

# Molecular dynamics simulations of the minor ampullate spidroin modular amino acid sequence from *Parawixia bistriata*: insights into silk tertiary structure and fibre formation

André M. Murad · Elíbio L. Rech

Received: 31 March 2010 / Accepted: 26 July 2010 / Published online: 11 August 2010  
© Springer-Verlag 2010

**Abstract** Spider fibres are primarily composed of proteins that are secreted by specialised glands found in different groups of arthropods. Because of their unique mechanical characteristics, it is of great interest to understand how the influence of repetitive modules within the fibres affects the final protein structure. Because each fibre is composed of a diverse set of repeated modular sequences, the differences between fibres allow for their structural comparison and, thereby, their functional comparison. Herein, we present molecular dynamics simulations of partial sequences from minor ampullate Spidroin (MiSp) silk of the Brazilian species *Parawixia bistriata*. Our data show that the formation of  $\beta$ -sheet structures is directly related to the N-termini alignment of the modules. The N-terminal alignment gives rise to a high number of hydrogen bonds whose formation is driven by repeated alanine (Ala) sequences, which, in turn, lead to increased fibre strength. This increased fibre strength contributes significantly to the final tertiary structure of the silk.

**Keywords** Biopolymer · Minor ampullate gland · *Parawixia bistriata* · Spider silk

## Introduction

In recent years, much research has been focused on the study of spider silk fibres [1]. These fibres have a unique

combination of toughness and elasticity compared to man-made aramid chemical polymers, such as Kevlar and Nomex [2]. Spiders have the capacity to produce silk and to utilise different silks for specialised purposes including prey capture, sustenance and reproduction [3]. Spiders have up to seven different glands that produce unique silks. However, some species can combine each gland to create different silk fibres to carry out their task of making traps and cocoons [4]. The fibre amino acid sequences are mainly composed of the repeated motif GPGXX (X is most often glutamine), Ala-rich sequences (An or (GA)n) and the GGX motif in which X is Ala, tyrosine (Tyr), leucine (Leu) or glutamine (Gln) [5–7]. The most well-studied spider silk is the dragline of the Major ampullate Spidroins (MaSp1 and MaSp2) from *Nephila clavipes* [8]. MaSp1 and MaSp2 are found in a highly concentrated state in these spiders' glands (approximately 30–50 wt %) [9, 10] and are extremely sensitive to shear stress, drying and hydration [9]. The *Parawixia bistriata* species is typically found in the Brazilian Cerrado region. Unlike other spiders, *P. bistriata* live in groups without cannibalism or territorialism surrounded by huge, strong communal webs in which each spider produces its own orbweb. However, there is limited information on the tertiary structure of the silk produced by this species (Paulo C. Motta, Zoology Department, University of Brasilia, personal communication). Many experimental studies have recently been undertaken to explain the mechanical properties of spider fibres based on their macromolecular and supramolecular structural organisation. Recent nuclear magnetic resonance (NMR) studies of proteins directly extracted from spiders' glandular sacs were compared with extruded fibres. It was found that inside the glands, proteins adopt several secondary structures [8]. Solid state NMR data of lyophilised gland dope showed no evidence of  $\beta$ -sheet

A. M. Murad · E. L. Rech (✉)  
Embrapa Recursos Genéticos e Biotecnologia,  
Parque Estação Biológica - PqEB,  
Av. W5 Norte (final), Caixa Postal: 02372, CEP: 70770-900  
Brasília, Brazil  
e-mail: rech@cenargen.embrapa.br

structures [11]. Solution state NMR, infrared Fourier transform and circular dichroism (CD) data suggested that the spidroins in the gland are in a highly metastable state of dynamically averaged helical conformations [9]. A NMR spectroscopic examination of native silk dope suggested a dynamically disordered structure, although other secondary structures with similar chemical shifts could not be excluded as possible alternative structures [12]. CD analysis of proteins freshly extracted from the tail, proximal and middle section of a silk gland indicated the presence of a structure that was poor in  $\alpha$ -helices and  $\beta$ -sheets, whereas proteins extracted from the distal part of the gland contained an abundance of  $\beta$ -sheet structures [13, 14]. These results indicate that the silk structure is polymorphic and flexible. Based on data from the *P. bistriata* functional spider silk genome, we conducted molecular dynamics simulations on the dragline MiSp amino acid sequence in water and vacuum environments.

