Secondary structure of peptides mimicking the Gly-rich regions of major ampullate spidroin protein 1 and 2

Geoffrey M. Gray, Brittany Thiessen and Arjan van der Vaart*
Department of Chemistry, University of South Florida, 4202 E. Fowler Ave. CHE 205,
Tampa, FL 33620, U.S.A.

Abstract

Spider dragline silk has highly desirable material properties, possessing high extensibility, strength, and biocompatibility. Before it is spun, the constituent proteins are stored in a concentrated dope that is void of fibrils. To investigate the structural properties of the amorphous fiber regions in the dope, computer simulations were performed on model peptides representing the *N. clavipes* Gly-rich regions. Analysis of the secondary structure found predominantly turns, bends and coils; a small 3_1-helical population decreased with increasing concentration. Interestingly, the population of 3_1-helices saw a large increase in octanol. These results indicate that the unusual 3_1-helical secondary structure of the Gly-rich region of the fiber is a consequence of the spinning process, and that the low dielectric environment of the fiber may assist in favoring this structure.

* Corresponding author. Email: avandervaart@usf.edu. Phone: +1-813-974-8762.

Keywords: Spider silk, dope, secondary structure, 3_1-helices, simulation

Introduction

Spider dragline silk is remarkably tough due to a combination of high extensibility and strength [1-3]. Dragline silk fibers are predominantly composed of two large proteins, major ampullate spidroin 1 (MaSp1) and major ampullate spidroin 2 (MaSp2) [4, 5]. The sequence of these proteins consist of non-repetitive termini, which are separated by a highly repetitive core that makes up the majority (~90%) of the sequence. The repetitive core consists of blocks of 40 to 200 residues that are repeated between 20 and 100 times; each block consists of a 4 to 10 residue poly-Ala or poly-Gly-Ala region that is adjacent to a Gly-rich region [2]. In the fiber, the poly-Ala and poly-Gly-Ala regions form 2×5×7 nm crystallites consisting of antiparallel β-sheets [6, 7]. These crystallites are aligned to the fiber axis and embedded in

an amorphous matrix [1]. The crystallites are responsible for the fiber's strength, while the amorphous Gly-rich regions provide its elasticity [8]. In MaSp1 the Gly-rich regions consist of GGX motifs [4], and in MaSp2 they consist of GPGXX motifs [5] (where X=Ala, Gly, Leu, Gln, Tyr). In silk fibers, the GGX motifs adopt 3_1-helical structures, also known as polyproline II (PPII) helices [9-19], while the GPGXX motifs adopt type II β-turn structures [20] that collectively enables formation of β-spirals [21]. 3_1-helices were first identified in fibrous collagen [22], but are relatively rare in folded proteins [23]. They are present in the glycine-rich anaplastic lymphoma kinase [24] and snow-flea antifreeze protein [25], for example, but not all glycine-rich protein motifs form 3_1-helices [26-31]. In contrast to the more familiar α- and 3_{10}-helices, 3_1-helices do not form intramolecular hydrogen bonds. They possess 3-fold symmetry, with three residues per turn, and their carbonyl oxygen atoms point away from the helical axis.

Before spider silk is spun into a fiber by a specialized organ (spinneret), the constituent proteins are stored in a highly concentrated solution in an abdominal gland [32]. Depending on the species, this gland consists of two or three different zones [33]. The A zone at the start of the gland is neutral or slightly alkaline and secretes and stores MaSp1 and MaSp2, while the other zones towards the spinnerets are acidic, secrete protein coating, and have a role in preprocessing of the fiber [34]. At concentrations that range between 30-50% [35-37], the proteins are believed to form micellar structures in the A zone, that congregate to form liquid crystal intermediates [38-40]. This viscous, gel-like solution is void of fibrils, through an unknown mechanism that is controlled by the pH [34, 39].

Insights into the protein microstructure in the A zone dope would aid the understanding of this mechanism, but so far, experimental data is somewhat conflicting. Early NMR, FTIR and CD measurements indicated that the dope lacks β-sheets, and consists of a dynamical ensemble of loosely formed α and 3_1-helices [37], while other CD measurements indicated a structure poor in α-helices and β-sheets [41]. CD indicated that the proteins are in a random coil conformation with some β-sheet structure in the gland [34], and a random coil in aqueous solution [42, 43]. NMR indicated that the dope is dynamically disordered without well-defined secondary structure [44-46] with fast backbone dynamics on the sub-nanosecond time scale [45]. Raman confocal spectromicroscopy and vibrational CD indicated a mixture of random coil and 3_1-helical segments, with some α-helical structures [47, 48]. Other Raman spectromicroscopy experiments indicated that the GGX and GPGXX motifs have indistinguishable conformations in the dope, which both form a disordered random coil with some α-helical contribution [49].

