JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT°

ELSEVIER

Novel Solution Structure of Porcine β-Microseminoprotein

Iren Wang[1,2], Yuan-Chao Lou[1], Kuen-Phon Wu[1], Shih-Hsiung Wu[2,3]* Wen-Chang Chang[2] and Chinpan Chen[1]*

[1]*Institute of Biomedical Sciences, Academia Sinica Taipei 115, Taiwan, ROC*

[2]*Institute of Biochemical Sciences, National Taiwan University, Taipei 106, Taiwan ROC*

[3]*Institute of Biological Chemistry, Academia Sinica Taipei 115, Taiwan, ROC*

*Corresponding authors

A number of β-microseminoproteins (MSPs) have been identified from different species. MSPs are all non-glycosylated and disulfide bond-rich, but show a relatively low level of conservation. Although all Cys residues are conserved, our previous study showed that the disulfide bond pairings differ in porcine and ostrich MSPs. Despite the variety of biological functions that have been suggested for MSPs, their real function is still poorly understood. Furthermore, no 3D structure has been reported for any MSP, so the determination of the structure and function of MSPs is an interesting and important task. In the present study, we determined the 3D solution structure of porcine MSP on the basis of 1018 restraints. The ensemble of 20 NMR structures was well defined, with average root-mean-square deviations of $0.83(\pm 0.16)$ Å for the backbone atoms and $1.37(\pm 0.17)$ Å for heavy-atoms in residues 2–90. The 3D structure showed that porcine MSP is clearly composed of two domains, an N-terminal domain consisting of one double-stranded and one four-stranded antiparallel β-sheet, and a C-terminal domain consisting of two double-stranded antiparallel β-sheet. The orientation of the two domains was derived mainly on the basis of long-range NOEs and verified using residual dipolar coupling data. No inter-domain hydrophobic interaction or H-bonding was detected. However, a number of charged residues were found in close proximity between the domains, indicating that electrostatic interaction may be the key factor for the orientation of the two domains. This is the first report of a 3D structure for any MSP. In addition, structural comparison based on distance matrix alignment (DALI), class architecture topology and homologous superfamily (CATH) and combinatorial extension (CE) methods revealed that porcine MSP has a novel structure with a new fold providing valuable information for future structural studies on other MSPs and for understanding their biological functions.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: β-microseminoprotein; prostate secretory protein; Na^+, K^+-ATPase inhibitor; benign prostatic hyperplasia; solution structure

Introduction

β-Microseminoprotein (MSP) was originally isolated from human seminal plasma.[1] It is also known as PSP94 (prostate secretory protein of 94 amino acid residues), and has subsequently been identified in several

other species.[2–4] MSPs are all non-glycosylated and disulfide-bond rich, but surprisingly show a relatively low level of conservation. Figure 1 shows that MSPs contain numerous charged residues with pI values in the range of 5.1–8.9 and that all the Cys residues are conserved. Porcine MSP shows 33–54% sequence identity with eight other MSPs, but no sequence identity greater than 30% with other proteins in the Databank. A variety of putative biological roles have been suggested for MSPs, e.g. porcine MSP was found to be a Na^+,K^+-ATPase inhibitor.[5] Most recently, human MSP was suggested as a novel and apparently non-toxic form of therapy for hormone-independent prostate cancer[6] and has been shown to decrease the growth of rat prostate tumor cells.[7] Also, PSP61, a modified isoform of human MSP consisting of the N-terminal 61 residues, was discovered in patients with benign prostatic hyperplasia (BPH) and suggested as a specific biological marker for the diagnosis of BPH.[8] These findings demonstrate the importance of MSP in clinical diagnosis in prostate pathophysiology. However, in general, the biological functions of MSPs are still poorly understood. To gain an insight into the structure/function relationship of MSPs, a detailed structural analysis of MSPs is obligatory. No 3D structure has been reported for any MSP in the Protein Data Bank, and structural studies on MSPs are scarce. Recently, Lazure *et al.*[9] determined the disulfide bond pairings of ostrich MSP using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. In our previous study on porcine MSP, in addition to showing that the secondary structure is composed mainly of β-strands, we found that the disulfide pairings of porcine MSP were different from those in ostrich MSP.[10] In the present study, we determined the tertiary structure of porcine MSP that is composed of two domains, an N-terminal domain composed of one double-stranded and one four-

stranded antiparallel β-sheet, and a C-terminal domain composed of two double-stranded antiparallel β-sheets. A search using DALI,[11] CATH[12] and CE[13] found no significant structural similarity in the Protein Data Bank, further demonstrating the novel protein fold of porcine MSP. The determination of this novel structure may provide valuable information for subsequent MSP structural studies and help in understanding the biological functions of MSPs, especially those of human MSP in prostate pathophysiology.