## Material and methods

### Amino acid sequence of *Parawixia bistriata* MiSp

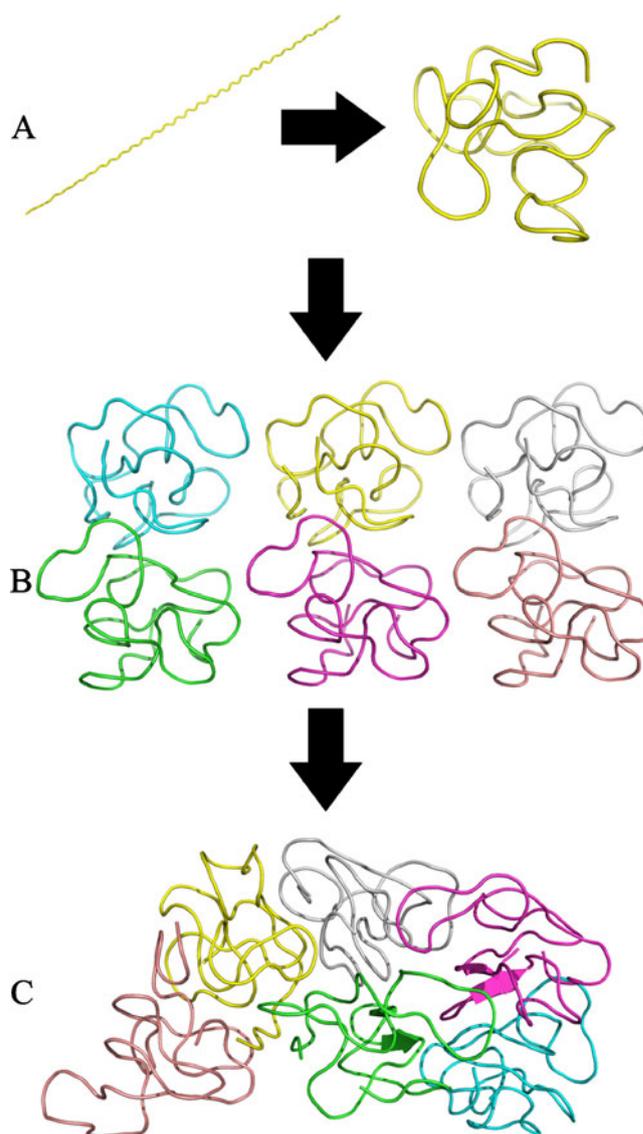
The amino acid sequence of MiSp was retrieved from GenBank accession number GQ275358. Only the partial translated gene sequence was used, which was given by GAGAGGGYGGGYGAGGGAGAGAGAGAGAGAGA GAGRGGAGGYGAGAGAGAGAAAAA.

### Molecular model building

For molecular dynamics simulations, a single primary sequence of MiSp was modelled onto a  $\beta$ -strand structure using the *DeepView* program v4.0 [15] by adjusting the phi and psi angles (Fig. 1a). This type of secondary structure was chosen based on results previously described by Hayashi et al. [16] in which MiSp from *Nephila clavipes* adopted a crystalline  $\beta$ -sheet conformation. Additionally, sequence modifications were made to increase the amount of Ala in the minor sequence (Figs. 5 and 6). The positions and numbers of sequences simulated are shown in Figs. 1, 5 and 6.

### Molecular dynamics (MD) simulations

The molecular dynamics simulations were performed using the GROMACS v3.3 package software [17] with a Solaris v10.0 operating system run on two Sun Fire X2200 M2s with two Opteron 2.6-GHz dual core processors (Sun Microsystems, CA, USA), which were connected to a 1-GB switch (a total of four processors with two cores each). All water box simulations were performed using the



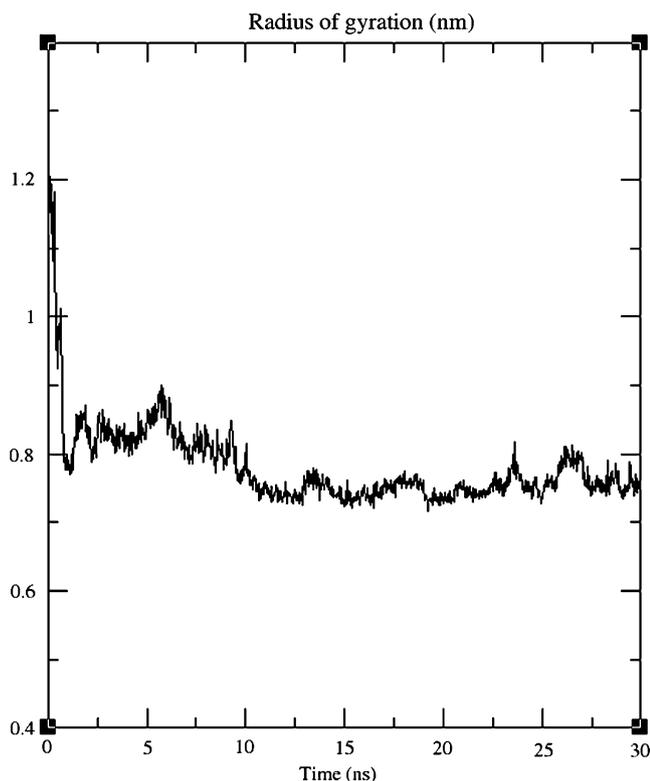
**Fig. 1** Initial and final structures of the MiSp module sequences. **a** Initial conformation followed by simulation, **b** six MiSp modules from previous simulations and **c** final structure after 30 ns. Arrows indicate the simulation step. Each colour corresponds to chains with identical sequences

