLPQQEKE	D <mark>LESIMDTLMSAIKGASAKGKSS</mark> GA	QLQAINMAVASSLA	EIVVAEDVGI	NQASMAVKTQALSGALE(QCFQAVMGRVDRKFIN	INDLISMFARQAATESNEI
AcSp [A.argentata]	LSRSP	ANPAKAGSLMNCLMSRIASSNVLPQQDKE	DLESIMDTLMSAIKGASAKGKSSAA	OLQAINMAVASSLA	EIVVAEDAGN	VQASIAVKTQALTGALGQ	OCFQAVMGTVDRKFIN	INDLITMFAKEAASESNEI
AcSp [A.aurantia]	LSRSP	ANPAKAGSLMNCLINRISSSNVLPOODKE	DLESIMDTLMSAITGASAKGKSSAA	OLOAINMAVASSLA	ELVVAEDAGI	NOASIAVKTOALSGALO	CFOAVMGTVDRKFIN	INDLITMFAREAASESNEI
AcSp [A.trifasciata]	ISRSP	ANPAKAGSLMNCLINRIASSNVLPÕÕEKE	DLESIMDTLMSAIKGASAKGKSSAA	ÕLÕAINMAVASSLA	ELVVAEDAGI	NÕASIAVKTÕALTGALGO	OCFOAVMGTVDRKFIN	INDLITMFAREAASDSNEI
AcSp [L.geometricus]	SSKSP	ANPAKANAFMKCLIOKIGTSPVFPÕÕEKE	DMEEIVETMMSAFSSMSTSGGSNAA	KLÕAMNMAFASSMA	ELVIAEDAD	VPDSISIKTEALAKSLOO	OCFKSTLGSVNRHFIA	TKDLIGMFAREAAAMEDTG
AcSp [L.hesperus]	SSKSP	ANPAKANAFMKCLIÕKISTSPVFPÕÕEKE	DMEEIVETMMSAFSSMSTSGGSNAA	KLÕAMNMAFASSMA		VPDSISIKTEALAKSLO	CFKSTLGSVNRHFIA	TKDLIGMFAREAAAMEEAG
AcSn [N clavines]	SSKSP	ANPKKANGEMKCLIOKISVSPVEPÕÕEKE	DMESIVETMMSAISGVSTSRGSSEA	TLOAMNMAFASSMA	ELVTAEDVNN	VPDSTAEKTEALSOALK	OCFRSTMGTVNROFTT	TKHLMTMFAAEAAOEAAAG
AcSn [P tenidariorum]	NSKSP		DMEETVETMMSAFSSMSSSCCSNAA	KMOAMNMAFASSMA	ELVIAEDADI	DIDSTNVKTEALAKALO	CFKSTLCTVNRHFTM	TKDLTGMFAKEAAANDDSE
AcSp [S grossa]	SSKSP		DMEEIVETMMSAFSSMSTSCCSNAA	KLOAMNMAFASSMA	ELVIAEDADI	PRGISIKTEALAKSLOO	CEKSTMCSVNRHETA	TKDLIGMEAREAASDDGA
FISn [A diadematus]	SSVTD	ADDYTAENEARAEVNNTVNSCEECAODAA	DEDDITOSLLOA - ONLHK - RHDSNA	KAKAMOMAFASSTA	FLVTAR SF	CANTORR STUSICAR	A MOSTTCC VDEEEMR	TEDLIHLESOESENEV
FISP [A argontata]	GTVAD	SDDNUAMUEAKSEVUNIA SCCEECSOCUE				CUNMOOKTDIVDNCHI	ALMSTTCOVDECEVS	TIDI UNI FOCOSNEV
ElSp [L bosporus]		DENERNEEECKI EVINI IECOVECOVDV		TERMATASSIA		INT CT VERTEN VED NOT	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	
FISP [L.nesperus]	TANCH	CNDNEARARA DEFUCNTUS COFFECA OCAR	DEEDITOSITON OCMCKCDUDEKA	TRATE MMALASSIA		CCDUODENTRAVIDAMA		
FISP [N.Clavipes]		SNPNIALAFARSFVSNIVSSGEFGAQGAE	DYNEWIICIENA EICCKODUNERA	KARAMU (ALASSIA		SGGD VORKINVISNALKI		TEDMUEVI CDUNCDCENEIDID
