



TITLE:

Unusual K_a Values Mediate the Self-Assembly of Spider Dragline Silk Proteins

AUTHOR(S):

Oktaviani, Nur Alia; Malay, Ali D.; Matsugami, Akimasa; Hayashi, Fumiaki; Numata, Keiji

CITATION:

Oktaviani, Nur Alia ...[et al]. Unusual K_a Values Mediate the Self-Assembly of Spider Dragline Silk Proteins. *Biomacromolecules* 2023, 24(4): 1604-1616

ISSUE DATE:

2023-04-10

URL:

<http://hdl.handle.net/2433/283276>

RIGHT:

Copyright © 2023 The Authors. Published by American Chemical Society; This is an open access article published under a Creative Commons Non-Commercial NoDerivative Works (CC-BY-NC-ND) Attribution License.

Unusual pK_a Values Mediate the Self-Assembly of Spider Dragline Silk Proteins

Nur Alia Oktaviani,^{||} Ali D. Malay,^{||} Akimasa Matsugami, Fumiaki Hayashi, and Keiji Numata*Cite This: *Biomacromolecules* 2023, 24, 1604–1616

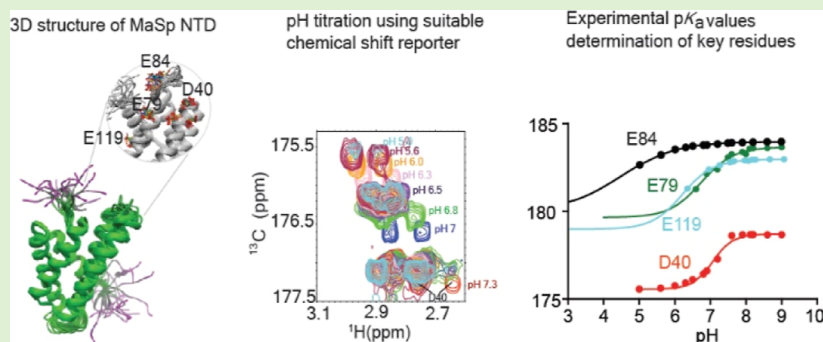
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



ABSTRACT: Spider dragline silk is a remarkably tough biomaterial and composed primarily of spidroins MaSp1 and MaSp2. During fiber self-assembly, the spidroin N-terminal domains (NTDs) undergo rapid dimerization in response to a pH gradient. However, obtaining a detailed understanding of this mechanism has been hampered by a lack of direct evidence regarding the protonation states of key ionic residues. Here, we elucidated the solution structures of MaSp1 and MaSp2 NTDs from *Trichonephila clavipes* and determined the experimental pK_a values of conserved residues involved in dimerization using NMR. Surprisingly, we found that the Asp40 located on an acidic cluster protonates at an unusually high pH (~ 6.5 – 7.1), suggesting the first step in the pH response. Then, protonation of Glu119 and Glu79 follows, with pK_a s above their intrinsic values, contributing toward stable dimer formation. We propose that exploiting the atypical pK_a values is a strategy to achieve tight spatiotemporal control of spider silk self-assembly.

INTRODUCTION

Spider dragline silk is a natural high-performance fiber that is well recognized for its outstanding mechanical properties and biocompatibility. Dragline silk is a hierarchically structured fiber that is primarily composed of spider silk proteins (spidroins) MaSp1 and MaSp2. Both proteins contain conserved N-terminal and C-terminal domains (NTD and CTD) flanked by a long repetitive domain (RD). Despite the predominance of repetitive sequences, the self-assembly of soluble spidroins into mature silk fibers is controlled *via* the small globular NTD and CTD (which have molecular weights of ~ 14 and 10 kDa, respectively).^{1,2} Inside the ampullate gland, spidroin is stored at a high concentration in a storage sac (up to 50% w/v).³ During the formation of native spider silk, the terminal domains are highly responsive to changes in pH and ion composition within the spinning ducts of the silk glands.^{2,4} The NTD maintains a monomeric structure at higher pH conditions found in the silk gland storage sac ($\sim \text{pH } 7$), whereas at lower pH values [as occurs in the spinning duct ($\sim \text{pH } 6$ – 5)], the NTD undergoes dimerization with an antiparallel orientation, a response also mediated by the NaCl concentration.^{1,5} In contrast, the CTD forms a stable dimer at

neutral pH and gradually unfolds at more acidic pH values.² The repetitive regions, on the other hand, maintain a dynamically disordered structure in solution,^{6,7} mainly composed of random coil conformations together with small amounts of PPII helices.^{8,9} The conformation of the isolated RD was not found to depend on the pH,⁹ but the intra- and intermolecular interactions of the RD were affected by the ion composition and concentration.¹⁰ Recently, it was also shown that liquid–liquid phase separation in response to anion and pH gradients is essential in controlling the self-assembly of spidroins into hierarchical structures.¹¹

The pH response of the NTD must be tightly regulated to ensure the correct timing of the structural changes leading to dimerization as the spidroins migrate along the multiple physicochemical gradients. Previous studies have identified an

Received: November 11, 2022

Revised: February 17, 2023

Published: March 29, 2023



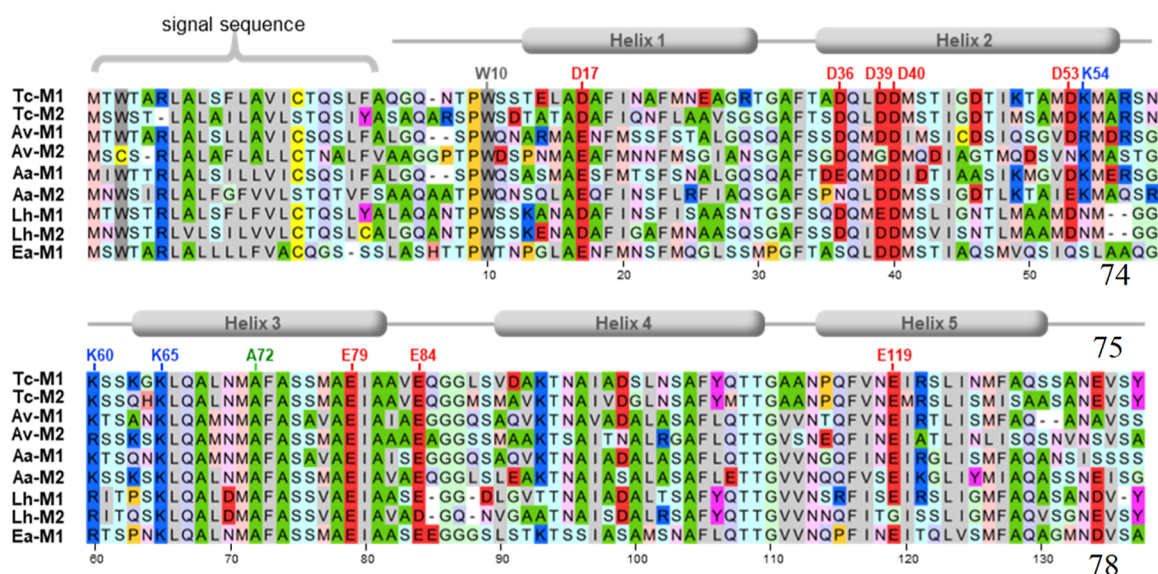


Figure 1. Alignment of MaSp1 (M1) and MaSp2 (M2) NTD amino acid sequences. Tc = *Trichonephila clavipes* (family: Araneidae); Av = *Araneus ventricosus* (Araneidae) and *Argiope argentata* (Araneidae); Lh = *Latrodectus hesperus* (Theridiidae); Ea = *E. australis* (Pisauridae). The locations of the N-terminal signal sequence (which is omitted from the constructs), α -helix, and loop elements are indicated on the top, and the residue numbers are shown on the bottom. The residue numbering for Tc-M2 has been shifted by -1 relative to the deposited structure (PDB ID: 8GS7) throughout this paper to facilitate comparisons with the other NTD variant.^{12,14} Several residues relevant to this study are indicated.

array of conserved anionic and cationic residues on the NTD surface that regulate different stages of the dimerization process (Figure 1).^{12,13} Most studies on the structure and function of spider NTD have been based on MaSp1 from *Euprosthenoops australis*, which contains a globular structure consisting of five α -helices, and a main acidic cluster that includes conserved residues D39, D40 (helix 2), and E79 (helix 3), which play important roles in dimerization, along with the more distally located E84 and E119 (helix 5).^{12,14,15} There is also a basic cluster that includes R60 and K65 on a loop between helices 2 and 3, which, upon antiparallel dimerization, are well positioned to form intermolecular interactions with the acidic cluster on the opposite subunit.^{1,12} Despite the conservation of sequence and overall structure, it is unclear whether the specifics of the dimerization mechanism are the same among different species as well as among different spider types, as discrepancies in the structures and mechanistic contributions of residues have been reported.^{13,15,16} To understand the contributions of the different anionic and cationic residues to the NTD dimerization mechanism, tryptophan (Trp) fluorescence spectroscopy, in conjunction with mutational analyses, has been used extensively.^{12–15} In addition, the monomer–dimer equilibrium of the *E. australis* MaSp1 NTD was investigated using fluorescence cross-correlation spectroscopy (FCCS),¹² while the kinetic dimerization on wild-type (WT) and mutant NTDs was studied using steady-state and time-resolved fluorescence experiments.¹⁷

While the interpretation based on Trp fluorescence spectroscopy provides useful insights, we consider that the results are less straightforward for the NTD dimerization since the technique shows the overall monomer–dimer equilibrium of WT and mutant NTDs; in contrast, the pH-dependent dimerization of NTD originates from intermolecular interaction between monomers, which is the result of ionization (protonation or deprotonation) events of several key residues, and this event is not well understood. To date, the detailed

molecular basis that explains the step-by-step mechanism has remained unclear because direct evidence of the relevant experimental pK_a values of individual key residues is still missing. Previously, computational methods have been used to predict the protonation behavior of NTD residues in a pH gradient.¹⁸ Furthermore, in another study using a coarse-grained model, the pK_a values of individual acidic and basic WT and mutant NTDs were also calculated.¹⁹ Although computational methods for estimating protein pK_a profiles are undoubtedly powerful, such approaches sometimes have limitations, and in some cases, the predictions may deviate from reality.^{20,21} Strictly speaking, such theoretical methods are valid only when the three-dimensional protein models (on which the calculations represent the conformational spaces sampled by the protein during its structural transition) are adequate. Furthermore, the theoretical calculations, which are based on static structures, are further complicated by the plasticity and conformational dynamics of the NTD.^{14,16,22}

To fabricate artificial spider silk, which mimics the mechanical properties of native silk, a comprehensive understanding of the spider silk self-assembly mechanism must be obtained, including the universal dimerization mechanism of the NTD. Here, we determined the solution structure and dynamics of MaSp1 and MaSp2 NTDs from *T. clavipes* and investigated the dimer formation of both NTDs. To pinpoint the sequential molecular mechanism of NTD dimerization, we performed pH titration on the MaSp1 and MaSp2 NTDs and followed the chemical shift changes of individual key amino acid residues using NMR spectroscopy, whereby we obtained the relevant experimental pK_a value and protonation state of individual amino acid residues. Additionally, we performed extensive mutational analysis on the NTDs and dimerization assays based on tryptophan fluorescence, the results of which largely converged with previous findings.^{12,13,15} Altogether, the results of this study provide clear and direct evidence of the pH-response dimerization of both MaSp1 and MaSp2 NTDs,

leading to a better understanding of the mechanism of NTD dimerization, which is crucial for spider silk spinning.