Here, we present molecular dynamics (MD) simulations to further assess the structural propensities of several peptide sequence motifs of the Gly-rich regions of *Nephila clavipes* MaSp1 and MaSp2. Five peptide sequences with GGX and GPGXX motifs from the amorphous region of *N. clavipes* were studied at differing concentrations to assess the inherent propensity for 3_1-helix formation under dope-like conditions. In order to fully sample the conformational space, simulations were performed with the multiple scaling replica exchange (MREST) method [50, 51]. This method combines replica exchange [52] with a temperature-scaled potential for the solvent-solvent and solvent-protein interactions [53] and Tsallis biasing [54] of all degrees of freedom for fast convergence. Simulations were also performed in octanol to mimic a low dielectric environment, as would be encounted in the fiber.

Methods

A total of five peptide sequences representative of the Gly-rich regions in MaSp1 [4] and MaSp2 [5] of *Nephila clavipes* were studied (Table 1). The MaSp1 sequences will be indicated by 1a-c, the MaSp2 sequences by 2a-b. Systems consisting of one strand and three strands per box were simulated in explicit water [55] using a solvent layer of 20 Å. In order to fully sample conformational space in reasonable time scales, sampling was performed with the MREST enhanced sampling method [50, 51]; details of this method can be found in the original paper [50]. The MREST simulations were performed with the CHARMM program [56], using six replicas with temperature and Tsallis q factors of (T, q) = (300 K, 1), (332 K, 1.00001), (368 K, 1.00002), (407 K, 1.00003), (451 K, 1.00004), and (500 K, 1.00005) for the single stranded systems, and 8 replicas with (T, q) = (300 K, 1), (312 K, 1.000010), (325 K, 1.000020), (339 K, 1.000030), (353 K, 1.000040), (368 K, 1.000050), (383 K, 1.000055), and (400 K, 1.000060) for the triple stranded systems. Systems were first heated in the NPT ensemble from 150 to 300 K in 50 K increments over a period of 400 ps, further heated (using 10 K increments of 100 ps) to the desired temperature, and equilibrated in the NVT ensemble for 2 ns. After a further equilibration using MREST for 1 ns, production runs of 40 ns per replica were performed. Swapping attempts were made every 5 ps. Simulations were run until the distribution of measured variables (backbone root mean square deviation (RMSD), secondary structure, and Ramachandran plots) did not change with additional simulation time; convergence was further tested by assessing swapping rates, round-trips of the entire (T,q) ladder for each replica, transitions between different structural clusters, and correlation times.

- Table 1 here -

The low dielectric environment inside a fiber was mimicked by simulations of a single strand in octanol [57]. The setup of this system was similar to the water solvated systems, except that 10 replicas were used with (T,q) = (300 K, 1), (317 K, 1.000025), (336 K, 1.000030), (355 K, 1.000040), (376 K, 1.000045), (398 K, 1.000050), (421 K, 1.000055), (446 K, 1.000065), (472 K, 1.000075), and (500 K, 1.000080).

The CHARMM 36 force field [58] was used for all simulations. To test whether this force field correctly handled 3_1-helices, a simulation of the 3_1-rich snow-flea antifreeze protein in the NPT ensemble was performed. The starting structure was taken from PDB ID 2PNE [25]. A solvent layer of 20 Å of TIP3 [55] water was added beyond the protein. The system was then minimized and heated with backbone restraints of 5 kcal/mol from 150 K to 300 K in 50 K increments over 1 ns each. The restraints were gradually released from 5 kcal/mol, 2.5 kcal/mol, 1 kcal/mol, 0.5 kcal/mol, 0.1 kcal/mol over 2 ns simulations each, followed by a 2 ns unrestrained equilibration and a 100 ns production run using OpenMM [59].

All simulations were performed with periodic boundary conditions using Langevin dynamics [60] and a Monte Carlo barostat for the NPT simulations [61]. A 12 Å cutoff was used for nonbonded interaction with a switching potential. Long-range electrostatics were treated using the particle-mesh Ewald method [62]. Bonds with hydrogen atoms were constrained using SHAKE [63], allowing a timestep of 0.002 ps. Snapshots were saved every 5 ps. Secondary structure was calculated using DSSP [64]. Since 3_1-helices are not recognized as a secondary structure by this method, an internal script was used to calculate the angle between adjacent C=O backbone vectors. Angles of 120° ± 20° were then classified as 3_1-helices. Clustering based on backbone RMSD was performed on all 300 K snapshots using the Art-2 algorithm in CHARMM with a radius of 1.5 Å and a maximum error of 10^{-4} [65].

Results

Because of the relative scarcity of 3_1-helices in known protein structures, and issues with simulating 3_1-helices in implicit solvent [13], MD simulations in explicit water were performed to test the ability of the CHARMM 36 force field to replicate this structural motif. For this purpose, the snow flea antifreeze protein, a protein with a high content of 3_1-helices, was simulated (Fig. 1). The protein was stable throughout the simulation and retained its secondary structure without experiencing significant structural changes. A small spike in RMSD was observed ~75 ns, which corresponded to a twist in the structure. The average fraction of residues in 3_1-helical conformation was 27%, which agreed well with the fraction

in the crystal structure (25%). These results show that the CHARMM 36 force field is able to replicate this structural motif with high fidelity.