FISP [P.tepidariorum]	NADISIL	SNPQRAREFARNFV SNLEDCGMFNIESVE					TFRQINCPENEDFIS	
FISP [S.grossa]	QUDADPS	DFNETNEEFAKIFVNNLIESGVFGRKGDK	DETAITESLLKAIQLESKGQNAPVS	TKKTFMMAFASSLA	ELIVSE SNI	NALNLIEKTRAVTDSMRI	DAYMKTSGNPNEALIO	VDFLVGIFLDVQGTADDDYEYE
Masp [A.aperta]	QSNPV	TDTATAESFISSVMSSVANQGCLSYDQID	DMQAVGDTMLATMDNLVRSGKSSSH	MLKAMNMAMGTSIA	EI - VADGGG-	NLGSKVSCISNALS	SAFLOTIGSVNTQFVN	I VSLISMFAQADINEV
MaSp [A.aquatica]	QSNNPV	SNSALAEQFISGFIGQMSYQTCFTVDQLE	DMASIGESIMASIDKMEKTGRSTPK	KLQALNMAFASGIA	ELAVSEDGG-	ASLDVKTNAIADSLSS	SAFMQTTGAVNQQFVN	EL RSLVGMFAQANANE I
MaSp [C.angustiarum]	GKSQVI	QSPQMAAAFMQTAIQSLYHNPNFTPDQIS	DMFSMADALIASVNNVAYSDRTSTR	RKNAIRMAMASSIA	ELAARDEGG-	PLSAKASAIANAMAI	DGFMETTGAVDETFIR	LIAMINMFALSSENSI
MaSp [D.canities]	MANTPL	ANKIMAEDFMNKFTNQLANSPYFSSQQKE	DMSSIKDELISVIESMDSAHKSSAA	KLQAMNMAFASAIA	JIAATEAYG-	ADISLETSAIANALSE	SAFLQTTGVVNKRFIS	CLQELIYMFAQDASVQS
MaSp 1 [A.diadematus]	LGQSP	KDASIAESFMNNFSNALARSGAFSSDQMD	D IMSICDSIQSGIVRMERSGKTSAN	KLQAMNMAFASAVA	EIAIAEGGGÇ	QSAQVKTNAVADALAS	SAFLQTTGVVNTQFIN	IRTLISMFAQANVIDS
MaSp 1 [A.argentata]	LGQSP	QSASMAESFMTSFSNALGQSQAFTDEQMD	DIDTIAASIKMGVDKMERSGKTSQN	KLQAMNMAFASAVA	EIAI <mark>SE</mark> GGGÇ	QSAQVKTNAIADALAS	SAFLQTTGVVNGQFIN	IRGLISMFAQANSISS
MaSp 1 [C.moluccensis]	LCQSP	QSASMAESFMTYFSEALGQSGAFTKEQID	DIDTIASSIKLGVDKMERSGKTSGS	K <mark>LQAMN</mark> IAFA <mark>S</mark> AVA	EIATTEGGEÇ	Q ——TAEVKTKAVADAL AB	FAFFQTKGAVNINFIN	EIKNLISMFAQTNTISS
→ MaSp 1 [E.australis]	SHTTP	TNPGLAENFMNSFMQGLSSMPGFTASQLD	D <mark>MSTIAQSMVQSIQSLAAQGRTSPN</mark>	KLQALNMAFASSMA	EIAASEEGGG	GSLSTKTSSIASAMS	NAFLQTTGVVNQPFIN	I TQLVSMFAQAGMNDV
MaSp 1 [K.hibernalis]	AANSP	SNAKTAEIFISKFISAILDSNAFTREQKE	DMMSIGETIIPAMEKMSGSSKSIHA	KLTALNMAFASSVA	EIAVVEEGGS	SDINEKTYAIVAALNQ	AFLDTTGKVNKQFIA	IRDLVKMFASANEENE
MaSp 1 [L.geometricus]	QANTP	SSKANADAFIGSFISSAONTGAFSTDOMD	DMSLIGNTLMAAMDNMGGRITPS	KLQAL DMAFASSVA	EIAAAEGG	DLGVTTNAIADALTS	AFYOTTGVVNSRFIS	IRSLINMFAQASANDV