following procedures. The minor modelled structure was first placed at a distance of 1 nm from the cubic box sides, and the protein was solvated with simple point charge (SPC) water molecules. Classical Ewald electrostatics was used with a cut-off distance for the short-range neighbour list of 1.0 nm and a Coulomb cut-off of 1.0 nm. No counter ions were included. Energy minimisation was performed using the steepest descent algorithm until the potential energy after minimisation was lower than  $-1.0 \times 10^5$  kJ.mol<sup>-1</sup>. Then, a short MD run with position restraints on the protein was conducted until the potential energy converged and stabilised. Berendsen temperature coupling was turned on with the temperature set to 310 K. No pressure coupling was

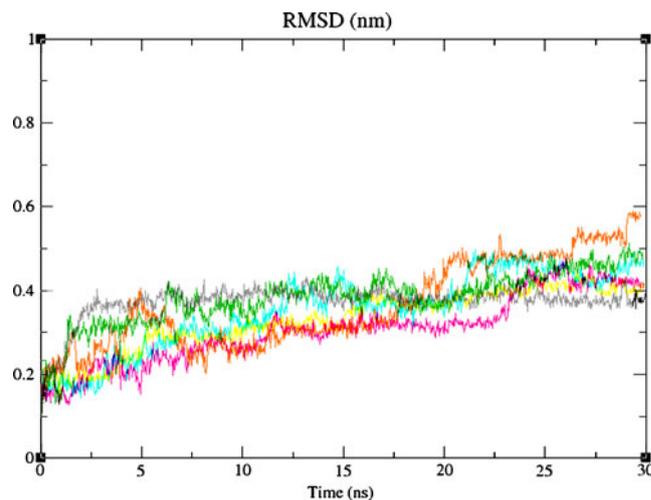
used, and MD was performed for approximately 100 ps. Finally, a full MD simulation without restraints was performed for 30 ns. Berendsen temperature and isotropic pressure coupling were used with a temperature of 310 K and a pressure of 1.0 atm. Periodic bound conditions were set to the XYZ coordinates, and Langevin dynamics were employed. Vacuum simulations were executed for 10 ns using Berendsen temperature and isotropic pressure coupling at 310 K and 1.0 atm. Again, periodic bound conditions were set to the XYZ coordinates to ensure that the atoms stayed inside the simulation box. Langevin dynamics were again employed.

### Trajectory analysis

The simulations were analysed using tools from the GROMACS package and were viewed and edited using Pymol software [18]. All trajectories were re-sampled, skipping 50 frames. The root mean square deviations (RMSDs) were calculated using `g_rms`. Cluster analysis was performed with `g_cluster` by setting the parameter `rmsmin` to 0.5 and `rmsd cutoff` to 0.5. Hydrogen bonds were calculated using `g_hbond`. The radius of gyration (Rg) of the protein was calculated using `g_gyrate`. All graphics were generated using `xmgrace` plotting software (<http://plasma-gate.weizmann.ac.il/Grace/>).



**Fig. 2** Radius of gyration (Rg). The black line indicates the Rg value for the MiSp module during the simulation

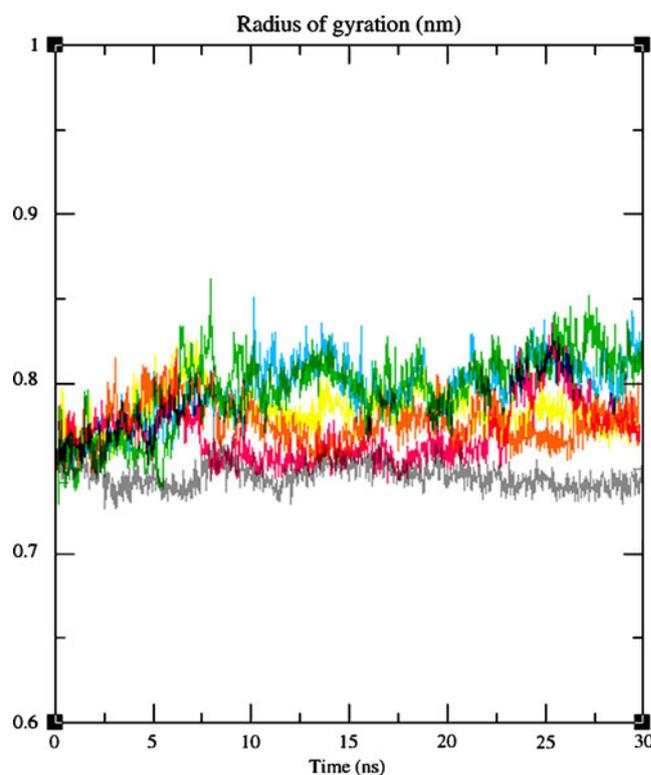


**Fig. 3** Root mean square deviation (RMSD) of the simulations. Each coloured line indicates a RMSD value for one chain during the simulation