- Figure 1 here -

Convergence of the MaSp1/2 Gly-rich motifs simulations was tested in a variety of ways. Each replica visited all (T, q) values within 1 ns of sampling, and round-trip traveling across the entire (T, q) ladder occurred for each replica in all simulations. Structural clustering of the trajectories showed frequent transitions between clusters and that each cluster was visited multiple times. Structures transitioned between clusters continuously throughout the simulation. Autocorrelation times for the backbone RMSD using the first frame as a reference were 1 ns for the single stranded systems in water, 1.5 ns in octanol, and 2 ns for the triple stranded systems in water.

Ramachandran plots are shown in Fig. 2. Right-handed 3_1-helices are typically associated with the region between (φ, ψ) = (-60°, 150°) to (-90°, 135°) [9, 12, 14-18]; however, this region of the Ramachandran plot overlaps with the region for β-turns [13]. The single stranded 1a-1c peptides had similar Ramachandran plots in water, with major populations centered around (-60°, 150°). Small populations were observed around the helical regions (corresponding to α-helices and 3_{10}-helices) of the Ramachandran plots as well, centered at (-60°, -150°) and (60°, ±150°). The single stranded 2a-2b peptides also had the largest population centered at (-60°, 150°) in water. However, the shape of this population differed from the 1a-1c systems, with additional density observed around (-50°, 180°). Populations at (60°, ±150°) were greatly reduced in 2a and reduced in 2b, which is likely due to the presence of Pro residues.

- Figure 2 here -

All triple stranded systems showed peaks around (-60°, 150°), as well as small populations in the α/3_{10}-helical region. 1a had a sizeable population around (60°, 150°), which was significantly reduced in the other systems. Peaks around (-60°, -180°) were also observed, and these populations were connected to the (-60°, 150°) populations. Several differences between the Ramachandran plots for the single and triple stranded systems in water were observed. For 1a-1c, the increase in concentration reduced the population of left-handed structures. This was evidenced by the loss of density at (60°, 150°). This loss was also observed for 2b. Increasing concentration also reduced the population in the α/3_{10}-helical regions.

Since the dihedral angles of 3_1-helices and β-turns overlap [13], the presence of 3_1-helices and other structural elements was further quantified by analysis of the carbonyl angles and DSSP (Fig. 3). This revealed high propensities for turns/bends and coils for the single stranded systems in water, with relatively small populations of 3_1-helices. Very few β-sheet and α-helices were found. Overall, 1a-1c had ~45% turns/bends, 45% coils and ~7% 3_1-helices. Typically, structural formation was local and did not result in the formation of large structural motifs. The distribution of these structural motifs was peaked around 2 residues for all structures, with tails extending to as many as 8 residues. Distributions for bends were the broadest, followed by turns and 3_1-helices. Distributions for 3_1-helices were very similar between sequences and ~90% of these helices were 2 residues in length. A similar percentage of turns/bends was seen for 2a-2b (45%), while its overall coil was slightly reduced (42%) and its 3_1-helical contents was slightly larger (11%) than 1a-1c. These differences are likely due to the presence of proline in peptides 2a-b, which is the only amino acid with a strong 3_1-helical propensity [23]. Of note was a large 3_1-helical propensity for the QQ motif of 2a (~45%) and the GPG motif of 2b (~40%).

For the aqueous triple stranded systems the 3_1-helical population was reduced. For 1a-c this reduction was modest, to 5%, but for 2a-b the reduction was severe (to 3%). These reductions are likely due to interstrand interactions. Coils comprised 35%, while turns/bends were were 55%. While $\alpha/3_{10}$-helices did not form, there was a slight increase in β-sheets. This was more pronounced for 2a-2b (7%), than for 1a-1c (5%). Compared to the single stranded systems, the triple stranded systems were relatively more compact and had less structural variability. Their RMSD distributions were shifted by 0.15 Å to lower values, and their radius of gyration were generally reduced as well.

- Figure 3 here -

Instead of maintaining their secondary structure throughout the simulations, the peptides frequently transitioned between different structural motifs. This observation indicated that the model peptides formed a highly dynamic ensemble, and is in agreement with NMR measurements of the dope [44-46]. High disorder, low α-helical and β-sheet propensity of the aqueous peptides also agreed with experimental findings of the dope [34, 37, 41-49], but the lack of experimental high resolution structural data complicated a detailed comparison.

In octanol the Ramachandran plot for systems 1a-1b showed high density in the (-60°, -45°) range, which corresponds to α-helices (Fig. 2). Both also had populations around (40°, 20°), suggesting the presence of left-handed 3_{10}-helices. Small populations around (-60°, 150°) were observed for 1b. 1c and 2a had