\rightarrow MaSp 1 [L.hesperus]	OANTP	SSKANADAFINSFISAASNTGSFSODOME	DMSLIGNTLMAAMDNMGGRITPS	KLOAL DMAFASSVA	EIAA <mark>S</mark> EGG	DLGVTTNAIADALTS	SAFYOTTGVVNSRFIS	IRSLIGMFAQASANDV
MaSp 1 [L.mactans]	ÕANTP	SSKANADAFINSFISAASNTGSFSÕDÕME	DMSLIGNTLMAAMDNMGGRITPS	KLÕALDMAFASSVA	EIAASEGG	DIGVTTNAIADALTS	SAFYÕTTGVVNSRFIS	IRSLIGMFAÕASANDV
MaSp 1 [N.clavipes]	ĜONTP	SSTELADAFINAFMNEAGRTGAFTÃDÕLD	DMSTIGDTIKTAMDKMARSNKSSKG	KLÕALNMAFASSMA	EIAAVEOGGI	LSVDAKTNAIADSLNS	SAFYÕTTGAANPOFVN	TIRSLINMFAÕSSANEV
MaSp 1 [P.tepidariorum]	GHRTP	ESAETAEAYVNCFMTNAERTGVFSAEÕMD	DMSSIGGLMKGAMGSMKGKATSH	KLÕALNTAFASAMA	EVATSESGG	NTGAVTDATTNSLNF	AFYÕTTGEGNPĜEVN	TRTLINMFAOSSGNST
MaSp 1 [S grossa]	OANTP	SSKANADSFTOAFMGAASKSGOFTTDOTD	DMSMIGNTIMTAMDNMGGRITPS	KLÕALDMAFASSVA	ETAAVEGO-	NLGASTNATADALOS	AFYOTTGVVNNOFTN	VRSLINMFAOVSDNEV
MaSn 2 [A diadematus]	VADSP	SSNEKADFFIRSFNEVISRSSAFTSOOID	DMSSIGETLISSIDNMAKNGRSSTK	KLOALNMAFASSMA	ETATAEOGG-		AFTRUSGSVNNEFTS	TROLILMESOVSMNDS
MaSp 2 [A argentata]	OAATP	ENSOLAEDEINSELREIAOSGAESPNOLD	DMSSIGDTLKTATEKMAOSRKSSKS	KLOALNMAFASSMA	ETAVAEOGG-		AFLETTGVVNOOFVS	TKGLTYMTAOASSNET
MaSp 2 [A bruennichi]		ENSOLAEDEINSELREIAOSCAESPNOLD		KT.OAT. NMAFASSMA	TAVAROCC-		AFLETTCEVNOOFVS	TKSLTVMTAOASSNET
MaSp 2 [A.bruennen] MaSp 2 [A trifacciata]	OCATP	ENSOLAESETSPEL PETCOSCAESDNOLD	DMSSIGDILKTAIEKMAOSPKSSKS	KLOAL NMAFASSMA	TAVAROCC-		AFLETTCYNOOFUN	TKULTEMIAOASSNET
MaSp 2 [A.titasciata]	QGAII	CCVDNADDETNAET OAACNCCAECCDOUD		VI OAT DMAEACOVA			A EXOTICULANOET C	
MaSp 2 [L.geofficiencus]		CEVENADAETCAEMNAACOCAECCDOTD	DMSVIGNILMIAMDNM CCRIMPS				AFIQIIGV VNNQFIS	T CCT T CMED OVCD NEV
MaSp 2 [L.nesperus]							SAF IQIIGV VNNQFII	
MaSp 2 [N.Clavipes]	QANTP	SDTATADAF IQNFLGAVSGSEAF TPDQLD	DMATVGDTIMSAIDKMARNNKSSKS	NEQSERMAPASSIA			SAFIMITGSINQUEVN	MRRLINMLSAASVNEL
Masp 2–like [N.I.madag.]	QANTP	SDTATADAF I ONFLGAV SGSGAF TPDOLD		KLQAL NMAFASSMA		USMDVKTNATANALDS	SAFIMITGSINQUFVN	MRSLINMLSAAAVNEV
MISP [A.diadematus]	QGQKI	INNPDAAMAMTNNLVQCAGRSGVLSADQMD	DLGMVSDTVNSQLRRMGANAPLH	KI RAVNTAVAAGVA	LVVASSPPQ-	SYSAVENTIGGCER	SMMQVTGSVDNTFTN	