## Results

### MiSp modular sequence under water conditions

Simulations of the structural behaviour of the MiSp module under a water solvation environment were conducted. Results of these simulations in which a single pre-modelled MiSp



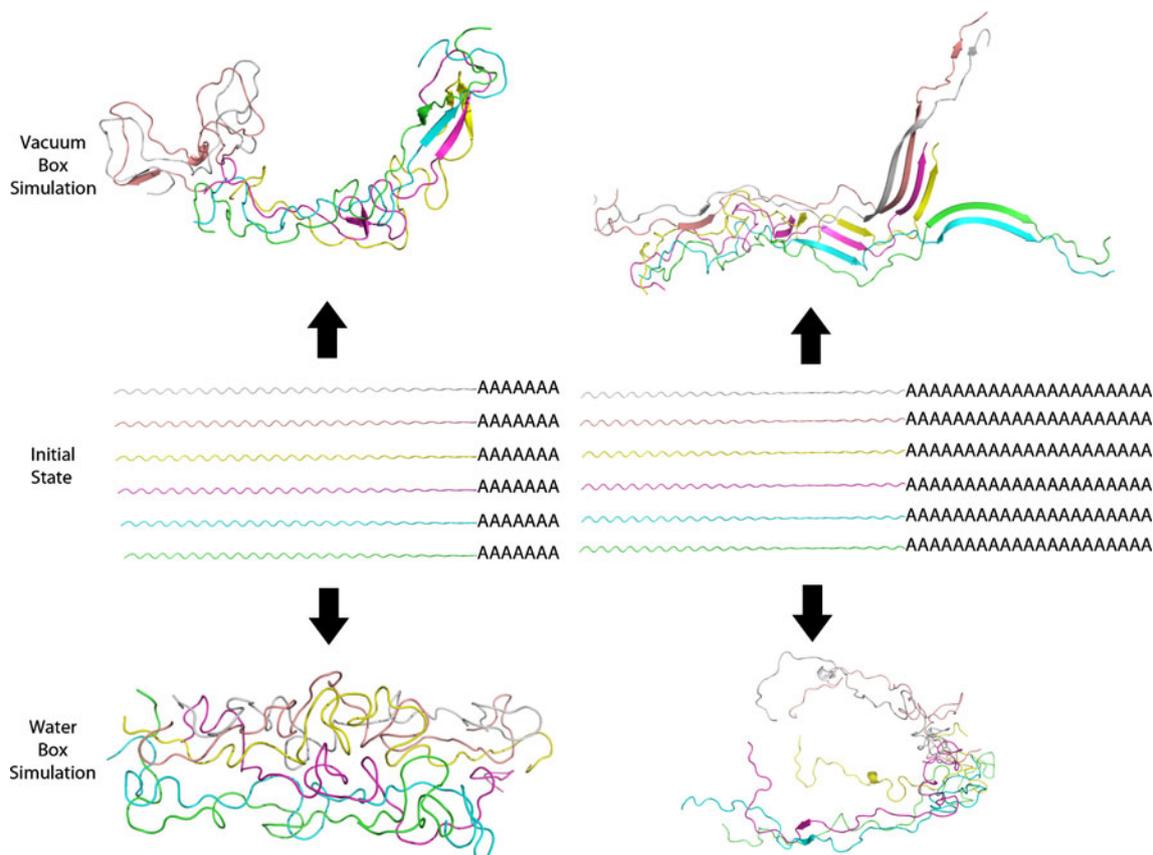
**Fig. 4** Radius of gyration (Rg). Each coloured line indicates a Rg value for one chain during the simulation

sequence was initially set to a  $\beta$ -strand secondary structure are shown in Fig. 1. After 30 ns of simulation in a cubic water box, the final structure of the MiSp module is shown in Fig. 1a. Upon initial visual analysis, this structure was composed entirely of coiled component and had a highly hydrophobic core. Changes in the structure over time are reported in the RMSD graph (data not shown) and show MiSp module changes that rapidly appear at the beginning of the simulations and stabilise around 0.15 nm over the 30 ns of simulation. The number of hydrogen bonds between the module and the solvent was also calculated (data not shown). The radius of gyration indicates that the minor module rapidly compacts its structure and remains compact with a value of 0.78 nm and low oscillation during the simulation (Fig. 2). These results indicate that the minor module has the same hydrophobic characteristics as the MiSp dragline from other spiders [5]. Specifically, the protein has an increased number of hydrogen bonds and a reduced number of contacts with the water molecules. Additionally, the structural behaviours of six MiSp module structures were simulated (Fig. 1b). The mean final calculated structure is represented in Fig. 1c, and the corresponding RMSD values for each chain are shown in Fig. 3. These results demonstrate that each module has flexibility and can refold in different