EXPERIMENTAL SECTION (MATERIALS AND METHODS)

Genetic Constructs. The spidroin NTDs used in the study were based on sequences from *T. clavipes*: MaSp1 (UniProt accession code B5SY55) and MaSp2 (UniProt accession code ACF19413). Constructs encoding the two WT sequences were purchased from GenScript as NdeI/XhoI inserts in the pET15b vector (Novagen).

Site-Directed Mutagenesis. Point mutations were generated *via* PCR-site-directed mutagenesis using primers featuring approximately 15 complementary bases on either side of the substitution site. Thirty amplification cycles were carried out using KOD+ polymerase (Toyobo), followed by DpnI digestion (Takara), and transformation into XL10-Gold ultracompetent *Escherichia coli* cells (Stratagene). Transformants were screened by colony PCR using GoTaq Master Mix (Promega) and T7 promoter/terminator primers, followed by DNA sequencing using the BigDye Terminator v3.1 reaction (Thermo Fisher Scientific).

Expression and Purification of WT and Mutant MaSp1 and MaSp2 NTDs. A plasmid of interest was transformed into BL21(DE3) *E. coli* cells and inoculated into a 100 mL culture at 37 °C in Luria Bertani (LB) medium (BD) with 100 $\mu\text{g mL}^{-1}$ ampicillin (Sigma) with overnight shaking at 180 rpm. This preculture was used to inoculate 1 L of fresh medium, which was grown at 37 °C with shaking until the OD₆₀₀ reached ~ 0.8 . Protein expression was induced by adding 0.4 mM IPTG (Wako) with overnight shaking, and the temperature was decreased to 30 °C. Cell pellets were harvested by centrifugation and were resuspended in 20 mM Tris-HCl pH 7.5, 0.15 M NaCl containing 10 μM (4-amidino-phenyl) methanesulfonyl fluoride (PMSF) as a protease inhibitor. The cells were lysed by sonication on ice using a QSonica Q500 instrument. The lysates were centrifuged at 8000 rpm for 30 min at 4 °C, and the supernatant fraction was transferred to fresh Falcon tubes and incubated for 30 min. The clear supernatant was loaded onto a 5 mL HisTrap column (GE Healthcare), and nonspecifically bound material was washed off with binding buffer (20 mM Tris-HCl pH 7.5, 20 mM imidazole, 500 mM NaCl) using an AKTA Explorer (GE Healthcare). The bound protein was recovered with elution buffer (20 mM Tris-HCl, 500 mM imidazole, and 500 mM NaCl, pH 7.5). The eluted product was concentrated and buffer-exchanged against 20 mM Tris-HCl, pH 7.5, with 0.15 M NaCl at 4 °C by centrifugation at 8000 rpm in a VivaSpin centrifugal concentrator with a 10,000 kDa cutoff (GE Healthcare). The poly-histidine tag was removed by overnight digestion with thrombin (Sigma) at 4 °C, and the protein was further washed and concentrated in the same buffer using VivaSpin to a final volume of 0.1–0.3 mL. The protein concentration was calculated by measuring the absorbance at 280 nm using a NanoDrop instrument (Thermo Fisher Scientific).

Fluorescence Spectroscopy Measurements. NTD dimerization in response to pH changes was monitored according to shifts in tryptophan fluorescence emission spectra. Each assay contained 5 μM NTD, an appropriate level of NaCl, and a universal buffer system to control the pH,²³ which consisted of sodium acetate, 2-(*N*-morpholino)ethanesulfonic acid, and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (20 mM each of), with the pH adjusted to the appropriate values (increments of 0.25 pH units). The assays (100 μL) were applied onto a 96-well black plate (Iwaki), and fluorescence spectra were measured at 25 °C using a Spectramax M3 instrument (Molecular Devices) with excitation at 280 nm and emission spectra from 300 to 400 nm. Response curves were generated by taking the ratio of emission intensity between 339 and 351 nm and plotting the results as a function of pH. The fitting and midpoint between the monomer and dimer forms (*pK*) of WT and mutant NTDs for this type of curve were generated according to the following equation

$$I_{\text{obs}} = I_{\text{dimer}} + \Delta I \frac{10^{n(\text{pH}-\text{pK})}}{1 + 10^{n(\text{pH}-\text{pK})}} \quad (1)$$

where I_{obs} is the observed ratio emission intensity between 339 and 351 nm at a certain pH value, I_{dimer} is the ratio emission intensity between 339 and 351 nm at low pH or in dimeric form, n is the slope of the curve, and *pK* is the midpoint between monomer and dimer forms of NTD. This fitting equation is adopted from the Henderson–Hasselbalch equation.

Protein NMR Sample Preparation. For the protein NMR samples, recombinant WT and mutant MaSp1 and MaSp2 NTDs were doubly labeled (¹³C, ¹⁵N) by growing BL21(DE3) *E. coli* in M9 minimal medium containing ¹³C-glucose (Cambridge Isotope Lab) and ¹⁵N-ammonium chloride (Cambridge Isotope Lab). First, the cells were grown in 5 mL of LB medium containing 100 $\mu\text{g/mL}$ ampicillin. Then, this culture was transferred into 200 mL of M9 minimal medium containing 100 $\mu\text{g/mL}$ ampicillin and grown overnight with shaking at 160 rpm at 37 °C. This preculture was transferred into 2 L of the main culture of M9 minimal medium containing 100 $\mu\text{g/mL}$ ampicillin. The cells were grown until they reached OD₆₀₀ ~ 1 . Then, 0.4 mM IPTG (Wako) was added to the cell cultures, and the cells were grown overnight at 30 °C with shaking at 160 rpm. The (¹³C, ¹⁵N) WT and mutant MaSp NTDs were purified in the same way as the unlabeled NTDs. The NMR samples contained ~ 0.5 –1 mM WT and mutant MaSp NTDs in 10 mM phosphate buffer, pH 7, 10% D₂O, and 0.1 mM 4,4-dimethyl-4-silaphentane-1-sulfonate (DSS).

NMR Measurements for Assigning Chemical Shifts. NMR spectra were recorded using a triple resonance TCI cryogenic probe and a *z*-axis gradient coil with a Bruker spectrometer (700 MHz). All NMR experiments were conducted at 25 °C. Backbone and side chain chemical shifts (¹³C, ¹⁵N) of the recombinant NTD MaSp1 at pH 7 in the presence of 300 mM NaCl were assigned based on 2D and 3D NMR experiments: 2D ¹H–¹⁵N HSQC, 2D ¹H–¹³C HSQC aliphatic, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, 3D HN(CA)CO, 3D HBHA(CO)NH, 3D HBHANH, 3D H(CCO)NH, 3D (H)C(CO)NH, 3D H(C)CH TOCSY, and 3D (H)CCH TOCSY spectra in the aliphatic region. The side-chain carboxyl group resonances of aspartic and glutamic acids and the side-chain carbonyl group resonances of asparagine and glutamine of MaSp2 NTD were assigned based on the 2D H₂(C)CO spectrum.^{24,25} The ¹⁵N ζ –¹H δ correlation of the lysine side chain and ¹⁵N δ –¹H ϵ of the arginine side chain resonances were assigned based on 2D H₂(C)N experiments.²⁶ Aromatic side chain chemical shifts were assigned based on 2D CG(CB)HB,²⁷ 2D CB(CGCD)HD, and 2D CB(CGCDCE)HE.²⁸ The assigned chemical shifts H δ and H ϵ in the aromatic rings from these experiments were used to assign the ¹H–¹³C HSQC aromatic spectrum. The chemical shift assignment of *T. clavipes* MaSp1 NTD at pH 7 in the presence of 300 mM NaCl was deposited at BMRB under accession number 50972.

The backbone and side chain chemical shifts of MaSp2 NTD at pH 7 in the presence of 300 mM NaCl were assigned based on our previous report (BMRB accession number: 50353).²⁹ All spectra were processed using NMRPipe³⁰ and analyzed using NMRFAM-SPARKY.³¹ All chemical shifts were referenced to DSS according to International Union of Pure and Applied Chemistry (IUPAC) recommendations.³² The structural propensities of both MaSp1 and MaSp2 NTDs were calculated using the neighbor-corrected structural propensity calculator.³³

3D NMR Structure Calculations. Distance restraints of both the MaSp1 and MaSp2 NTDs were obtained by measuring 3D ¹H–¹⁵N NOESY HSQC and 3D ¹H–¹³C NOESY HSQC experiments. Structure calculations were performed using CYANA.³⁴ The 20 conformers calculated with CYANA were used as starting structures for further refinement in Cartesian coordinates using XPLOR-NIH.³⁵ The solution NMR structures of MaSp1 and MaSp2 NTDs from *T. clavipes* were deposited in the Protein Data Bank (PDB) under accession codes 7W10 and 8GS7, respectively.

NMR Measurement for Protein Dynamics. ¹H–¹⁵N heteronuclear NOE experiments were performed to characterize the protein backbone dynamics on the picosecond–nanosecond time scale. The ¹H–¹⁵N steady-state NOE values were measured by recording the following spectra: an initial spectrum recorded without