shifted helical populations closer to (-40°, 0°), indicating greater 3_{10}-helical content. 1c also had a left-handed helical population, and a large population at (-60°, 150°) and (-90°, 135°), indicative of right-handed 3_1-helices. Similar populations were observed in 2a, while a population at (-60°, 150°) was observed in 2b. 2b had small populations in the $\alpha/3_{10}$-helical regions of the plot, with larger populations at (-60°, 180°) and (60°, 150°). Further secondary structure analysis confirmed that the peptides were more α-helical in octanol (Fig. 3). 1a and 1b had ~12 α-helix content, while 1c had ~5%. 1c had similar coil, turns/bends and β-sheet content as in water. 1a-c had dramatic increases in 3_1-helices: a four-fold increase over water for 1a-b to ~25%, and a two-fold for 1c to 12%. Nearly all residues of 1a showed significant 3_1-helical propensity, while the propensity was especially large for the GAGQ motif of 1b and GLG of 1c. The overall 3_1-helical propensity was similar in octanol and water for 2a, but double in octanol for 2b (20%). Especially the PGG motif of 2b showed large propensity (over 50%). The GPG motif of 2b was mostly in the bend conformation (80%). The RMSD distributions of 2a and 2b were shifted to slightly lower values than the single stranded water systems, and the distributions widths were reduced. The radii of gyration were also reduced with lower widths compared to the single stranded systems in water, while the distributions had higher widths than for the triple stranded systems in water. More hydrogen bonding was observed in octanol than in water, with broader distributions centered around 0.3-0.5 hydrogen bonds per residue.

Conclusion

Simulations of the snow-flea antifreeze protein demonstrated that the force field used in this study was able to accurately replicate 3_1-helical secondary structures. Simulations of peptides corresponding to the Gly-rich regions of MaSp1 and MaSp2 showed that these peptides predominantly form coils, turns and bends in water. 3_1-helices were found at a small fraction, and this fraction decreased with increasing peptide concentration. This data indicates that the Gly-rich regions of MaSp1 and MaSp2 mostly form coils, bends and turns in the aqueous dope. These regions have little inherent propensity toward the 3_1-helix, and this propensity does not increase with concentration.

These observations therefore suggest that formation of 3_1-helices in the fiber is a result of the spinning process. While spinning involves a complex and poorly understood interplay of chemical and mechanical factors, our simulations identified a contribution that may aid the formation of 3_1-helices. We observed that 3_1-helices were formed significantly more in octanol than in water, and in octanol the 3_1-helices contributed to a significant fraction of the overall structure. These findings indicate that a lower dielectric environment

favors 3_1-helices, and imply that the low dielectric environment of the spider silk fiber may directly contribute to the formation and stabilization of these unusual helices.

Acknowledgments

This project was supported through a University of South Florida Nexus Initiative (UNI) Award, as well as the National Science Foundation (grant number MCB-1919096). Computer time was provided by USF Research Computing.

References

[1] J.L. Yarger, B.R. Cherry, A. van der Vaart, Uncovering the structure-function relationship in spider silk, Nature Reviews Materials, 3 (2018) 8000.
[2] R.V. Lewis, Spider silk: Ancient ideas for new biomaterials, Chemical Reviews, 106 (2006) 3762-3774.
[3] D. Ebrahimi, O. Tokareva, N.G. Rim, J.Y. Wong, D.L. Kaplan, M.J. Buehler, Silk-Its Mysteries, How It Is Made, and How It Is Used, Acs Biomaterials Science & Engineering, 1 (2015) 864-876.
[4] M. Xu, R.V. Lewis, Structure of a protein superfiber - spider dragline silk, Proceedings of the National Academy of Sciences of the United States of America, 87 (1990) 7120-7124.
[5] M.B. Hinman, R.V. Lewis, Isolation of a clone encoding a 2nd dragline silk fibroin - Nephila-clavipes dragline silk is a 2-protein fiber, Journal of Biological Chemistry, 267 (1992) 19320-19324.
[6] D.T. Grubb, L.W. Jelinski, Fiber morphology of spider silk: The effects of tensile deformation, Macromolecules, 30 (1997) 2860-2867.
[7] C. Riekel, C. Branden, C. Craig, C. Ferrero, F. Heidelbach, M. Muller, Aspects of X-ray diffraction on single spider fibers, International Journal of Biological Macromolecules, 24 (1999) 179-186.
[8] Y. Termonia, Molecular modeling of spider silk elasticity, Macromolecules, 27 (1994) 7378-7381.
[9] G.P. Holland, M.S. Creager, J.E. Jenkins, R.V. Lewis, J.L. Yarger, Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy, Journal of the American Chemical Society, 130 (2008) 9871-9877.
[10] G.P. Holland, J.E. Jenkins, M.S. Creager, R.V. Lewis, J.L. Yarger, Quantifying the fraction of glycine and alanine in beta-sheet and helical conformations in spider dragline silk using solid-state NMR, Chemical Communications, (2008) 5568-5570.