three-dimensional structures, adapting its shape to accommodate other MiSp modules. Thus, each module is able to form a highly hydrophobic quaternary structure as seen from the graph of the radius of gyration (Fig. 4). The number of hydrogen bonds on all chains remained constant (data not shown), and the structure remained coiled. These conformations are not fibre-like structures and are, thus, unique structures that are adopted in the interior of the gland sac [8].

#### Six MiSp modular sequences under water and vacuum conditions

Given that the simulation of a single module did not shed light on fibre formation, we simulated six modules' N-termini aligned under water and under vacuum environments with and without modifications to the number of Ala residues in their sequences, as shown in Fig. 5. After simulation, we found that the presence of more Ala residues assisted fibre formation only in the vacuum environment. The water box simulation demonstrated that modules agglomerate in the Ala-rich region and, thus, no crystalline  $\beta$ -sheet is able to be seen (Fig. 5). However, in the vacuum simulation, the formed structure was similar to the structure of spider silk (Fig. 5), indicating that dehydration is



**Fig. 5** Initial states and final structures for the six MiSp module sequences simulated in water and in the vacuum box. Arrows indicate the simulation step. Each colour corresponds to chains with identical sequences

necessary for fibre formation because the water-mediated hydrogen bonds are more accessible to modules with extended structures. This phenomenon occurs in the spider spinning duct, which effectively removes water from the fibre through fine channels and, thus, assists in the fibre’s formation [19]. The formation of  $\beta$ -sheets in this case occurs in a (GA) $x$  region rather than at the poly-Ala residues.

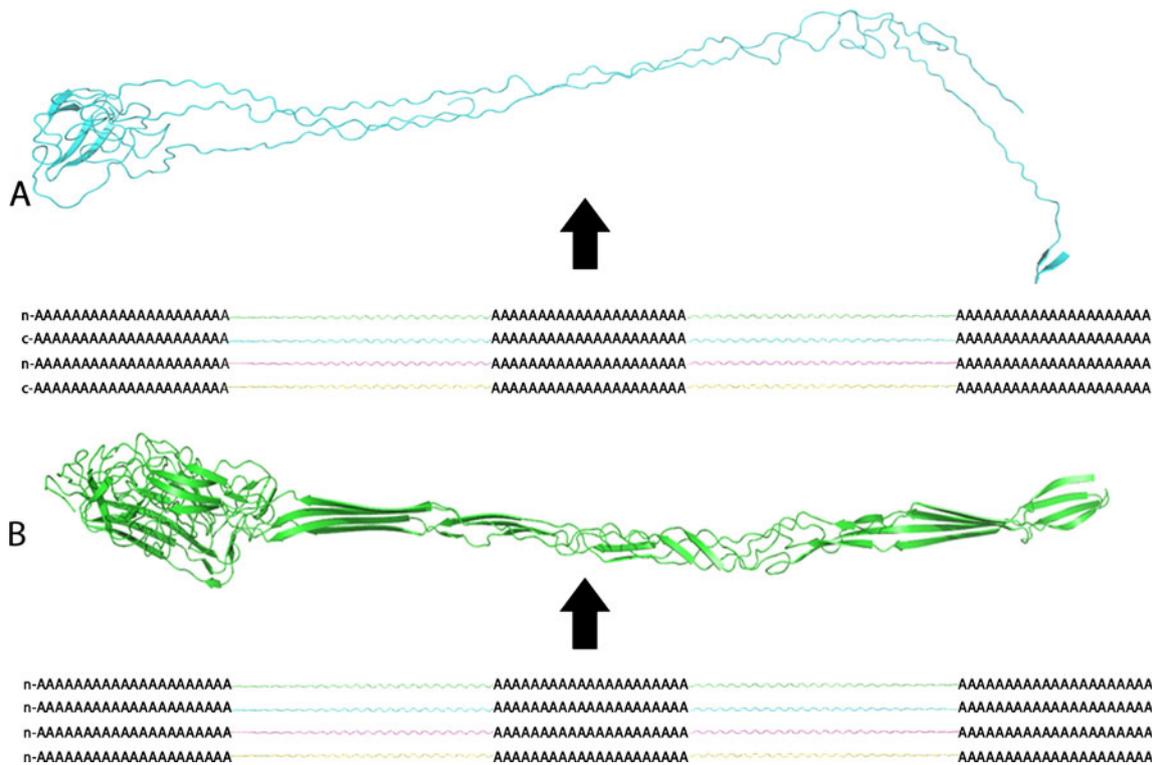
Four MiSp modular sequence modules with initial structural modifications under vacuum conditions