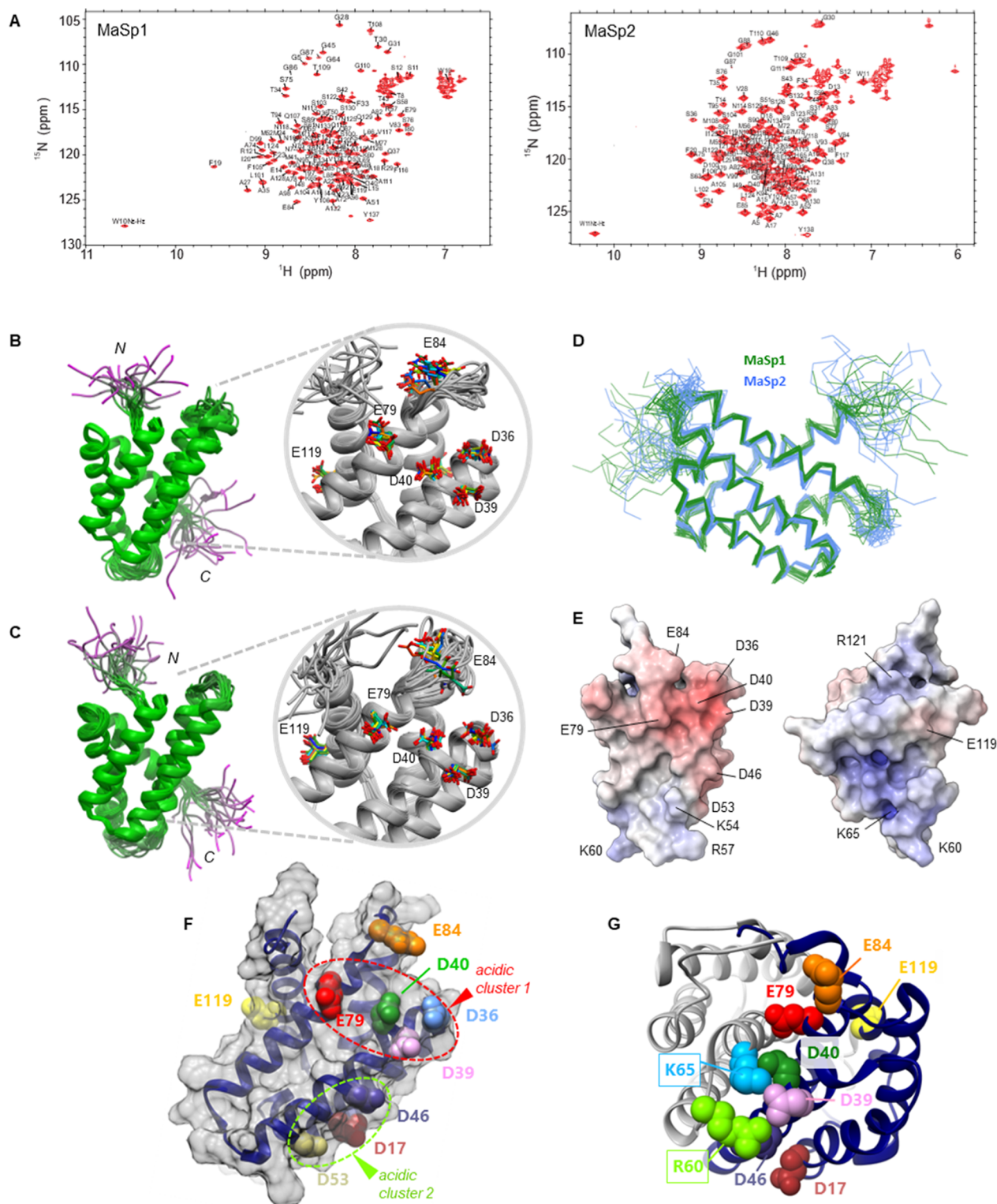


Figure 2. Determination of the NMR structures of MaSp1 and MaSp2 monomeric NTDs from *T. clavipes* at pH 7.0. (a) ^1H - ^{15}N HSQC spectra of MaSp1 NTD (spectral resolutions for ^1H and ^{15}N are 10.1 and 7.1 Hz, respectively) and MaSp2 NTD (spectral resolutions for ^1H and ^{15}N are 10.1 and 6.0 Hz, respectively) in 10 mM phosphate buffer pH 7.0 and 300 mM NaCl. The chemical shift assignments for MaSp1-NTD and MaSp2-NTD were deposited in the BMRB with accession codes 50972 and 50353, respectively. (b,c). Superimposed ensemble structures of MaSp1 NTD (b) and MaSp2 NTD (c) (PDB ID: 7W10 and 8GS7, respectively). Left, 20 superimposed conformers that are colored according to the root-mean-square deviation (RMSD) of all atoms (green = low; magenta = high RMSD). The right portion shows the main acidic cluster region on the monomer surface, and residue side chains are represented as sticks. (d) Superimposed structures of all 20 MaSp1 and 20 MaSp2 NTD conformers, showing the $\text{C}\alpha$ positions, with MaSp1 and MaSp2 in green and blue, respectively. (e) Surface representation of the MaSp2 NTD core region (residues 8–131) in two views, which is colored by the electrostatic surface potential and was calculated using APBS. The relative orientation

Figure 2. continued

(rotation) of the structure is 180° along the *y*-axis. (f) Representative conformer of the MaSp2 NTD is shown in cartoon and surface representations, with the conserved acidic residues depicted as spheres. The two acidic clusters are shown, with cluster 1 comprising D36, D39, D40, and E79 and cluster 2 comprising D17, D46, and D53. (g) Representative conformer of the MaSp1 NTD dimer structure from *E. australis* at pH 5.5 (PDB ID: 2LTH). The backbone of subunit A is shown in dark blue, and the conserved acidic residues are depicted as in (f). The subunit B backbone is shown in gray, and the basic residues R60 and K65 that come in close contact with the acidic residues of the subunit in the dimer form are indicated.

the initial proton saturation and a second spectrum recorded with initial proton saturation (3 s). The $\{^1\text{H}\}-^{15}\text{N}$ steady-state NOE values of the MaSp1 and MaSp2 NTDs were determined using 700 MHz NMR machines (Bruker). The steady-state NOE values were calculated based on the ratios of the average intensities of the peaks with and without proton saturation. The standard deviations of the NOE values were calculated from the background noise level using the following formula

$$\frac{\sigma\text{NOE}}{\text{NOE}} = \sqrt{\left(\frac{\sigma I_{\text{sat}}}{I_{\text{sat}}}\right)^2 + \left(\frac{\sigma I_{\text{unsat}}}{I_{\text{unsat}}}\right)^2} \quad (2)$$

where I_{sat} and I_{unsat} are the measured intensities of the peaks in the presence and absence of proton saturation, respectively. The noise in the background regions of the spectra, which were measured with initial proton saturation and without initial proton saturation, is indicated by σI_{sat} and σI_{unsat} , respectively.

Chemical Shift Perturbation Induced by Specific Mutations. Chemical shift perturbations induced by specific mutations were calculated by the following equation³⁶

$$\text{CSP} = \sqrt{(\Delta H)^2 + \left(\frac{\Delta N}{6.45}\right)^2} \quad (3)$$

pK_a Determination. The pH was changed in steps of 0.25–0.35 pH units by adding a few μL of HCl or NaOH solution. The pH meter (HORIBA) was calibrated using buffers of pH 4.0, 7.0, and 9.0. WT and mutant NTDs were titrated in a pH range between 9 and 5. The chemical-shift titration data were analyzed using the Henderson–Hasselbach equation, which is suitable for the rapid exchange of the nuclei between environments and is associated with neutral and charged states of the side chain as follows

$$\delta_{\text{obs}} = \delta_{\text{AH}} + \Delta\delta \frac{10^{n\text{H}(\text{pH}-\text{pK}_a)}}{1 + 10^{n\text{H}(\text{pH}-\text{pK}_a)}} \quad (4)$$

where δ_{AH} is the chemical shift for the protonated form, and $\Delta\delta = \delta_{\text{A}} - \delta_{\text{AH}}$ is the change in chemical shift upon deprotonation. At lower pH values, some peaks broaden due to intermediate exchange, and samples tend to aggregate; therefore, to estimate the pK_a value of aspartic acid, the $\Delta\delta$ of C γ was assumed to be 3.2 ppm, while to estimate the pK_a value of glutamic acid, the $\Delta\delta$ of C δ was assumed to be 4.1 ppm.³⁷ We used the 2D H₂(C)CO spectrum to follow the $^{13}\text{C}\gamma$ and $^{13}\text{C}\delta$ chemical shifts of aspartic acid and glutamic acid, respectively, as a function of pH. As a control experiment to check whether the pH meter was well calibrated, we determined the pK_a value of His-2 of WT and mutant NTDs, which are located close to the N-terminus of the proteins. To determine the pK_a value of histidine, $^{13}\text{C}\epsilon$ chemical shifts obtained from $^1\text{H}-^{13}\text{C}$ HSQC aromatics were used as a chemical shift reporter.

RESULTS AND DISCUSSION

MaSp1 and MaSp2 NTD Structures from *T. clavipes*. The structures of *T. clavipes* MaSp1 and MaSp2 NTDs at pH 7.0 were solved by solution NMR spectroscopy (Figure 2; the refinement statistics are summarized in Table S1). This is the first report of a MaSp2 NTD structure from any species, while the MaSp1 NTD model supplements previously published structures obtained from different species and under different

conditions.^{1,14,15,38} The 2D $^1\text{H}-^{15}\text{N}$ heteronuclear single quantum coherence (HSQC) spectra mapping the backbone amide proton and amide nitrogen chemical shifts showed some divergence between the MaSp1 and MaSp2²⁹ data (Figure 2a) due to several amino acid differences on the helix regions (Figure 1), which encouraged to experimentally determine the structure of NTD MaSp2.

Interestingly, despite some differences on the fingerprint 2D $^1\text{H}-^{15}\text{N}$ HSQC spectra of both NTD MaSp1 and MaSp2, the translation of the backbone chemical shifts into secondary structures of both MaSp1 and MaSp2 demonstrated five helical regions connected by flexible loops (Figure S1). The final models were deposited in the protein data bank (PDB) with accession codes 7W1O and 8GS7 for MaSp1 NTD and MaSp2 NTD, respectively. The two proteins showed similar overall structures due to high sequence similarity (75% sequence identity) (Figure 2b–d), as well as similarity to other previously reported NTD structures (Figure S2),^{13,14,38} which was consistent with a high degree of functional conservation across spider taxa and spidroin types.

Within the core structure (residues 9–131, excluding the flexible loops at either end), the positions of the helices were constant for all conformers, while the highest degree of conformational flexibility occurred in two loop regions (between helix 2 and helix 3 and between helix 3 and helix 4). As found in previous studies, a distinct asymmetry in the distribution of charged residues was observed in the *T. clavipes* structures, leading to a pronounced dipole moment within each monomer (Figure 2e). The positions of the conserved acidic residues on the surface of the MaSp2 NTD are shown in Figure 2f. These acidic residues are grouped mainly into two clusters; acidic cluster 1 contains residue D40 in the center, which is flanked closely by D39 and D36 on one side and by E79 on the opposite side. Notably, D36 is not present in the well-studied *E. australis* MaSp1. The E84 residue on the loop between helices 3 and 4 is located close to this cluster in some of the conformers; however, the varying side chain orientations suggest that E84 does not form an integral part of this cluster. A smaller cluster (acidic cluster 2), consisting of D17, D46, and D53, is also apparent. The presence of the Asp/Glu residue at position 17 is well conserved among MaSp NTDs, while D53 is also conserved (although *E. australis* MaSp1 has a Gln at the equivalent position) (Figure 2f), and D46 is found to be conserved within the Araneidae family. Interestingly, the intramolecular “handshake” that was reported between D17 and D53 in the intermediate pH 6.5 dimer model of the *T. clavipes* MaSp1 NTD³⁸ was not apparent in our pH 7 structures. The conserved E119 residue on helix 5 is situated on the opposite side of the subunit surface and is not part of either acidic cluster. As a comparison, Figure 2g shows the dimeric structure of the *E. australis* MaSp1 NTD elucidated at pH 5.5 (PDB ID: 2LTH). Residues D40, E79, and E119 are located at the dimer interface, and dimerization in the anti-parallel orientation shifts cationic groups from the opposite