[11] J.D. van Beek, S. Hess, F. Vollrath, B.H. Meier, The molecular structure of spider dragline silk: Folding and orientation of the protein backbone, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002) 10266-10271.
[12] T. Asakura, M.Y. Yang, T. Kawase, Structure of characteristic sequences in Nephila clavipes dragline silk (MaSp1) studied with C-13 solid state NMR, Polymer Journal, 36 (2004) 999-1003.
[13] G.M. Gray, A. van der Vaart, C.C. Guo, J. Jones, D. Onofrei, B.R. Cherry, R.V. Lewis, J.L. Yarger, G.P. Holland, Secondary Structure Adopted by the Gly-Gly-X Repetitive Regions of Dragline Spider Silk, International Journal of Molecular Sciences, 17 (2016) 2023.
[14] T. Asakura, M.Y. Yang, T. Kawase, Y. Nakazawa, C-13 solid-state NMR study of structural heterogeneity in peptides containing both polyalanine and repeated GGA sequences as a local structural model of Nephila clavipes dragline silk (Spidroin 1), Macromolecules, 38 (2005) 3356-3363.
[15] J. Ashida, K. Ohgo, K. Komatsu, A. Kubota, T. Asakura, Determination of the torsion angles of alanine and glycine residues of model compounds of spider silk (AGG)(10) using solid-state NMR methods, Journal of Biomolecular Nmr, 25 (2003) 91-103.
[16] J.E. Jenkins, S. Sampath, E. Butler, J. Kim, R.W. Henning, G.P. Holland, J.L. Yarger, Characterizing the Secondary Protein Structure of Black Widow Dragline Silk Using Solid-State NMR and X-ray Diffraction, Biomacromolecules, 14 (2013) 3472-3483.
[17] J. Kummerlen, J.D. van Beek, F. Vollrath, B.H. Meier, Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance, Macromolecules, 29 (1996) 2920-2928.
[18] I. Marcotte, J.D. van Beek, B.H. Meier, Molecular disorder and structure of spider dragline silk investigated by two-dimensional solid-state NMR spectroscopy, Macromolecules, 40 (2007) 1995-2001.
[19] M.Y. Yang, Y. Nakazawa, K. Yamauchi, D. Knight, T. Asakura, Structure of model peptides based on Nephila clavipes dragline silk spidroin (MaSp1) studied by C-13 cross polarization/magic angle spinning NMR, Biomacromolecules, 6 (2005) 3220-3226.
[20] J.E. Jenkins, M.S. Creager, E.B. Butler, R.V. Lewis, J.L. Yarger, G.P. Holland, Solid-state NMR evidence for elastin-like beta-turn structure in spider dragline silk, Chemical Communications, 46 (2010) 6714-6716.
[21] C.Y. Hayashi, R.V. Lewis, Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks, Journal of Molecular Biology, 275 (1998) 773-784.
[22] M.D. Shoulders, R.T. Raines, Collagen structure and stability, Annu Rev Biochem, 78 (2009) 929-958.
[23] A.A. Adzhubei, M.J. Sternberg, A.A. Makarov, Polyproline-II helix in proteins: structure and function, J Mol Biol, 425 (2013) 2100-2132.

[24] T. Li, S.E. Stayrook, Y. Tsutsui, J. Zhang, Y. Wang, H. Li, A. Proffitt, S.G. Krimmer, M. Ahmed, O. Belliveau, I.X. Walker, K.C. Mudumbi, Y. Suzuki, I. Lax, D. Alvarado, M.A. Lemmon, J. Schlessinger, D.E. Klein, Structural basis for ligand reception by anaplastic lymphoma kinase, Nature, 600 (2021) 148-152.
[25] B.L. Pentelute, Z.P. Gates, V. Tereshko, J.L. Dashnau, J.M. Vanderkooi, A.A. Kossiakoff, S.B.H. Kent, X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers, Journal of the American Chemical Society, 130 (2008) 9695-9701.
[26] B.P. Tripet, K.E. Mason, B.J. Eilers, J. Burns, P. Powell, A.M. Fischer, V. Copié, Structural and biochemical analysis of the Hordeum vulgare L. HvGR-RBP1 protein, a glycine-rich RNA-binding protein involved in the regulation of barley plant development and stress response, Biochemistry, 53 (2014) 7945-7960.
[27] G.E. González-Páez, D.W. Wolan, Ultrahigh and high resolution structures and mutational analysis of monomeric Streptococcus pyogenes SpeB reveal a functional role for the glycine-rich C-terminal loop, J Biol Chem, 287 (2012) 24412-24426.
[28] R. Fukunaga, S. Yokoyama, Structure of the AlaX-M trans-editing enzyme from Pyrococcus horikoshii, Acta Crystallogr D Biol Crystallogr, 63 (2007) 390-400.
[29] D. Sarkar, I. Chakraborty, M. Condorelli, B. Ghosh, T. Mass, M. Weingarth, A.K. Mandal, C. La Rosa, V. Subramanian, A. Bhunia, Self-Assembly and Neurotoxicity of β-Amyloid (21-40) Peptide Fragment: The Regulatory Role of GxxxG Motifs, ChemMedChem, 15 (2020) 293-301.
[30] K. Saito, T. Kigawa, S. Koshiba, K. Sato, Y. Matsuo, A. Sakamoto, T. Takagi, M. Shirouzu, T. Yabuki, E. Nunokawa, E. Seki, T. Matsuda, M. Aoki, Y. Miyata, N. Hirakawa, M. Inoue, T. Terada, T. Nagase, R. Kikuno, M. Nakayama, O. Ohara, A. Tanaka, S. Yokoyama, The CAP-Gly domain of CYLD associates with the proline-rich sequence in NEMO/IKKgamma, Structure, 12 (2004) 1719-1728.
[31] I. Chakraborty, R.K. Kar, D. Sarkar, S. Kumar, N.C. Maiti, A.K. Mandal, A. Bhunia, Solvent Relaxation NMR: A Tool for Real-Time Monitoring Water Dynamics in Protein Aggregation Landscape, ACS Chem Neurosci, 12 (2021) 2903-2916.
[32] D.B. Peakall, Synthesis of silk mechanism and location, American Zoologist, 9 (1969) 71-&.
[33] M. Andersson, L. Holm, Y. Ridderstrale, J. Johansson, A. Rising, Morphology and Composition of the Spider Major Ampullate Gland and Dragline Silk, Biomacromolecules, 14 (2013) 2945-2952.
[34] C. Dicko, F. Vollrath, J.M. Kenney, Spider silk protein refolding is controlled by changing pH, Biomacromolecules, 5 (2004) 704-710.
[35] X. Chen, D.P. Knight, F. Vollrath, Rheological characterization of Nephila spidroin solution, Biomacromolecules, 3 (2002) 644-648.