The vacuum environment simulation demonstrated that fibre formation occurred more readily in vacuum. We wished to verify whether or not the presence of Ala residues was the only factor that contributed to the formation of the  $\beta$ -sheets. Therefore, two conditions were tested: first, four sequences with two modules containing 21 Ala residues inserted at three positions in the N-terminus was tested (Fig. 6b), and second, the same structure as described before but with intercalating N-terminal sequences within the C-terminal sequences was explored (Fig. 6a). These simulations indicated that not only were the Ala residues important for module alignment and fibre formation, but also the orientation of the sequences was essential (Fig. 6). A highly condensed  $\beta$ -sheet was observed only when N-terminal sequences were simulated (Fig. 6b). In

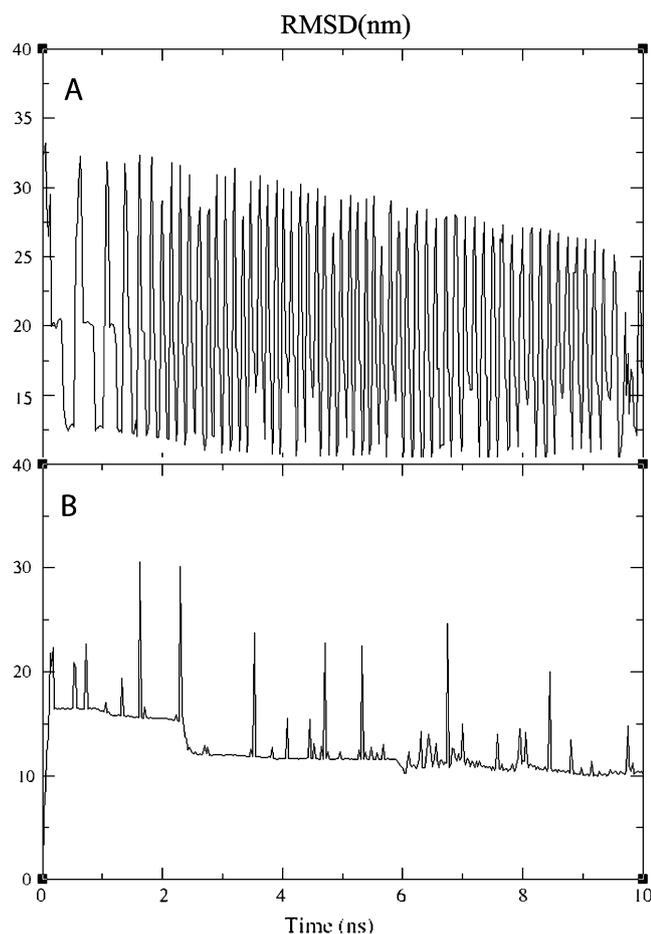
these regions, we found that the (GA) $n$  repeated motif aligned and resolved the secondary structure mainly because of the presence of the poly-Ala sequences. The RMSD values for each simulation were also calculated and show that the inserted N-terminal sequences tended to stabilise the fibre (Fig. 7b), unlike the intercalated sequences, which adopted no significant tertiary structure (Fig. 7a). In this case, the molecules possible reorganise to achieve a more energetic favourable structure may explain the RMSD oscillation (Fig. 7a). Another notable quality of our simulated structures is that the number of hydrogen bonds found in the inserted N-terminal sequences numbered 250 compared with the 60 hydrogen bonds in the intercalated sequences.

Discussion

In recent years, various efforts have been made to elucidate the molecular mechanisms behind the unique properties spider silk. Vehoff et al. [20] carried out mechanical measurements and clarified the underlying causes for some spider dragline silk properties from *Nephila clavipes* and *Nephila senegalensis*. For example, they explored the effects of humidity, hysteresis and relaxation in spider silks and showed that water can easily permeate the fibres and be incorporated into the amorphous matrix. Vehoff et al.