	SUSUPIN	TNPNAAMTMTNNLVQCASRSGVLTADQMD	DMGMMADSVNSQMQKMGPNPPQH	RLRAMNTAMAAEVA	EVVATSPPQ-	SISAVLNTIGACLRE	SMMQATGSVDNAFTN	SVMQLVKMLSADSANEV
MISp [A.argentata]	LGQQFV	SDRGSAMAMTNGFVGCLKQSGYFSSEQMD	DLDLIIDTTAMQLDKMGNRVSQH	KLKALTMGMAGSMA	ETVVSSPYG-	VSPDLMNSIDSCLTC	SMLQVTGNIDQTFRT	
MISp [L.geometricus]	DAASV	DSTAKAEAFIGSFNSGMERSGVLSRSQME	DIASISDTIISAIERNPNNSRS	KLQALNMAFASSVS	ELAFSENNG-	ISNGAKIQAIIDALRO	GAFLQTIGTVDQTFLN	CLASLVKMFSQVSAENEV
MiSp [L.hesperus]	DAASV	DSTATAEAFIGSFNSGMERSGVLSRSQMD	DISSISDTIISAIERNPNNSKS	KLQALNMAFASSVS	ELAFSENNG-	ISNSAKIQAIIDALRO	GAFLQTIGTVDQTFLN	CLSSLVKMFSQVSAENAV
MiSp [L.tredecimguttatus]	DAASV	DSTATAEAFIGSFNSGMERSGVLSRSQMD	DISSISDTIISAIERNPNNSKS	KLQALNMAFASSVS	EIAFSENNG-	ISNSAKIQAIIDALRO	GAFLQTIGTVDQTFLN	ISSLVKMFSQVSAENAV
MiSp [M.grandiosa]	AHGHI	GTPGAGKSVTGSIVQCAGQSGVFSGDQMQ	DLGDMADAVNRQLDRLGPNAPDH	RLKGVTTMMAAGIA	DAAVNSPGQ-	SLDVMINTISGCMT(DAMSQAVGYVDQTLIR	VAEMVNMLANENANAVS
MiSp [N.clavipes]	ALSASI	SSTSMAESFMQSFTTTLGQKGVLSGDQMD	DIASIGDTLMGAVEKSGGKKN	KLQALNMAFASSVA	EIAFADMTG-	LPADVKTNAILNSLSE	EAFLQTTGFVDNYFIQ	GGLINMFAEATANEAS
MiSp [P.tepidariorum]	KAASP	DNTQLAEGFIGAFNAAMAQSGTLSSTQMD	DISSISDTIILAIERNPSNSRS	KLQALNMAFASSVS	EIAFSENIG-	VPNSVKIQAIIDALRO	GAFLQTTGTVDQTFLN	GIGSLVRMFSQVSAENEV
MiSp [S.grossa]	DAASV	DSTATAEAFIGSFNSAMERSGVLSSSQVE	DIASISDTIISAIERNPNNSRS	KLQALNMAFASSVS	EIAFSESNG-	ISNSLKIQAIINALKO	GAFLQTTGTVDQIFLN	IGSLVRMFSQVSAENEI
MiSp [U.diversus]	QGASV	SSPQMAENFMNGFSMALSQAGAFSGQEMK	DFDDVRDIMNSAMDKMIRSGKSG RG	AMRAMNAAFGSAIA	EIVAANGG	KEYQIGAVLDAVT	NTLLQLTGNADNGFLN	ISRLITLFSSVEANDVS
PiSp [A.diadematus]	SSKSW	QDETTAIYFLEHVLENMRECGVLKIDQER	DAIDTMFQ-MSSMFRQNA-KVS	KIKHSIASKLA	DIVIEGMEGI	DS-DPSYKMDCVTQAIAS	SAFEATTGKVDESFIE	SVNELIVVIYNNDIEEKMEEEIEEIS
PiSp [A.argentata]	LAKSW	QDEGTAIYFLNRAIEYLRECGVLKIDQER	DAIETMFE-TSSLYQKNA-KPS	KIKHSVASKLA	EIIIEGLEG <i>I</i>	AS-DTSYKMDCVSQAIAS	SASEATTGTVDDNFIE	VQELVVVMYNNDIEDKLEEIEEIA
PiSp [A.aquatica]	RLPKT	KDQRSFNSFVEKLVYYFKDFNVLHAGDEG	EFLEA-LNYIFMTNEDKET	RVLKHALASEIA	OVIGKAIE-A	AGVNWKYCLANAVQTTAE	FAFEDVYETVDDSFLE	IIRMVTDKYDDDCKIHECEEVSNDGSF