[36] P.R. Laity, S.E. Gilks, C. Holland, Rheological behaviour of native silk feedstocks, Polymer, 67 (2015) 28-39.
[37] D.H. Hijirida, K.G. Do, C. Michal, S. Wong, D. Zax, L.W. Jelinski, C-13 NMR of Nephila clavipes major ampullate silk gland, Biophysical Journal, 71 (1996) 3442-3447.
[38] H.J. Jin, D.L. Kaplan, Mechanism of silk processing in insects and spiders, Nature, 424 (2003) 1057-1061.
[39] A.A. Walker, C. Holland, T.D. Sutherland, More than one way to spin a crystallite: multiple trajectories through liquid crystallinity to solid silk, Proceedings of the Royal Society B-Biological Sciences, 282 (2015) 20150259.
[40] L.R. Parent, D. Onofrei, D. Xu, D. Stengel, J.D. Roehling, J.B. Addison, C. Forman, S.A. Amin, B.R. Cherry, J.L. Yarger, N.C. Gianneschi, G.P. Holland, Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks, Proceedings of the National Academy of Sciences of the United States of America, 115 (2018) 11507-11512.
[41] J.M. Kenney, D. Knight, M.J. Wise, F. Vollrath, Amyloidogenic nature of spider silk, European Journal of Biochemistry, 269 (2002) 4159-4163.
[42] Z.Z. Shao, F. Vollrath, Y. Yang, H.C. Thogersen, Structure and behavior of regenerated spider silk, Macromolecules, 36 (2003) 1157-1161.
[43] C. Dicko, D. Knight, J.M. Kenney, F. Vollrath, Structural conformation of spidroin in solution: a synchrotron radiation circular dichroism study, Biomacromolecules, 5 (2004) 758-767.
[44] M. Hronska, J.D. van Beek, P.T.F. Williamson, F. Vollrath, B.H. Meier, NMR characterization of native liquid spider dragline silk from Nephila edulis, Biomacromolecules, 5 (2004) 834-839.
[45] D. Xu, J.L. Yarger, G.P. Holland, Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy, Polymer, 55 (2014) 3879-3885.
[46] J.E. Jenkins, G.P. Holland, J.L. Yarger, High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders, Soft Matter, 8 (2012) 1947-1954.
[47] T. Lefevre, J. Leclerc, J.F. Rioux-Dube, T. Buffeteau, M.C. Paquin, M.E. Rousseau, I. Cloutier, M. Auger, S.M. Gagne, S. Boudreault, C. Cloutier, M. Pezolet, Conformation of spider silk proteins in situ in the intact major ampullate gland and in solution, Biomacromolecules, 8 (2007) 2342-2344.
[48] T. Lefevre, S. Boudreault, C. Cloutier, M. Pezolet, Conformational and orientational transformation of silk proteins in the major ampullate gland of Nephila clavipes spiders, Biomacromolecules, 9 (2008) 2399-2407.