**Fig. 6** Initial states and final structures of the four MiSp module sequences simulated in the vacuum box. **a** Initial and final structures of the intercalated aligned modules and **b** initial and final structures of the N-terminal inserted modules. Arrows indicate the simulation step



**Fig. 7** RMSDs of the simulation. **a** RMSD values for the intercalated aligned modules and **b** RMSD values for the N-terminal inserted modules

assumed that the hydration interfered with the hydrogen bonding between the amorphous chains, leading to a loss of rigidity. Based upon our simulations, water does indeed permeate the MiSp modules and causes separation of the chains. In contrast, in vacuum simulations, the chains remain bonded. Holland et al. [11] investigated major and MiSp molecule properties in native and hydrated states of *N. clavipes* by solid state NMR. Their results indicated that water plasticises MaSP and MiSP silks, thus increasing the chain dynamics observed in regions containing Gly, Glu, Ser, Tyr and Leu when the silk was hydrated. They also observed that in poly(Ala) and poly(Gly Ala) motifs, the silks have predominately rigid  $\beta$ -sheet structures, indicating that water did not penetrate these domains. Our molecular dynamics experiments showed a high movement in Gly regions. The modules demonstrated high environmental adaptation capabilities both in water and in vacuum. However, when tandem repeat sequences were utilised, fibres were observed only when sequences were N-terminally inserted in the vacuum environment. The insertion of Ala residues, in both cases, led to the formation

of more hydrogen bonds in the poly(Gly, Ala) region, raising the number of  $\beta$ -sheets in the three-dimensional structure and, therefore, increasing the fibres' strength. These results were in agreement with most biophysical data in the literature [8, 9, 11]. Furthermore, Hronska et al. [12] stated that dynamic disordered structures can be found in silk dope. This statement was confirmed in our water experiments. Most of the previously recorded data were obtained from major ampullate silk, due to its unique combination of elasticity and high tensile strength [1, 9]. X-ray diffraction, circular dichroism and NMR data revealed that  $\beta$ -sheets are the main secondary structural conformation present in spider silk, and that they are oriented approximately in parallel with the fibre axis [21–25]. Furthermore, some molecular dynamics simulations have been performed with peptides similar to those found in spider silks. Ma and Nussinov [26] modelled a Syrian hamster prion peptide, an amyloidogenic precursor of certain diseases including Alzheimer's, and simulated its structure and stability in several situations. The simulated peptides had the sequences AGAAAAGA and AAAAAAAA ( $A_8$ ) and corresponded to the tails of our models. Our results indicate that these peptides form a highly hydrophobic  $\beta$ -sheet structure when they are inserted into the N-termini of several MiSp ampullates. This phenomenon has also been described by Kenney et al. [13]. When  $\beta$ 2-microglobulin peptides were simulated, we observed that they also had chaotic structures during the simulation. However, when Kenney et al. simulated more peptides, the structures converged to form highly organized  $\beta$ -sheet structures with increased numbers of hydrogen bonds in a similar fashion to the structures found in our study [27]. Becker et al. [28] simulated the flagelliform protein sequences from *Araneos gemmoides* and *Nephila clavipes* with 18 amino acid residues using steered molecular dynamics. The *A. gemmoides* sequence was shown to be less regular than the *N. clavipes* sequence, but both sequences had more flexibility than the dragline. The same properties occurred in our simulations of sequences with seven Ala residues, which were responsible for the formation of highly tensile fibres, as previously reported by Hayashi et al. [16] and Lewis [1]. Another interesting hypothesis regarding fibre formation suggests that the strength of the  $\beta$ -sheet fibre structures relies not only on the hydrophobic forces driven by the repeated Ala sequences but also on the arrangement of the hydrogen bond clusters [29]. We were able to visualise the increase in hydrogen bond number between modules and in the N-termini molecules with inserted sequences. No significant differences were found in either simulation. However, there exists the possibility that other structural elements may be responsible for the silk strength, as proposed by Vehoff et al. [20]. Another possibility is that a mixture of monomers and aligned sequences may influence the final fibre strength and flexibility. The data

reported here may contribute to the understanding of the molecular engineering involved in spider silk modules and may be useful for the *in vitro* production of synthetic fibres.

## Summary

We modelled and simulated several MiSp sequence modules (GenBank accession number GQ275358) in water and vacuum box environments. Using a constant temperature and a constant pressure of 310 K and 1.0 atm, respectively, our results indicated that no fibres formed in the water conditions. MiSp modules simulated under vacuum conditions formed highly organized  $\beta$ -sheet secondary structures, which were modified to be similar to spider silk structures in two ways. First, one block of 21 Ala residues at the C-terminus of the modules was added and simulated. Second, a modified structure containing three blocks of 21 Ala residues that were inserted between two modules and aligned with the N-terminus was examined. The formation of the  $\beta$ -sheet secondary structures of these modified structures indicates that Ala residues are not the main factor in fibre formation, but that the alignment of the modules and the absence of water molecules are also important factors.

**Acknowledgments** This work was supported by Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo a Pesquisa do Distrito Federal (FAP-DF).