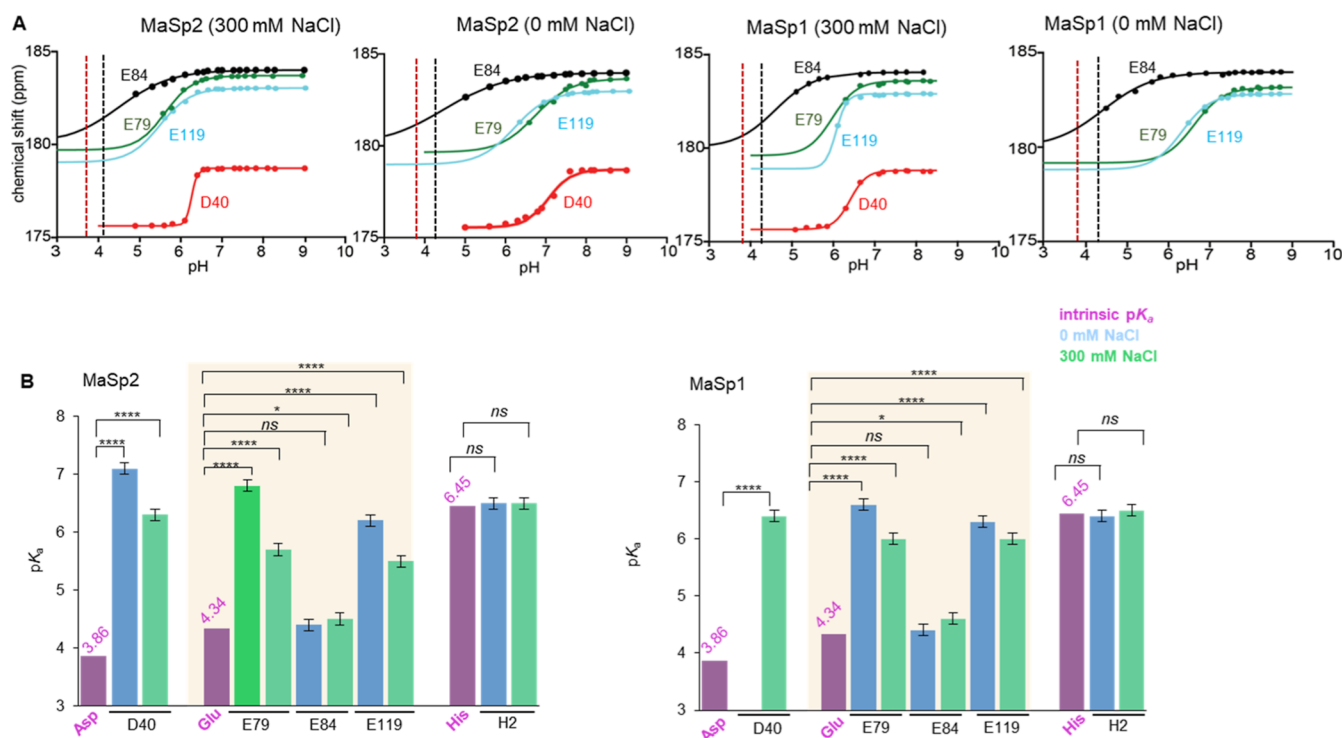


Figure 3. Experimental pK_a values of key MaSp1 and MaSp2 NTD residues obtained by NMR spectroscopy. (a) Titrations were performed in a pH range of 9–4 and under two different salt conditions (0 and 300 mM NaCl). Curve fitting to the chemical shift data was performed using the Henderson–Hasselbalch equation. The conserved residues D40, E79, E84, and E119 were probed, and the nonnative residue H2 (remains at the N-terminus after cleavage from tag purification) was used as a control. The vertical dotted lines show the intrinsic pK_a of the side chains of Asp (3.86, red) and Glu (4.34, gray).³⁷ (b) Comparison of the experimentally measured pK_a values with the intrinsic pK_a values. Unusually elevated pK_a values were detected for D40, E79, and E119 compared to the intrinsic values. In some cases, proper curve fitting was prevented because the line of experimental data was excessively broadened in the low pH range, which is indicated as ND. The data are arranged according to residue type. Statistically significant differences relative to the intrinsic pK_a value of each residue type are indicated by asterisks (**** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$; ns = not significant).

Table 1. Summary of the Estimated Individual pK_a Values of the Conserved Acidic NTD MaSp1 and NTD MaSp2 from *T. clavipes* by NMR Spectroscopy^a

residue	MaSp1 NTD ^{WT}				MaSp2 NTD ^{WT}			
	0 NaCl		300 mM NaCl		0 NaCl		300 mM NaCl	
	pK_a^{est}	n_{Hill}	pK_a^{est}	n_{Hill}	pK_a^{est}	n_{Hill}	pK_a^{est}	n_{Hill}
D40	N.D	N.D	6.4 ± 0.1	2.2 ± 0.2	7.1 ± 0.1	1.5 ± 0.2	6.3 ± 0.1	6.1 ± 0.3
E79	6.6 ± 0.1	1.2 ± 0.1	6.0 ± 0.1	1.5 ± 0.1	6.8 ± 0.1	1.0 ± 0.1	5.7 ± 0.1	1.3 ± 0.1
E84	4.4 ± 0.1	0.7 ± 0.1	4.6 ± 0.1	0.9 ± 0.1	4.4 ± 0.1	0.6 ± 0.1	4.5 ± 0.1	0.7 ± 0.1
E119	6.3 ± 0.1	1.3 ± 0.1	6.1 ± 0.1	3.4 ± 0.4	6.2 ± 0.1	1.0 ± 0.1	5.5 ± 0.1	1.2 ± 0.1
H2	6.4 ± 0.1	0.8 ± 0.1	6.5 ± 0.1	0.9 ± 0.1	6.5 ± 0.1	1.0 ± 0.1	6.5 ± 0.1	1.1 ± 0.1

^aND means not determined due to line broadening of NMR signals.

subunit (in this case R60 and K65) near acidic cluster 1. Notably, for both *T. clavipes* MaSp1 and MaSp2, position 60 is occupied by Lys (K60) instead of Arg, and additional basic residues occupy nearby positions as well (K54 and R58).

Experimental pK_a Measurements of Key Acidic Residues in WT NTD. Using solution-state NMR, we measured the individual side chain pK_a values of conserved acidic residues D40, E79, E84, and E119 for both MaSp1 and MaSp2 WT NTDs (NTD^{WT}), as these side chains are highly conserved and located at the dimer interface. Regarding residue D39, unfortunately, the side chain chemical shift of this residue overlaps with other aspartic acids (D18 and D54) (Figure S3). During the pH titration, we could not observe the chemical shift changes other than D40. Therefore, we consider

that D39 does not undergo protonation in the relevant pH range. Consistently, $C\beta$ chemical shift of residue D39 (BMRB ID: 18480) of the previously reported *E. australis* NTD dimer structure (PDB ID 2LTH) is 41.828 ppm,¹² which is a typical chemical shift of deprotonated aspartic acid.³⁷

The titration experiments were performed over a pH range of 9–5 (with intervals of 0.25 to 0.35 pH units) in the absence and presence of 300 mM NaCl, and the resulting chemical shift data were fitted against the Henderson–Hasselbalch equation (see the experimental section in Supporting Information). The chemical shifts of ¹³C γ and ¹³C δ were used as reporters for the pK_a values of Asp and Glu, respectively. To avoid overlapping signals, we performed 2D NMR experiments that correlated $C\gamma$ – $H\beta$ resonances for Asp and $C\delta$ – $H\gamma$ resonances for

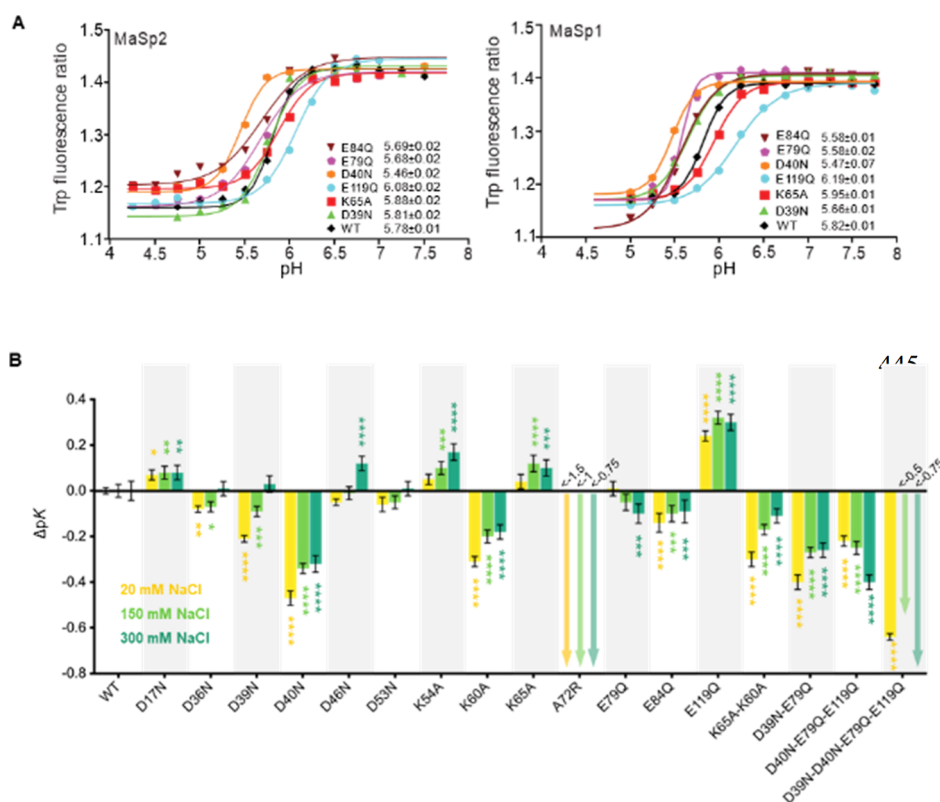


Figure 4. Investigation of pH-dependent dimerization of NTD *via* tryptophan fluorescence spectroscopy. (a) pH response of MaSp1 NTD and MaSp2 NTD as determined by changes in the fluorescence emission intensity ratios between 339 and 351 nm. The measured inflection points (pK) are provided (\pm SE), corresponding to the midpoint of the sigmoidal response curves. The results are shown for the WT NTD as well as for key variants (D39N, D40N, K65A, E79Q, E84Q, and E119Q) against a background of 150 mM NaCl. The full set of mutants and NaCl conditions tested are presented in Figure S4b. (b) Change in the pH-dependent dimerization midpoints for the MaSp2 NTD mutants relative to WT at three NaCl concentrations (20, 150, and 300 mM NaCl). The data are plotted as Δ pK on the y-axis and are calculated as $pK^{WT} - pK^{mut}$. For A72R and the D39N/D40N/E79Q/E119Q quadruple mutant, the values of some inflection points (and hence of Δ pK) could not be precisely determined, although minimum values can be estimated from the data and are indicated by arrows (e.g., for A72R at 150 mM NaCl, the inflection is at least 1 pH unit lower than that for the WT). Statistically significant differences relative to the respective WT value at each concentration are indicated by asterisks (**** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$).