[49] J. Dionne, T. Lefevre, M. Auger, Major Ampullate Spider Silk with Indistinguishable Spidroin Dope Conformations Leads to Different Fiber Molecular Structures, International Journal of Molecular Sciences, 17 (2016) 1353.
[50] H. Kamberaj, A. van der Vaart, Multiple scaling replica exchange for the conformational sampling of biomolecules in explicit water, Journal of Chemical Physics, 127 (2007) 234102.
[51] J. Spiriti, H. Kamberaj, A. Van Der Vaart, Development and Application of Enhanced Sampling Techniques to Simulate the Long-Time Scale Dynamics of Biomolecular Systems, International Journal of Quantum Chemistry, 112 (2012) 33-43.
[52] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chemical Physics Letters, 314 (1999) 141-151.
[53] P. Liu, B. Kim, R.A. Friesner, B.J. Berne, Replica exchange with solute tempering: A method for sampling biological systems in explicit water, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005) 13749-13754.
[54] I. Andricioaei, J.E. Straub, On Monte Carlo and molecular dynamics methods inspired by Tsallis statistics: Methodology, optimization, and application to atomic clusters, Journal of Chemical Physics, 107 (1997) 9117-9124.
[55] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, Journal of Chemical Physics, 79 (1983) 926-935.
[56] B.R. Brooks, C.L. Brooks, III, A.D. Mackerell, Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: The Biomolecular Simulation Program, Journal of Computational Chemistry, 30 (2009) 1545-1614.
[57] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. MacKerell, Jr., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields, Journal of Computational Chemistry, 31 (2010) 671-690.
[58] J. Huang, A.D. MacKerell, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data, Journal of Computational Chemistry, 34 (2013) 2135-2145.
[59] P. Eastman, M.S. Friedrichs, J.D. Chodera, R.J. Radmer, C.M. Bruns, J.P. Ku, K.A. Beauchamp, T.J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M.R. Shirts, V.S. Pande, OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation, Journal of Chemical Theory and Computation, 9 (2013) 461-469.

[60] R.M. Levy, M. Karplus, J.A. McCammon, Diffusive Langevin dynamics of model alkanes, Chemical Physics Letters, 65 (1979) 4-11.
[61] K.H. Chow, D.M. Ferguson, Isothermal Isobaric Molecular-Dynamics Simulations with Monte-Carlo Volume Sampling, Comput. Phys. Commun., 91 (1995) 283-289.
[62] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, Journal of Chemical Physics, 103 (1995) 8577-8593.
[63] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical-integration of cartesion equations of motion of a system with constraints - molecular-dynamics of n-alkanes, Journal of Computational Physics, 23 (1977) 327-341.
[64] D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment, Proteins, 23 (1995) 566-579.
[65] G.A. Carpenter, S. Grossberg, Neural Networks - Introduction to the 1 December 1987 Issue of Applied Optics, Appl. Optics, 26 (1987) 4909-4909.

Figures

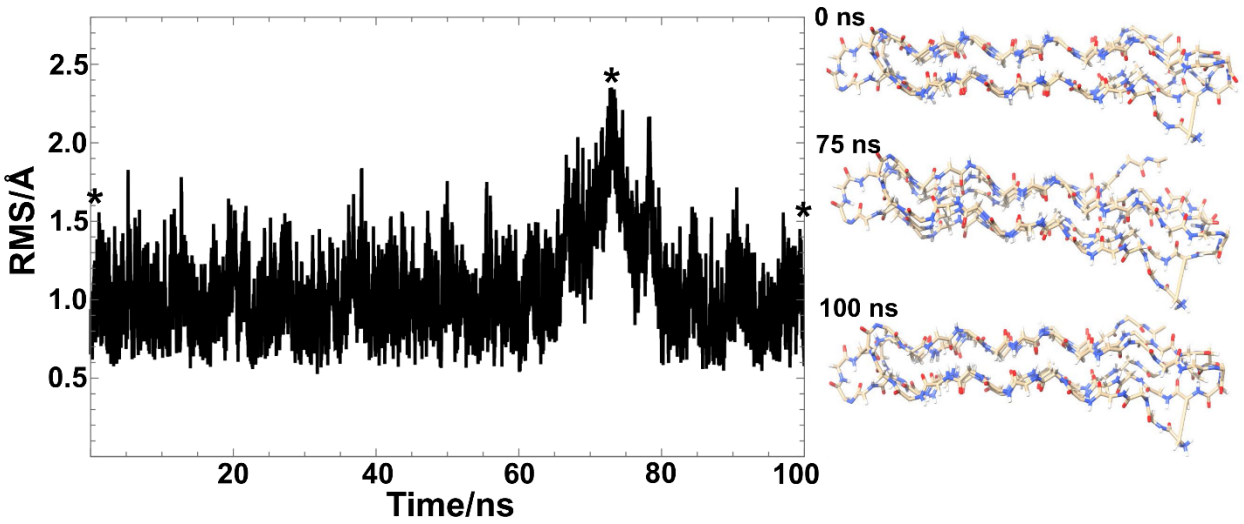

Figure 1. RMSD for the snow-flea antifreeze protein relative to the crystal structure. Snapshots at 0, 75 and 100 ns are shown on the right and indicated with stars on the RMSD graph.