PiSp [L.hesperus]	SAKAL	QEEKSTVYFLDRFVEYLRACDVLKADTER	IAIDSLYS-TRKLFYKNA-SPT	KIRHSLVTKIA	EVITDGLE-1	FGQDVTYKLDC SNQALAA	AFQDASGALDDSFIE	SVQEMVVVMYNNVVEEVIDELEEKQ
PiSp [N.clavipes]	ASKSM	QDEGTTLYLMERMVDNLKDCGIVKIEQEA	DAINSLFE-ISGLYQKNA-KPS	KIKHSMASKIA	DIVIEGLEG	DGDTSYKLDCVNQAAA	AMEETAGVVDDNFIE	IQELVVVMYNNDVEEQQELEQE
PiSp [S.grossa]	SAKAI	QEERSTVYFLDRFVDYLRACDVLKADTER	IAIDSLFS-VRKFFHKNV-SPT	KIRHSLVTKIA	EIITDGLE-7	FGKDVTYKLDCSNQALA	AFQDVSGSLDDTFIE	VQEMVVVMYNNVVEEVIDELEEKQ
TuSp 1 [A.aperta]	TAGIRNIE	GNPNTANNFVDCLKGGIQASPAFPROEOA	DIQSIASSILSAGNTATKSKA	IEQALSTALASSLA	EIVITESGGC	QDYSKQITDLNGILSN	ICFIQTTGVENKRFVN	JIONLIRLLAESAVSET
TuSp 1 [A.ventricosus]	TAVPSVE	SSSTLASSFLOCLTSGIGSSPAFPSÕEÕE	DLDAIAOVILSAVSSNTGATSSA	RAOALSTALASSLT	DLLISESAE	GNYDNOLSTLTSILSN	CFVTTTGSDSPAFVS	IKSLISVLSONNNAI
TuSp 1 [A.argentata]	AVPSVE	SSPNLASGFLÕCLTFGIGNSPAFPTÕEÕO	DLDAIAÕVILNAVSTNTGATASA	RAOALSTALASSLT	DLLIAESAES	SNYNNÕLSELTGILSN	ICFIOTTGSDNPAFVS	IOSLISVLSÕNTDVNII
TuSp 1 [A.bruennichi]	AVPSVE	SSPNLASGFLÕCLTFGIGNSPAFPTÕEÕÕ	DLDAIAÕVILNAVSSNTGATASA	RAÕALSTALASSLT	LLIAESAES	SNYSNOLSELTGILS	OCFIOTTGSDNPAFVS	IOSLISVLSÕNADTNII
TuSp 1 [L.hesperus]	ASVNT	NSPNAATSFLNCLRSNTESSPAFPFOEOA	DLDSIAEVILSDVSSVNTASSA	TSLALSTALASSLA	ELLVTESAEF	EDIDNOVVALSTILS	CEVETTGSPNPAEVA	VKSLLGVLSOSASNYEFV
TuSp 1 [N.antipodiana]	SVPSVE	SSPSLASGFLGCLTTGIGOSPDFPFOEOO	DLDDLAOVILSAVT SNTDTSKSA	RAOALSTALASSLA	DLLISESSG	SSYOTOISALTNILS	OCFVTTTGSNNPAFVS	VOTLIAVLSÖSSSNAT
TuSp 1 [N clavata]	PVPSVI	SSPSLASGELGCLTTGIGLSPAEPEOEOO	DLDDLAKVILSAVT - SNTDTSKSA	RAOALSTALASSLA	DLLTSESSG	SSYOTOTSALTNILS	CEVTTTGSNNPAEVS	NOTLIGVISOSSSNAT
TuSn 1 [N clavines]	ΟΔΤΩΥΓ	SSPSIASSFLGCLTNGTGOSPVFPSOFOT	DI.DDI.AKVII.SAVT - SNTDISKSA	RAOALSTALASSLA	TITSESCO	ISYONOTSALNUTLS	ICEVTTTCSNNPAEVC	NOTLIGVESOSSSNAT
TuSn 1 [S grossa]	SAVNVE	TSPSAASSFLOCI, TSSLCSSPAFPTOOD	DI.DSTAKATI.SDVSSVSSAPTTA	TAOALSAALASSMA	ELLVTES A FT	PETEEOVSALSCILS		WKSLLGVLSOSSSSAOAT
145p I [5.91055a]		TOT DUMPOL PÄGTT DOT GODT VLT TÄÄÄD	•				501 TÕT TOUT MI AT AD	TOTTO I DO CODORIÓNI
]		1	I	I I			
	W	0 D	40 k	(65 E	79 E84		E1	19