Glu^{24,39} to follow the chemical shift changes as a function of pH (Figure S3). Figure 3a shows the data and the curve-fitting results, and the individual side chain pK_a values that were estimated are shown in Table 1, in which similar trends were observed for MaSp2 and MaSp1. The pK_a of D40 NTD MaSp1 in the absence of NaCl could not be determined since the D40 side chain chemical shift experienced severe exchange broadening due to monomer–dimer equilibrium and caused less data points, particularly for the fitting on the transition region. On the other hand, the side chain chemical shifts (Cδ–Hγ) of E79 and E119 are in full protonated form, could not be obtained due to the exchange broadening at acidic pH (Figure S4) and sample instability at pH less than 5. In this study, upon pH titration, the chemical shift changes ($\Delta\delta$) of those glutamates were more than 0.7 ppm. Considering, the chemical shift difference in the well-defined baseline of the deprotonation region is not more than 0.1 ppm, therefore, the chemical shift differences for more than 0.6 ppm in the transition region are sufficient for the fitting and estimate the pK_a values with assumption of $\Delta\delta$ as described previously.³⁷

Strikingly, residues D40, E79, and E119 displayed pK_a values that were considerably higher than the intrinsic pK_a values (reported as 3.86 and 4.34 for Asp and Glu, respectively) (Figure 3b).³⁷ The deviation was most prominent in D40, in which the experimentally measured pK_a value was approx-

imately 2.5–3 pH units higher than the intrinsic value (6.4 ± 0.1 for MaSp1 at 300 mM NaCl, 6.3 ± 0.1 for MaSp2 at 300 mM NaCl, and 7.1 ± 0.1 for MaSp2 at 0 NaCl). This result contrasts with a previous computational study, in which the pK_a of D40 was estimated to be approximately 4.1, that is, very close to the intrinsic value for Asp.¹⁸ For E79 and E119, on the other hand, the measured pK_a values were approximately 1–2 pH units higher than the intrinsic values, while for E84, located on the solvent-exposed helix 2–3 loop, the observed pK_a's (4.4–4.6) were very close to the intrinsic values. In agreement with our data, the previously reported Cβ chemical shift of D40 and Cγ chemical shifts of E79 and E119 of the *E. australis* NTD MaSp1 dimer at pH 5.5 are 39.312, 33.080, and 33.423 ppm (BMRB ID: 18480),¹² respectively, which are typical side chain chemical shifts of protonated aspartate and glutamate, as mentioned previously.³⁷

Furthermore, we found that the ¹⁵Nζ K65, K60, and ¹⁵Nε R58 chemical shifts of the MaSp2 NTD at neutral pH were approximately 33.5, 32.8, and 85.2 ppm, respectively (Figure S3), which are typical chemical shifts for protonated Lys and Arg, respectively.⁴⁰ However, the side chain Lys and Arg signals of the MaSp2 NTD at acidic pH could not be observed because the signals were broadened beyond detection. On the other hand, in the final NTD dimer structure (Figure 2g), the K65 side chain was in proximity to D40 (the D40–K65

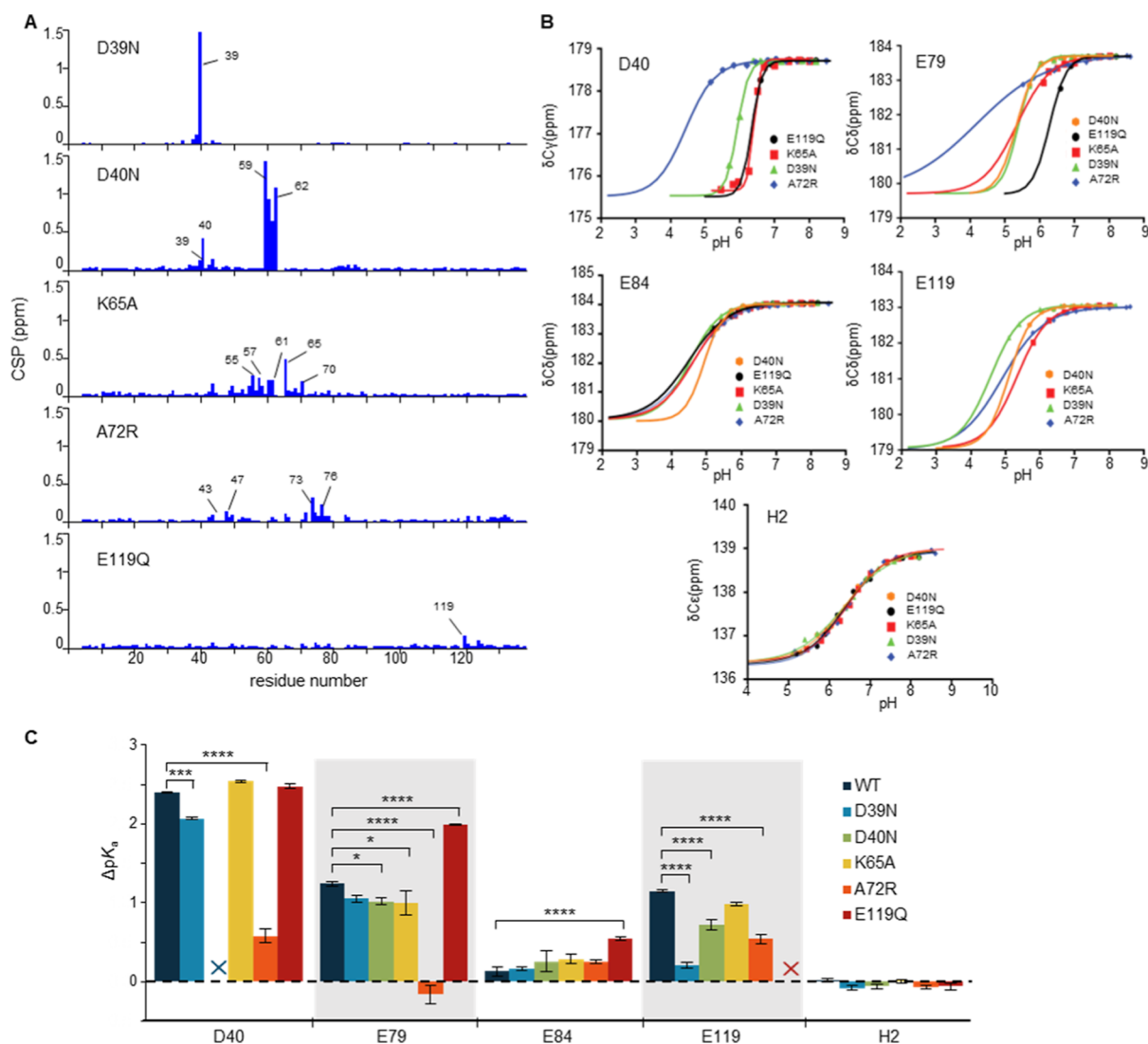


Figure 5. Probing effects of key NTD mutations on the conformation and pK_a of conserved acidic residues using NMR spectroscopy. (a) CSPs induced by specific mutations (at pH 7 in the presence of 300 mM NaCl). Combined chemical shift differences were calculated using eq 4 (see the Supporting Information). For the D39N and E119Q mutants, CSPs (>0.1 ppm) were localized around the mutated residues. For D40N, perturbations were observed around the mutated site, although the highest CSP values were found in the region around residues 59–62. For the K65A and A72R mutants, the effects were distributed across a wide range of residues that include the mutated site (although in the latter case, data could not be measured at position 72). (b,c) Experimental pK_a measurement for the conserved acidic residues of selected NTD mutants. (b) Experimental pK_a values for D40, E79, E84, E119, and H2 were measured for the MaSp2 NTD variants D39N, D40N, K65A, A72R, and E119Q at 300 mM NaCl. Chemical shift measurements against pH titration were carried out, and curve fitting was performed according to the Henderson–Hasselbalch equation. (c) Effect of MaSp2 NTD mutations on the experimental pK_a values of key acidic residues, which are expressed in terms of ΔpK_a , *i.e.*, the difference between measured and intrinsic pK_a values ($pK_a^{\text{exp}} - pK_a^{\text{int}}$). For the MaSp2 NTD residues with deviant pK_a values relative to the intrinsic values (notably D40, E79, and E119), the introduction of a mutation at a different site led to varying degrees of either reversion toward the intrinsic pK_a value (seen most prominently in D40 with the A72R variant) or a more pronounced deviation from the intrinsic value (*e.g.*, in E79 with the E119Q variant). As a control, H2 pK_a values are close to the intrinsic pK_a value of His in all variants. Statistically significant differences relative to the respective WT ΔpK_a value are indicated (**** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$).

distance is less than 4 Å),¹² suggesting that a salt-bridge interaction formed between the two residues.

Probing pH-Induced Dimerization via Tryptophan Fluorescence Spectroscopy. We carried out Trp fluorescence spectroscopy^{1,14} on NTD^{WT} and on a wide range of mutants to monitor the overall monomer-to-dimer equilibrium as a function of pH. The present data covered a wider array of residues than that of previous studies^{12,13,15} and under three different salt conditions. Figure 4a shows the spectra of a subset of MaSp2 and MaSp1 variants at 300 mM, while the entire set of curves is presented in Figure S5. We calculated the

pH transition midpoint (pK , or equilibrium point between monomer and dimer forms) by fitting a curve against the ratio of the fluorescence emission intensity ratio at 339 versus 351 nm (excitation = 280 nm). For WT MaSp1 and MaSp2, we observed a consistent trend of decreasing pK with increasing NaCl concentration (pK of 6.48 ± 0.03 , 6.06 ± 0.02 , and 5.82 ± 0.01 for MaSp1 NTD at 7.5, 150, and 300 mM NaCl, respectively, and 6.37 ± 0.01 , 5.98 ± 0.02 , 5.78 ± 0.03 for MaSp2 NTD at 20, 150, and 300 mM NaCl, respectively), indicating the stabilization of the monomer form of NTD with

increasing concentrations of NaCl, which is in line with previous findings.^{12,41}

In the mutational analysis, for the acidic residues, the carboxylate side chains were substituted with their non-titratable amide counterparts (Asp to Asn; Glu to Gln), whereas the basic residues (Arg or Lys) were replaced with uncharged Ala. We also analyzed the A72R mutant, which prevents dimer formation through steric and repulsive effects at the interface.¹⁴ Figure 4b presents the MaSp2 NTD variants in terms of ΔpK , which is a shift in the pH inflection point introduced by each mutation relative to the WT value. Among the acidic point mutations, D40N produced the largest change in the pH midpoint value, shifting the pK value by approximately -0.4 relative to that of the WT. As shown by the experimental pK_a , D40 undergoes the earliest protonation event in the acidic gradient among the different residues probed (Figure 3). Our results suggest that D40 plays a crucial role in maintaining the NTD dimerization response at the relatively high pH levels found in the silk spinning ducts. Interestingly, E79Q produced an overall decrease in the pH transition point (ΔpK of approximately -0.24 and -0.1 for MaSp1 and MaSp2, respectively, at 300 mM NaCl conditions), whereas the same mutation was reported to increase the midpoint pK of *E. australis* MaSp1,¹² which might reflect differences in the functions of NTD residues between spider species. Mutations at the other sites on the acidic cluster (D36N, D39N, and E84Q) exhibited only mild or negligible effects on the dimerization curves. Mutagenesis at the second acidic patch (D17N, D46N, and D53N) caused only slight perturbations in the pH response. On the other hand, the E119Q mutation produced a marked positive effect on ΔpK (at least $+0.3$ relative to WT at 300 mM NaCl), consistent with previous findings.¹² E119 is located away from the acidic clusters, and upon dimerization, the side chain inserts into a hydrophobic pocket on its binding partner, which could explain the ΔpK effects. Among single-point mutations of the basic residues, K60A yielded the most prominent results, with a ΔpK perturbation of approximately -0.2 . Notably, our K65A results diverged somewhat from those of previous reports; we observed a slight increase in ΔpK (approximately $+0.1$), whereas other studies showed a decrease in ΔpK .^{12,16} Interestingly, among the single point mutations, A72R produced an almost flat line for the MaSp2 data from pH 7 to 5, suggesting that this variant stabilized the monomer conformation. Below pH 5, a gradual fluorescence shift that was consistent with dimerization was observed, although insufficient data points in the lower pH range were obtained, thus precluding a precise calculation of pK values. A number of double, triple, and quadruple point mutations were also investigated (K60A/K65A, D39N/E79Q, D40N/E79Q/E119Q, and D39N/D40N/E79Q/E119Q), which produced significant downward shifts in ΔpK , although none of these mutations produced a perturbation greater than that of D40N alone (with the exception of the quadruple mutant), highlighting the important role of D40 in governing the pH-dependent dimerization mechanism.