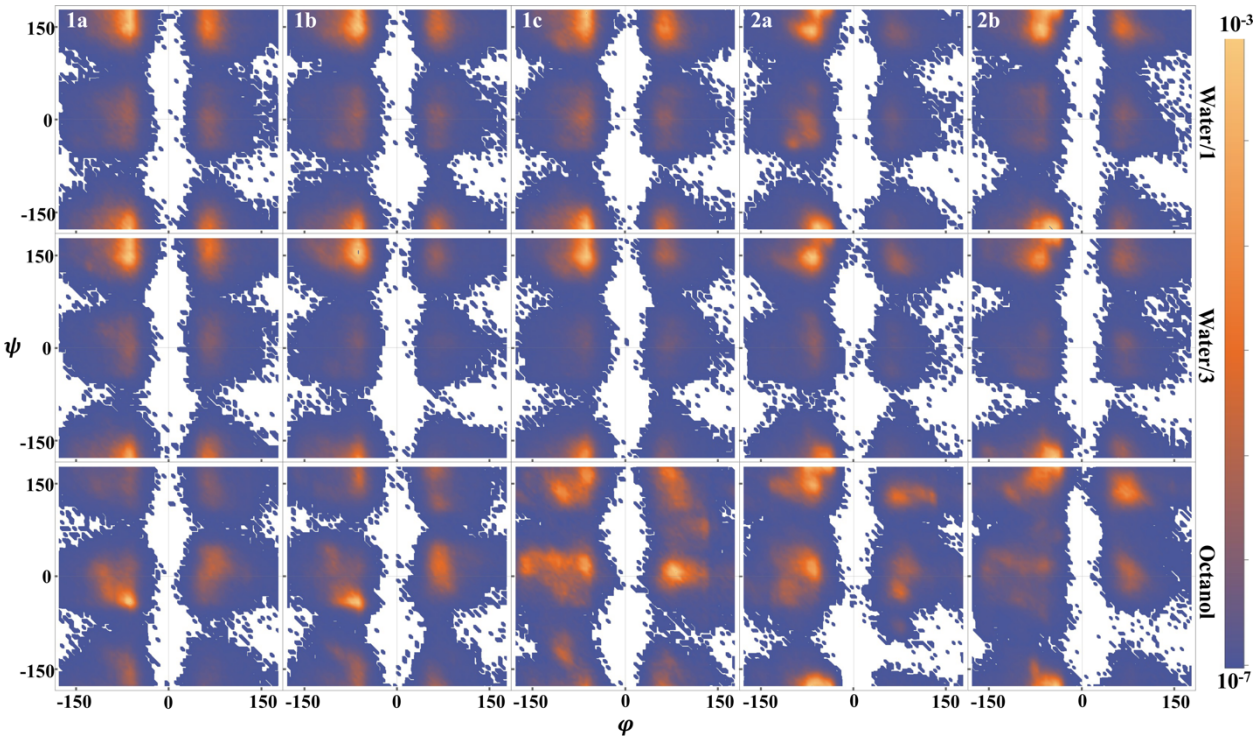

Figure 2. Ramachandran plot for Gly-rich peptides. Colors indicate the probabilities; white regions are unsampled and sterically forbidden.

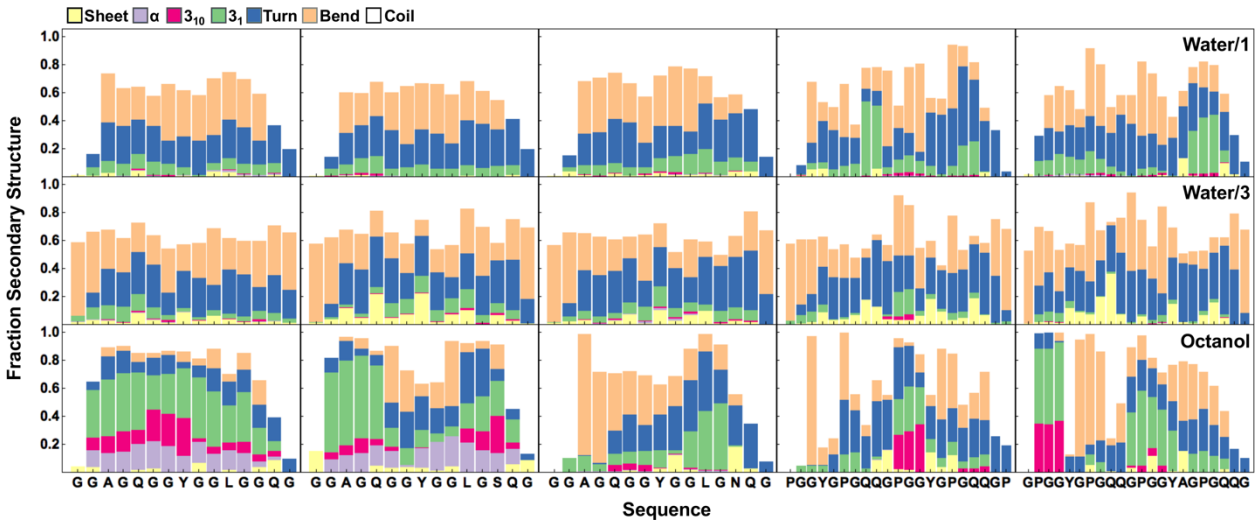

Figure 3. Secondary structure of Gly-rich peptides in water and octanol.

Tables

Table 1. Studied peptide sequences.

Name	Source	Sequence
1a	MaSp1	GGAGQGGYGGLGSQG
1b	MaSp1	GGAGQGGYGGLGSGG
1c	MaSp1	GGAGQGGYGGLGNGG
2a	MaSp2	PGGYGPGQQGPGGYGPGQQGP
2b	MaSp2	GPGGYGPGQQGPGGYAPGQQGP