Figure S2: Superposition of the 20 conformers with the lowest target functions.

Superpositions of the polypeptide backbone heavy atoms of residues 11-55 and 61-130 for **(A)** the bundle of conformers representing the NMR structure of wt FlSp NT monomer, **(B)** the bundle of conformers representing the NMR structure of wt FlSp NT dimer and **(C)** the wt FlSp NT monomer (orange and dimer (green) structures showing the 20 conformers of Phe11 side chain.



Figure S3: Size exclusion chromatography of wt FISp NT. (A) Overlay of chromatograms at pH 5.5 (black) and at pH 8.0 (green). **(B)** Dimer to monomer ratio calculated from the apparent molecular weights of wt FISp NT at pH 5.5 and pH 8.0, as determined from the elution volumes compared to a set of calibrants at each pH.



Figure S4: ESI-MS analysis of wt FISp NT dimers at different concentrations. Spectra measured at pH 5.5 show predominantly dimers at concentrations ranging from $7 - 125 \mu$ M.



Figure S5: Dimerization studies of Phe to Trp mutants. Analysis was performed for NT_{Trp} (red) and NT^*_{Trp} (blue) by (**A**) size exclusion chromatography at pH 8.0 (solid lines) and at pH 5.5 (dashed lines) or by tryptophan fluorescence spectroscopy of (**B**) NT_{Trp} and (**C**) NT^*_{Trp} . The fluorescence was measured at pH 5.2 (brown), pH 5.6 (red), pH 6.0 (pink), pH 6.4 (orange), pH 6.8 (green), pH 7.2 (blue), pH 7.6 (gray) and pH 8.0 (black).



Figure S6: NMR analysis of wt FISp NT and mutants. (A) Averaged backbone amide ¹H and ¹⁵N chemical shift differences $\Delta \delta_{av} = \sqrt{(0.1 \Delta \delta_N)^2 + (\Delta \delta_H)^2}$ between wt FISp NT at pH 7.2 and pH 5.5 (left panel). The chemical shift difference per residue (center panel) and the charge distribution of FISp NT are displayed in surface representations. HSQC spectra of (B) FISp NT* at pH 7.2 (green) and 5.5 (yellow), (C) FISp NT_{E79QE84QE119Q} at pH 7.2 (purple) and 5.5 (orange) and (D) FISp NT_{E79QE84QE119QE130Q} at pH 7.2 (black) and 5.5 (pink). (E) Estimation of the molecular weight of wt FISp NT and FISp NT mutants at pH 7.2 and pH 5.5 using ¹⁵N NMR relaxation analysis. The table shows (from left to right) the ¹⁵N NMR T₁ and T₂ relaxation times, the rotational correlation times (τ_c) and the estimated molecular weights (MW).