Chemical Shift Perturbation and pK_a Determination of Selected Mutants of the MaSp2 NTD by NMR Spectroscopy. Since the effects of mutations on the structure and residue-specific pK_a 's of key amino acids remained unclear, we measured the chemical shift perturbation (CSP) at pH 7.0 (Figure 5a). We selected five mutants that targeted conserved sites on the MaSp2 NTD (D39N, D40N, K65A, A72R, and

E119Q) and investigated the effect of those mutants on the structures. The CSPs of E119Q and D39N were found to localize around the mutated amino acids, suggesting that no large structural changes or long-range effects were induced by these mutations. Interestingly, the CSP in D40N was localized not only around residue 40 but also around residues 59–62, which are located in the loop between helices 2 and 3 (and containing conserved basic residues K60), indicating that the D40 side chain is close to K60.

Since D40 is not proximal to residues 60 in the monomer 3D structure, the unusual CSP result suggests that the intermolecular contact between two residues (weak dimer formation) occurs even at pH 7 and in the presence of 300 mM NaCl. Conversely, variant K65A was found to affect a broad range of residues (from residue 44 to 65), although the effect is subtle, implying that the K65A mutant causes minor structural rearrangements that are distributed across a wide span of residues. Similarly, the A72R mutant caused wider effects, localized around helix 3, where residue 72 is found, as well as at residues 44 and 48; however, the low CSP values suggest only minor structural perturbations occurred.

To understand the effect of site-directed mutagenesis on the pK_a of key residues, we also measured the experimental pK_a 's of ionizable residues against point mutation backgrounds (Figure 5b,c). A previous study showed that neighboring acidic residues, which act as clusters, might cause elevated pK_a values.⁴² For D40, which exhibits an unusually high pK_a in WT NTD (~ 6.3 for MaSp2 at 300 mM NaCl), we found that the D39N mutation at the adjacent site led to a lower pK_a (~ 5.9) for D40, although the value was still considerably higher than the intrinsic pK_a . In contrast, the A72R mutation caused a drastic reduction in the D40 pK_a to ~ 4.4 , close to the intrinsic pK_a of Asp (3.86). Contrary to the D40N variant, Trp fluorescence data showed the highest pK value of E119Q compared to those of other mutants, which indicates that this variant stabilized the dimer conformation (Figure 5S). NMR data showed that the pK_a of D40 of the MaSp2 E119Q variant NTD is similar to the side chain pK_a of the WT D40 MaSp1 NTD, but the side chain pK_a of E79 of the E119Q variant is higher than that of the WT. This phenomenon occurs since E119 is surrounded by hydrophobic residues and buried in the hydrophobic interior; therefore, the mutation of Glu to Gln at position 119 is more favorable for the dimer interaction.

Origin of the Elevated pK_a Values of D40, E79, and E119 of NTD MaSp1 and MaSp2. The A72R mutation, which prevents dimer interaction, cause drastic reduction on the pK_a value of D40 to ~ 4.4 (close to the intrinsic pK_a value of Asp), implies that the unusually high pK_a value of D40 around neutral pH is originated from the salt bridge interaction between oxygen from the carboxyl group of residue D40 with proton from the counterpart lysine or arginine that is close in the space (K60/R60/K65). A similar phenomenon also occurred in the previously reported study in BPTI protein, where the salt bridge caused the pK_a at a rather basic pH.⁴³ Interestingly, we also found weak dimer interaction between D40 and K60 which exists, even at pH 7 in the presence of 300 mM NaCl, as shown by CSP (Figure 5). This phenomenon also explains the drastic reduction on pK value of D40N since the conversion of the carboxyl group in aspartate to the carbonyl group in asparagine hinders the salt-bridge interaction and stabilizes the monomer conformation (Figure 4).

Even though the salt-bridge interaction could not be observed in the *T. clavipes* NTD dimer structure due to the

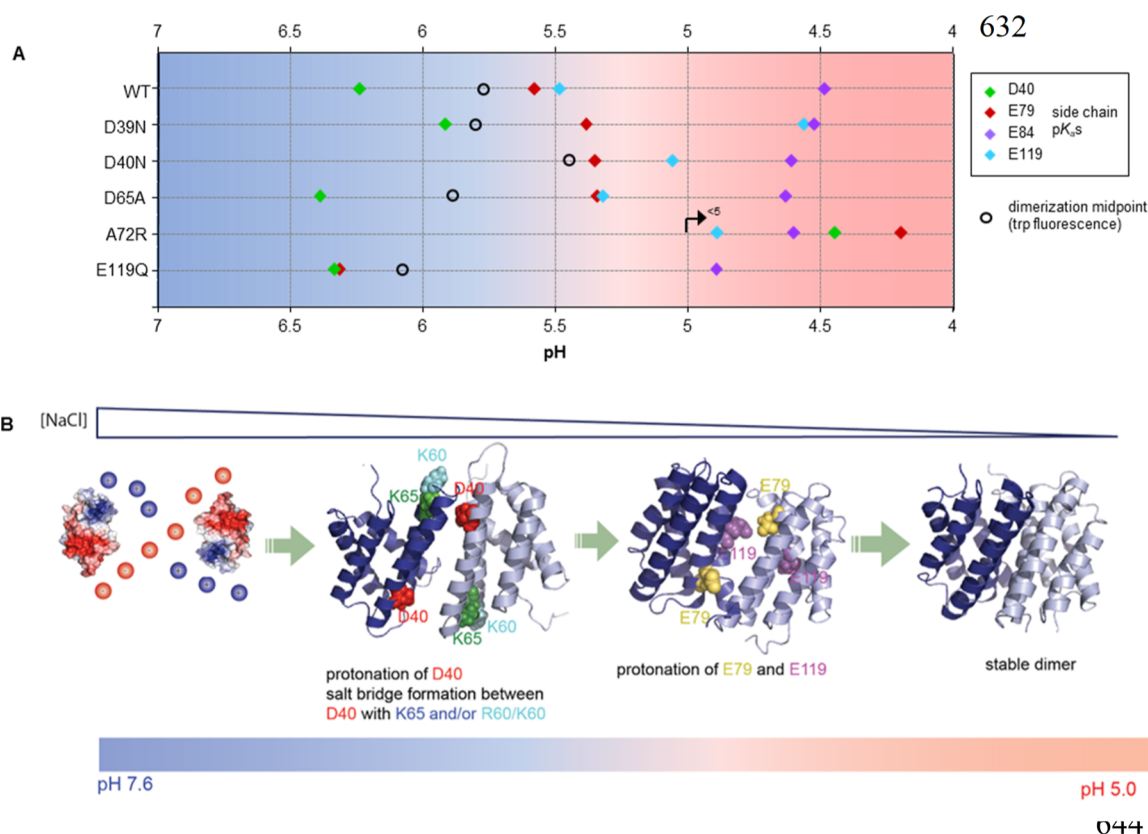


Figure 6. Proposed step-by-step dimerization mechanism of spidroin NTD. (a) Sequence of protonation events during MaSp NTD dimerization relative to the pH gradient. The x -axis corresponds to pH (neutral to acidic from left to right), along which the measured pK_a values (under 300 mM NaCl conditions) of key conserved acidic residues D40, E79, E84, and E119 are indicated for the MaSp2 NTD variant. Additionally, for each variant, the pH inflection point corresponding to the overall dimerization event as monitored by tryptophan fluorescence spectroscopy is indicated with black circles. Protonation of D40 along the pH gradient precedes the dimerization event and is likely crucial for the initial step, which locks the dimer in two positions. The protonation of E79 and E119 occurs downstream of the Trp-associated conformational change, suggesting that these residues stabilize the formation of the dimer structures. E84 undergoes protonation below pH 5 and thus a change in its ionization state is unlikely to play a role in the dimerization mechanism. (b) Distribution of acidic and basic residues on NTD produces positive and negative poles, which cause the NTD to be oriented in an antiparallel manner. As the NaCl concentration and pH decrease, dimerization will be initiated by protonation of D40, which possibly forms salt bridge interactions with K65 and/or R60/K60. This process is subsequently followed by the protonation of E79 and E119.

lack of the dimer structure; however, this finding is also supported by the fact that the *E. australis* NTD MaSp1 dimer structure distance between the side chain oxygen of D40 and the side chain proton (H_{ζ}) K65 of the *E. australis* NTD MaSp1 dimer (PDB ID: 2LTH) is 3.2 Å (Figure S6), which is qualified as salt bridge interaction for the distance less than 4 Å.⁴⁴

Once D40 is protonated and a salt bridge is formed with counterpart basic residues, the E79 and E119 residues will be buried at relatively hydrophobic interior environment. The subsequent protonation of those residues at elevated pH is required to stabilize the dimer interaction.

Our experimental study provides a direct glimpse into the protonation states of key residues in the MaSp1 and MaSp2 NTDs as a function of pH, which is essential to properly understand the mechanism behind the pH relay system during spider dragline silk self-assembly. Since the pK_a values of those conserved acidic residues are relevant to the pH gradient of the spider gland,⁴ this finding suggests that D40, E79, and E119 participate in the NTD dimerization mechanism. This study directly pinpointed the unusual pK_a value of D40 (~6.5–7.1), suggesting that protonation of D40 is important in the initial step to lock the dimer and the protonation of two glutamic

acid residues, E79 and E119, which have estimated pK_a values of ~5.4–6.2, are required to stabilize the NTD dimer. In contrast to previous studies, our NMR data showed that the pK_a value of E84 was ~4.4–4.6 (Table 1, Figure 5c), which is close to the intrinsic pK_a value of glutamic acid and is not affected by the salt concentration, suggesting that E84 is a solvent-exposed residue. This finding is also in line with the solvent exposed of E84 residue of previously reported NMR structure of a homologous monomer NTD MaSp1 (PDB ID: 2LPJ).¹⁴

Interestingly, some striking discrepancies were observed between the experimental measurements and the computational predictions (Table S2). Continuous constant pH molecular dynamics (CpHMD)¹⁸ and PROPKA⁴⁵ predicted only small perturbations in the pK_a of D40 (within -0.7 to +1.4 pH units relative to the intrinsic value) when applied to known NTD dimer structures, in contrast to the large shift measured by the NMR titration experiments (within +2.4 to +2.6 pH units relative to the intrinsic value). On the other hand, the two computational methods reported positive deviation pK_a values for E79 and E119 relative to the intrinsic pK_a value of glutamic acid, consistent with the experimental findings.