## References

- Lewis RV (2006) Spider silk: ancient ideas for new biomaterials. *Chem Rev* 106:3762–3774
- Ko FK, Kawabata S, Inoue M, Niwa M, Fossey S, Song JW (2002) Engineering properties of spider silk. Proceedings, MRS Annual Meeting, Drexel University
- Gosline JM, DeMont ME, Denny MW (1986) The structure and properties of spider silk. *Endavour* 10:37–43
- Scheibel T (2004) Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb Cell Fact* 3:14
- Xu M, Lewis RV (1990) Structure of a protein superfiber: spider dragline silk. *Proc Natl Acad Sci USA* 87:7120–7124
- Hinman MB, Lewis RV (1992) Isolation of a clone encoding a second dragline silk fibroin, *nephila clavipes* dragline silk is a two protein fiber. *J Biol Chem* 267:19320–19324
- Oroudjev E, Soares J, Arcidiacono S, Thompson JB, Fossey AS, Hansma HG (2002) Segment nanofibers of spider dragline silk: atomic force microscopy and single-molecule force spectroscopy. *Proc Natl Acad Sci USA* 99:6460–6465
- Lefevre T, Boudreault S, Cloutier C, Pezolet M (2008) Conformational and orientational transformation of silk proteins in the major ampullate gland of *nephila clavipes* spiders. *Biomacromolecules* 9:2399–2407
- Hijirida DH, Do KG, Michal C, Wong S, Zax D, Jelinski LW (1996) NMR of *nephila clavipes* major ampullate silk gland. *Biophys J* 71:3442–3447
- Chen X, Knight DP, Vollrath F (2002) Rheological characterization of *nephila* spidroin solution. *Biomacromolecules* 3:644–648
- Holland GP, Jenkins JE, Creager MS, Lewis RV, Yarger JL (2008) Solid-state nmr investigation of major and minor ampullate spider silk in the native and hydrated states. *Biomacromolecules* 9:651–657
- Hronska M, van Beek JD, Williamson PT, Vollrath F, Meier BH (2004) NMR characterization of native liquid spider dragline silk from *nephila edulis*. *Biomacromolecules* 5:834–839
- Kenney JM, Knight D, Wise MJ, Vollrath F (2002) Amyloidogenic nature of spider silk. *Eur J Biochem* 269:4159–4163
- Hedhammar M, Rising A, Grip S, Martine AS, Nordling K, Casals C, Stark M, Johansson J (2008) Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin I from *euprosthenops australis*: implications for fibre formation. *Biochemistry* 47:3407–3417
- Guex N, Peitsch MC (1997) Swiss-model and the swiss-pdbviewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723
- Hayashi CY, Shipley NH, Lewis RV (1999) Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int J Biol Macromol* 24:271–275
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) Gromacs: fast, flexible, and free. *J Comput Chem* 26:1701–1718
- DeLano WL (2002) The pymol molecular graphics system 0.9 edit. DeLano Scientific, San Carlos
- Vollrath F, Knight D (2001) Liquid crystalline spinning of spider silk. *Nature* 410:541–548
- Vehoff T, Glisovic A, Schollmeyer H, Zippelius A, Salditt T (2007) Mechanical properties of spider dragline silk: humidity, hysteresis, and relaxation. *Biophys J* 93:4425–4432
- Simmons AH, Michal CA, Jelinski LW (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* 271:84–87
- Riekel C, Branden C, Craig C, Ferrero C, Heidelbach F, Miller M (1999) Aspects of x-ray diffraction on single spider fibers. *Int J Biol Macromol* 24:179–186
- van Beek JD, Kummerlen J, Vollrath F, Meier BH (1999) Supercontracted spider dragline silk: a solid-state nmr study of the local structure. *Int J Biol Macromol* 24:173–178
- Brooks AE, Stricker SM, Joshi SB, Kamerzell TJ, Middaugh CR, Lewis RV (2008) Properties of synthetic spider silk fibers based on *argiope aurantia* msp2. *Biomacromolecules* 9:1506–1510
- Dicko C, Porter D, Bond J, Kenney JM, Vollrath F (2008) Structural disorder in silk proteins reveals the emergence of elastomericity. *Biomacromolecules* 9:216–221
- Ma B, Nussinov R (2002) Molecular dynamics simulations of alanine rich beta-sheet oligomers: insight into amyloid formation. *Protein Sci* 11:2335–2350
- Liang C, Derreumaux P, Wei G (2007) Structure and aggregation mechanism of beta(2)-microglobulin (83-99) peptides studied by molecular dynamics simulations. *Biophys J* 93:3353–3362
- Becker N, Oroudjev E, Mutz S, Cleveland JP, Hansma PK, Hayashi CY, Makarov DE, Hansma HG (2003) Molecular nanosprings in spider capture-silk threads. *Nat Mater* 2:278–283
- Keten S, Buehler MJ (2008) Geometric confinement governs the rupture strength of h-bond assemblies at a critical length scale. *Nano Lett* 8:743–748