Figure S7: Intensity decay profile for PFG-NMR self-diffusion measurements of wt NT at pH 7.2 and pH 5.5. The data have been fitted to an exponential equation using MestReNova 14.



Figure S8: Refolding capacity and thermal stability measured with circular dichroism spectroscopy. (A) Spectra were measured in 20 mM sodium phosphate at 25°C (red), 95°C (black), and the ability to refold was evaluated from spectra after cooling the samples to 25°C (blue). The data are presented as a smoothed average of five measurements. (B) The thermal stability of wt NT (black), NT* (red) and NT_{E79QE84QE119Q} (blue) was assessed by temperature-induced denaturation. Ellipticity at 222 nm was measured as a function of temperature. The respective T_m of the NT variants are indicated between brackets.



Quantity	Monomer at pH	Dimer at pH 5.5 ^a
	7.2 ^a	
Extent of chemical shift assignment		
backbone H, 15 N, 13 C $_{\alpha}$, H $_{\alpha}$ atoms	92.1% ^b	99.6%
non-labile side chain protons	97.2%	98.6%
NOE upper distance limits	1387	2300
intra-residual (i-j =0)	466	838
short-range	365	650
medium-range	291	368
long-range	265	408
intermolecular	-	36
Residual NOE violations		
number ≥ 0.1 Å	9 ± 3	20 ± 3
maximum [Å]	0.22 ± 0.05	0.27 ± 0.09
PARALLHDG energies [kcal/mol]		
total	-4951 ± 76	-10715 ± 174
van der Waals	-455 ± 25	-1027 ± 32
electrostatic	-5332 ± 79	-11352 ± 159
R.m.s.d. from mean coordinates [Å]		
backbone (residues 11–55,61–130) ^c	0.74 ± 0.12	1.00 ± 0.14
all heavy atoms (residues 11–55,61–130) ^c	1.13 ± 0.09	1.41 ± 0.16
Ramachandran plot statistics		
most favored regions [%]	94.3	88.9
additional allowed regions [%]	5.3	10.6
generously allowed regions [%]	0.3	0.4
disallowed regions [%]	0.1	0.1

Table S1. Input for structure calculations and structural statistics of the NMR structures ofFISp NT monomer and dimer at pH 7.2 and pH 5.5, respectively.

^aExcept for the top eight entries, average values and standard deviations for the 20 energy-minimized conformers are given.

^bMissing assignments for the first 9 residues (including five N-terminal residues GSGNS stemming from the expression and purification tag) and residues 57-62 located in a loop between helices 2 and 3. ^cFor the dimer, values given are for both chains.

Data collection and processing					
X-ray source	MAX IV BEAMLINE BioMAX				
Wavelength (Å)	0.9788				
Space group	P2 ₁				
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å), β (°)	41.88, 58.89, 46.33, 106.65				
Resolution limits (Å)	44.39-1.80 (1.90-1.80) ^[a]				
Observations	100490 (6887) ^[a]				
Unique reflections	18569 (1906) ^[a]				
Completeness (%)	92.6 (66.0) ^[a]				
R _{merge}	$0.094 \ (0.324)^{[a]}$				
⟨ Ι/σΙ ⟩	10.7 (2.0) ^[a]				
Multiplicity	5.4 (3.6) ^[a]				
Refinement					
R _{work}	0.186 (0.259) ^[a]				
R _{free}	0.2173 (0.3390) ^[a]				
Average B factor (Å ²)	26.3590				
(B) from Wilson plot (Å ²)	23.9				
No. atoms					
Protein	1932				
Water	71				
Ion	0				
R.m.s. deviations from ideal values					
Bond lengths (Å)	0.016				
Bond angles (°)	1.9				
Outliers in Ramachandran plot (%)	0				
PDB code	700M				

 Table S2. X-ray data collection, processing and refinement statistics.

^[a] Values in brackets refer to the highest resolution shell.