By integrating the results of the experimental pK_a measurements and the dimerization assays, we show that D40 protonation precedes the midpoint of the fluorescence curves that report on the conformational changes accompanying dimerization in response to the decreasing pH gradient (Figure 6a), therefore suggesting that the protonation of D40 represents an initial step in the pH response mechanism. On the other hand, the protonation of E79 and E119, while within the pH range found in the spinning ducts, is shown to occur after the fluorescence inflection point, suggesting that these residues perform stabilizing functions toward the formation of the final dimer structures.

The protonation of D40 at the beginning of the pH gradient seems to be necessary for maintaining the relatively high midpoint for dimerization of the NTD, whereas in the D40N mutant, the fluorescence assay inflection point shifts drastically lower, to approximately pH 5.5, suggesting that D40N stabilizes the monomer conformation of the NTD. This effect of the D40 mutation is consistent with previous reports,^{12,13,17} although the implications have been relatively elusive in previous models. As previously mentioned, the A72R mutation caused the NTD to stay in monomeric form from pH 7 to 5,¹⁴ thus, the fact that the A72R mutant decreased the pK_a of D40 drastically to a pK_a of D40 \sim 4.4 highlights the unusually high pK_a value of D40 in the WT MaSp2 NTD, originating from the interaction between D40 and its counterpart in the dimeric form of the NTD rather than from its neighboring acidic residues. This interpretation is also supported by the CSP results, where long-range perturbations were observed in the D40N mutant relative to the WT. However, we also found that the K65A mutation does not alter the pK_a of D40, which suggests that D40 might interact with multiple residues (such as K60 and possibly also with R58 in the case for *T. clavipes* MaSp1/MaSp2). In a similar manner, for residue E79, the A72R mutation leads to a large decrease in the pK_a (\sim 5.6 in the WT) to \sim 4.2, close to the intrinsic value for Glu side chains, and likely reflects the effects of intermolecular contacts.

Intriguingly, considering the important role of D40, E79, and E119, mutations of those residues (D40N–E79Q–E119Q) do not abolish dimer formation at pH 5, as shown by Trp fluorescence spectroscopy data (Figure S5). The mutations of D40N and E119Q have opposite effects. D40N stabilized the monomeric form, while E119Q stabilized the dimeric form. Comparison of the 2D ^1H – ^{15}N HSQC spectra of WT, D40N, E119Q, and D40N–E79Q–E119Q at pH 7, 6, and 5 demonstrates that the 2D spectrum of triple mutants D40N–E79Q–E119Q at pH 6 is still quite similar with 2D spectrum of D40N, suggesting that the resultant effect of the triple mutant is to stabilize the monomeric form at pH 6 (Figure S7). However, at pH 5, this triple mutant experiences exchange broadening due to monomer–dimer equilibrium. Furthermore, when quadruple mutations are introduced on cluster acidic residues (D40N–D39N–E79Q–E119Q), a greater effect on stabilizing monomer conformation is observed (Figure S5). This result implies that the ability of the NTD to form an antiparallel dimer is not limited to one key residue, but the entire acid and basic clusters function to generate positive and negative poles, which cause electrostatic interactions and lead to the formation of an antiparallel dimer of the NTD. This finding is also supported by the observation of accelerated association of NTD that is insensitive to charge screening in case of a dipole–dipole interaction.¹⁷ A decrease in the salt concentration, which is relevant to the condition in the

spinning duct of the spider gland,⁴⁶ can be interpreted as increasing the probability of the monomers encountering the correct orientation, thereby increasing the rate at which the monomer transforms into the dimer.

CONCLUSIONS

We propose the mechanism of pH-induced NTD dimerization (Figure 6b). Naturally, the distribution of acidic and basic residues on NTDs generates the positive and negative poles, which allow electrostatic interaction, thereby leading to an antiparallel orientation of the NTDs. As the salt concentration and pH decrease, dimerization is initiated by the protonation of D40, which forms an interaction with its basic counterparts (K65 and/or R60/K60). This is the initial step that locks the dimer in two positions. Subsequently, E79 and E119 are protonated to form a stable dimer since both residues are buried in the hydrophobic interior. Here, we also demonstrate that this molecular mechanism is conserved on MaSp1 and MaSp2 NTDs. In this study, we clarify the sequential step of NTD dimerization in both MaSp1 and MaSp2 NTDs by showing direct evidence based on experimental pK_a values obtained from individual amino acids. This pH-response NTD dimerization is connecting spidroins *via* intermolecular interactions to form rapid self-assembly of hierarchically spider silks and is thereby required for synthesizing strong artificial spider silk. This study also elucidates the molecular mechanism and introduces the possibility of making hybrid MaSp1/MaSp2 spider silks *via* NTD dimerization, which results in artificial spider silks with tunable properties.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biomac.2c01344>.

Secondary structure and dynamics of MaSp1 NTD and MaSp2 NTD in 10 mM phosphate buffer pH 7.0 and 300 mM NaCl; NMR structures of *T. clavipes* MaSp1 and MaSp2 NTD overlaid with previously reported structures; 2D NMR correlation spectra of side-chain NTD in 10 mM phosphate buffer pH 7 and 300 mM NaCl; overlay of NMR signals in different pH and salt concentrations; pH-dependent dimerization of NTD *via* tryptophan fluorescence spectroscopy; salt bridge interaction on the previously reported *E. australis* NTD MaSp1 dimer structure; overlay of 2D ^1H – ^{15}N HSQC spectra of NTD MaSp2 WT, D40N, E119Q, and triple mutants D40N–E79Q–E119Q at pH 7, 6, and 5; restraints and structural statistics of NTD MaSp1 and MaSp2 from *T. clavipes*; and comparison of the experimental pK_a values with *in silico* prediction methods (PDF)

AUTHOR INFORMATION

Corresponding Author

Keiji Numata – *Biomacromolecules Research Team, RIKEN Center for the Sustainable Resource Sciences, Wako, Saitama 351-0198, Japan; Department of Material Chemistry, Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan; Institute for Advanced Bioscience, Keio University, Yamagata 997-0017, Japan; orcid.org/0000-0003-2199-7420; Email: keiji.numata@riken.jp*

Authors

Nur Alia Oktaviani – Biomacromolecules Research Team, RIKEN Center for the Sustainable Resource Sciences, Wako, Saitama 351-0198, Japan

Ali D. Malay – Biomacromolecules Research Team, RIKEN Center for the Sustainable Resource Sciences, Wako, Saitama 351-0198, Japan

Akimasa Matsugami – Advanced NMR Application and Platform Team, NMR Research and Collaboration Group, NMR Science and Development Division, RIKEN SPring-8 Center, Yokohama, Kanagawa 230-0045, Japan

Fumiaki Hayashi – Advanced NMR Application and Platform Team, NMR Research and Collaboration Group, NMR Science and Development Division, RIKEN SPring-8 Center, Yokohama, Kanagawa 230-0045, Japan

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.biomac.2c01344>

Author Contributions

[†]N.A.O. and A.A.D.M. contributed equally as first authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Mami Goto for preparing the doubly labeled (¹³C and ¹⁵N) NTD MaSp2 D40N–E79Q–E119Q sample. We also thank the RIKEN SPDR fellowship, Japan Society for Promotion of Science (JSPS) Kakenhi for early career scientist for N.A.O. The study was also supported by Kakenhi Grant-In-Aid for Scientific Research (C), JSPS, and a Mitsubishi Foundation Research Grant in the Natural Sciences (A.A.D.M.). This work was also supported by grants from the ImPACT Program of Council for Science, Technology, and Innovation (Cabinet Office, Government of Japan), JST ERATO grant number JPMJER1602, Grant-in-Aid for Transformative Research Areas (B), and MEXT Program: Data Creation and Utilization-Type Material Research and Development Project grant number JPMXP1122714694 to K.N.

REFERENCES

- (1) Askarieh, G.; Hedhammar, M.; Nordling, K.; Saenz, A.; Casals, C.; Rising, A.; Johansson, J.; Knight, S. D. Self-Assembly of Spider Silk Proteins Is Controlled by a PH-Sensitive Relay. *Nature* **2010**, *465*, 236–238.
- (2) Hagn, F.; Eisoldt, L.; Hardy, J. G.; Vendrely, C.; Coles, M.; Scheibel, T.; Kessler, H. A Conserved Spider Silk Domain Acts as a Molecular Switch That Controls Fibre Assembly. *Nature* **2010**, *465*, 239–242.
- (3) Hijirida, D. H.; Do, K. G.; Michal, C.; Wong, S.; Zax, D.; Jelinski, L. W. ¹³C NMR of Nephila Clavipes Major Ampullate Silk Gland. *Biophys. J.* **1996**, *71*, 3442–3447.
- (4) Andersson, M.; Chen, G.; Otkovs, M.; Landreh, M.; Nordling, K.; Kronqvist, N.; Westermarck, P.; Jörnvall, H.; Knight, S.; Ridderstråle, Y.; Holm, L.; Meng, Q.; Jaudzems, K.; Chesler, M.; Johansson, J.; Rising, A. Carbonic Anhydrase Generates CO₂ and H⁺ That Drive Spider Silk Formation Via Opposite Effects on the Terminal Domains. *PLoS Biol.* **2014**, *12*, No. e1001921.
- (5) Gaines, W. A.; Sehorn, M. G.; Marcotte, W. R. Spidroin N-Terminal Domain Promotes a PH-Dependent Association of Silk Proteins during Self-Assembly. *J. Biol. Chem.* **2010**, *285*, 40745–40753.
- (6) Hronska, M.; van Beek, J. D.; Williamson, P. T. F.; Vollrath, F.; Meier, B. H. NMR Characterization of Native Liquid Spider Dragline Silk from Nephila Edulis. *Biomacromolecules* **2004**, *5*, 834–839.

(7) Jenkins, J. E.; Holland, G. P.; Yarger, J. L. High Resolution Magic Angle Spinning NMR Investigation of Silk Protein Structure within Major Ampullate Glands of Orb Weaving Spiders. *Soft Matter* **2012**, *8*, 1947–1954.

(8) Lefèvre, T.; Leclerc, J.; Rioux-Dubé, J.-F.; Buffeteau, T.; Paquin, M.-C.; Rousseau, M.-E.; Cloutier, L.; Auger, M.; Gagné, S. M.; Boudreault, S.; Cloutier, C.; Pézolet, M. Conformation of Spider Silk Proteins In Situ in the Intact Major Ampullate Gland and in Solution. *Biomacromolecules* **2007**, *8*, 2342–2344.

(9) Oktaviani, N. A.; Matsugami, A.; Malay, A. D.; Hayashi, F.; Kaplan, D. L.; Numata, K. Conformation and dynamics of soluble repetitive domain elucidates the initial β -sheet formation of spider silk. *Nat. Commun.* **2018**, *9*, 2121.

(10) Oktaviani, N. A.; Matsugami, A.; Hayashi, F.; Numata, K. Ion effects on the conformation and dynamics of repetitive domains of a spider silk protein: implications for solubility and β -sheet formation. *Chem. Commun.* **2019**, *55*, 9761–9764.

(11) Malay, A. D.; Suzuki, T.; Katashima, T.; Kono, N.; Arakawa, K.; Numata, K. Spider Silk Self-Assembly via Modular Liquid-Liquid Phase Separation and Nanofibrillation. *Sci. Adv.* **2020**, *6*, No. eabb6030.

(12) Kronqvist, N.; Otkovs, M.; Chmyrov, V.; Chen, G.; Andersson, M.; Nordling, K.; Landreh, M.; Sarr, M.; Jörnvall, H.; Wennmalm, S.; Widengren, J.; Meng, Q.; Rising, A.; Otzen, D.; Knight, S. D.; Jaudzems, K.; Johansson, J. Sequential PH-Driven Dimerization and Stabilization of the N-Terminal Domain Enables Rapid Spider Silk Formation. *Nat. Commun.* **2014**, *5*, 3254.

(13) Bauer, J.; Schaal, D.; Eisoldt, L.; Schweimer, K.; Schwarzinger, S.; Scheibel, T. Acidic Residues Control the Dimerization of the N-terminal Domain of Black Widow Spiders' Major Ampullate Spidroin 1. *Sci. Rep.* **2016**, *6*, 34442.

(14) Jaudzems, K.; Askarieh, G.; Landreh, M.; Nordling, K.; Hedhammar, M.; Jörnvall, H.; Rising, A.; Knight, S. D.; Johansson, J. pH-Dependent Dimerization of Spider Silk N-Terminal Domain Requires Relocation of a Wedged Tryptophan Side Chain. *J. Mol. Biol.* **2012**, *422*, 477–487.

(15) Otkovs, M.; Chen, G.; Nordling, K.; Landreh, M.; Meng, Q.; Jörnvall, H.; Kronqvist, N.; Rising, A.; Johansson, J.; Jaudzems, K. Diversified Structural Basis of a Conserved Molecular Mechanism for PH-Dependent Dimerization in Spider Silk N-Terminal Domains. *ChemBioChem* **2015**, *16*, 1720–1724.

(16) Atkison, J. H.; Parnham, S.; Marcotte, W. R.; Olsen, S. K. Crystal Structure of the Nephila Clavipes Major Ampullate Spidroin 1A N-Terminal Domain Reveals Plasticity at the Dimer Interface. *J. Biol. Chem.* **2016**, *291*, 19006–19017.

(17) Schwarze, S.; Zwettler, F. U.; Johnson, C. M.; Neuweiler, H. The N-Terminal Domains of Spider Silk Proteins Assemble Ultrafast and Protected from Charge Screening. *Nat. Commun.* **2013**, *4*, 2815.

(18) Wallace, J. A.; Shen, J. K. Unraveling A Trap-and-Trigger Mechanism in the PH-Sensitive Self-Assembly of Spider Silk Proteins. *J. Phys. Chem. Lett.* **2012**, *3*, 658–662.

(19) da Silva, F. L. B.; Pasquali, S.; Derreumaux, P.; Dias, L. G. Electrostatics Analysis of the Mutational and PH Effects of the N-Terminal Domain Self-Association of the Major Ampullate Spidroin. *Soft Matter* **2016**, *12*, 5600–5612.

(20) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Erik Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pK_a values in proteins. *Proteins* **2011**, *79*, 3260–3275.

(21) Kilambi, K. P.; Gray, J. J. Rapid Calculation of Protein pK_a Values Using Rosetta. *Biophys. J.* **2012**, *103*, 587–595.

(22) Heiby, J. C.; Goretzki, B.; Johnson, C. M.; Hellmich, U. A.; Neuweiler, H. Methionine in a Protein Hydrophobic Core Drives Tight Interactions Required for Assembly of Spider Silk. *Nat. Commun.* **2019**, *10*, 4378.

(23) Brooke, D.; Movahed, N.; Bothner, B. Universal Buffers for Use in Biochemistry and Biophysical Experiments. *AIMS Biophys.* **2015**, *2*, 336–342.

- (24) Oda, Y.; Yamazaki, T.; Nagayama, K.; Kanaya, S.; Kuroda, Y.; Nakamura, H. Individual Ionization Constants of All the Carboxyl Groups in Ribonuclease HI from *Escherichia Coli* Determined by NMR. *Biochemistry* **1994**, *33*, 5275–5284.
- (25) Oktaviani, N. A.; Otten, R.; Dijkstra, K.; Scheek, R. M.; Thulin, E.; Akke, M.; Mulder, F. A. A. 100% Complete Assignment of Non-Labile (1)H, (13)C, and (15)N Signals for Calcium-Loaded Calbindin D(9k) P43G. *Biomol. NMR Assignments* **2011**, *5*, 79–84.
- (26) André, I.; Linse, S.; Mulder, F. A. A. Residue-Specific pK_a Determination of Lysine and Arginine Side Chains by Indirect ¹⁵N and ¹³C NMR Spectroscopy: Application to Apo Calmodulin. *J. Am. Chem. Soc.* **2007**, *129*, 15805–15813.
- (27) Prompers, J. J.; Groenewegen, A.; Hilbers, C. W.; Pepermans, H. A. M. Two-Dimensional NMR Experiments for the Assignment of Aromatic Side Chains in ¹³C-Labeled Proteins. *J. Magn. Reson.* **1998**, *130*, 68–75.
- (28) Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. Two-Dimensional NMR Experiments for Correlating Carbon-13.β. and Proton.Δ. Chemical Shifts of Aromatic Residues in ¹³C-Labeled Proteins via Scalar Couplings. *J. Am. Chem. Soc.* **1993**, *115*, 11054–11055.
- (29) Oktaviani, N. A.; Malay, A. D.; Matsugami, A.; Hayashi, F.; Numata, K. Nearly Complete ¹H, ¹³C and ¹⁵N Chemical Shift Assignment of Monomeric Form of N-Terminal Domain of Nephila Clavipes Major Ampullate Spidroin 2. *Biomol. NMR Assignments* **2020**, *14*, 335–338.
- (30) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (31) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31*, 1325–1327.
- (32) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wuthrich, K. Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *J. Biomol. NMR* **1998**, *256*, 1–15.
- (33) Tamiola, K.; Mulder, F. A. A. Using NMR Chemical Shifts to Calculate the Propensity for Structural Order and Disorder in Proteins. *Biochem. Soc. Trans.* **2012**, *40*, 1014–1020.
- (34) Güntert, P.; Mumenthaler, C.; Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program Dyana. *J. Mol. Biol.* **1997**, *273*, 283–298.
- (35) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. The Xplor-NIH NMR Molecular Structure Determination Package. *J. Magn. Reson.* **2003**, *160*, 65–73.
- (36) Ferguson, F. M.; Dias, D. M.; Rodrigues, J. P. G. L. M.; Wien, H.; Boelens, R.; Bonvin, A. M. J. J.; Abell, C.; Ciulli, A. Binding Hotspots of BAZ2B Bromodomain: Histone Interaction Revealed by Solution NMR Driven Docking. *Biochemistry* **2014**, *53*, 6706–6716.
- (37) Platzer, G.; Okon, M.; McIntosh, L. P. pH-dependent random coil ¹H, ¹³C, and ¹⁵N chemical shifts of the ionizable amino acids: a guide for protein pK_a measurements. *J. Biomol. NMR* **2014**, *60*, 109–129.
- (38) Atkison, J. H.; Parnham, S.; Marcotte, W. R.; Olsen, S. K. Crystal Structure of the Nephila Clavipes Major Ampullate Spidroin 1A N-Terminal Domain Reveals Plasticity at the Dimer Interface. *J. Biol. Chem.* **2016**, *291*, 19006–19017.
- (39) Oktaviani, N. A.; Pool, T. J.; Yoshimura, Y.; Kamikubo, H.; Scheek, R. M.; Kataoka, M.; Mulder, F. A. A. Active-Site pK_a Determination for Photoactive Yellow Protein Rationalizes Slow Ground-State Recovery. *Biophys. J.* **2017**, *112*, 2109–2116.
- (40) André, I.; Linse, S.; Mulder, F. A. A. Residue-Specific PK_a Determination of Lysine and Arginine Side Chains by Indirect ¹⁵N and ¹³C NMR Spectroscopy: Application to Apo Calmodulin. *J. Am. Chem. Soc.* **2007**, *129*, 15805–15813.
- (41) Hagn, F.; Thamm, C.; Scheibel, T.; Kessler, H. pH-Dependent Dimerization and Salt-Dependent Stabilization of the N-terminal

Domain of Spider Dragline Silk-Implications for Fiber Formation. *Angew. Chem., Int. Ed.* **2011**, *50*, 310–313.

(42) Flocco, M. M.; Mowbray, S. L. Strange Bedfellows: Interactions between Acidic Side-Chains in Proteins. *J. Mol. Biol.* **1995**, *254*, 96–105.

(43) Lauritzen, C.; Skovgaard, O.; Erik Hansen, P. E.; Tüchsen, E. Effects of N-Terminal Extension Peptides on the Structure and Stability of Bovine Pancreatic Trypsin Inhibitor Studied by ¹H n.m.r. *Int. J. Biol. Macromol.* **1992**, *14*, 326–332.

(44) Kumar, S.; Nussinov, R. Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**, *3*, 604–617.

(45) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(46) Knight, D. P.; Vollrath, F. Changes in Element Composition along the Spinning Duct in a Nephila Spider. *Naturwissenschaften* **2001**, *88*, 179–182.

Recommended by ACS

Structural and Dynamical Properties of Elastin-Like Peptides near Their Lower Critical Solution Temperature

Tatiana I. Morozova, Jean-Louis Barrat, *et al.*

MARCH 06, 2023
BIOMACROMOLECULES

READ 

β-Sheet Structure Formation within Binary Blends of Two Spider Silk Related Peptides

Mirjam Hofmaier, Martin Müller, *et al.*

JANUARY 12, 2023
BIOMACROMOLECULES

READ 

Inviting C5-Trifluoromethylated Pseudoprolines into Collagen Mimetic Peptides

Anaïs Terrien, Emeric Miclet, *et al.*

FEBRUARY 14, 2023
BIOMACROMOLECULES

READ 

Soy Protein Amyloid Fibril Scaffold for Cultivated Meat Application

Zhengxun Wei, Ning Xiang, *et al.*

MARCH 14, 2023
ACS APPLIED MATERIALS & INTERFACES

READ 

Get More Suggestions >