Results

Conformational stability based on circular dichroism (CD) analysis of porcine MSP

The CD spectra showed that the conformation of recombinant porcine MSP was pH-independent over the range of 3–8 and nearly identical with that of the native protein. The secondary structure, determined using the CONTIN-LL, SELCON3, and CDSSTR programs,[14] was found to be composed mainly of β-sheet structures. To analyze the thermostability of porcine MSP, equilibrium CD titration experiments at 231 nm were carried out between 10 °C and 95 °C, and showed that the protein was not completely denatured even at 95 °C, so the T_m value could not be measured accurately (Figure 2). However, it is clear that porcine MSP is highly thermostable. Moreover, the conformation of porcine MSP was found to be thermally reversible, as the CD spectrum (195–260 nm) at 25 °C and that obtained at 25 °C after cooling from 95 °C were identical. The chemical stability of porcine MSP was examined by equilibrium unfolding at different concentrations of guanidine-HCl (Gdn·HCl), but, again, no suitable fitting curve for C_m measurement could be obtained
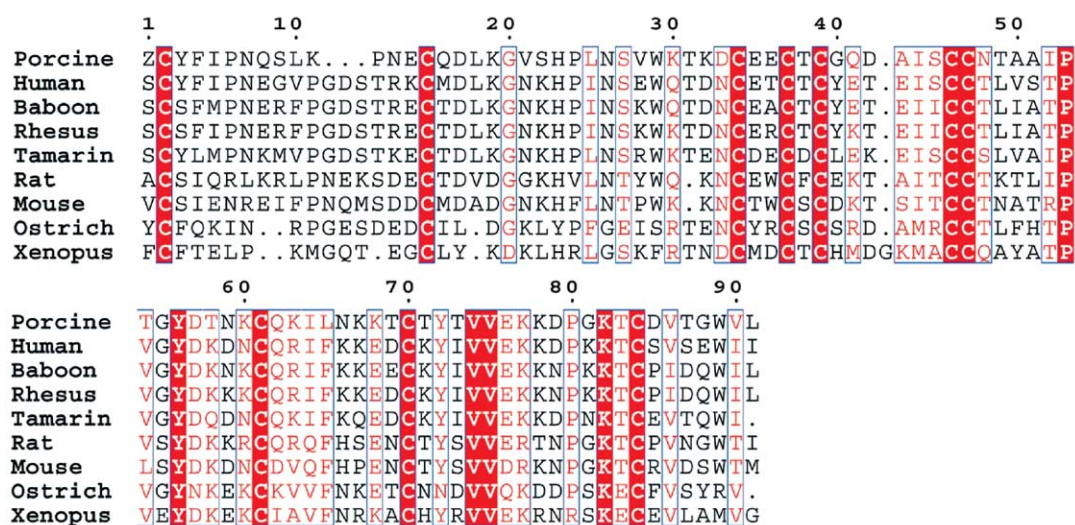

Figure 1. Sequence alignment of β-microseminoproteins from different species generated using programs CLUSTAL-W and ESPrint.
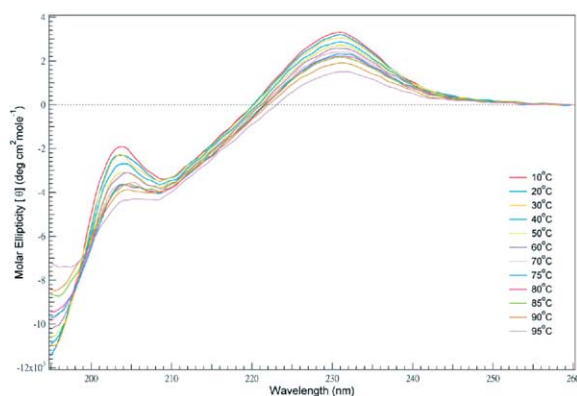
Figure 2. CD spectra of the recombinant porcine MSP at different temperatures acquired with 20 μM proteins in 20 mM phosphate buffer at pH 5.0.

due to incomplete denaturation in 8 M Gdn·HCl (data not shown), showing that porcine MSP is also chemically stable.

Resonance assignment and structure determination by NMR

Since the CD data indicated no significant structural differences in porcine MSP at different pH values, we chose to use pH 5.0 for the structural

study, as the quality of the NMR data obtained at this pH was better. All resonance assignments, including backbone and side-chain atoms, the *cis/trans* conformations of proline residues, and disulfide-bridge pairings, have been previously completely defined.[10] A consensus chemical shift index (CSI)[15] generated using the $^1H^\alpha$, $^{13}C^\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts revealed that the secondary structure was primarily composed of nine β-strands and no α-helical structure (see Figure S1 in Supplementary Material). The existence of β-sheets was also detected based on crossover β-strand nuclear Overhauser enhancements (NOEs) of $d_{\alpha\alpha}(i,j)$ in the ^{13}C NOE spectroscopy (NOESY) data. The amide proton exchange rate showed that most of the slowly exchanged amide protons were located in secondary structure regions. In total, a set of 1018 restraints that contained 891 distances, 91 dihedral angles, and 36 hydrogen bond restraints was derived and included in simulated annealing and energy minimization calculations using the XPLOR-NIH program. The 20 structures with the lowest energy were chosen to represent the ensemble of NMR structures in the final stage of refinement. These structures displayed no restraint violation greater than 0.3 Å for distance restraints and 3° for dihedral angle restraints. The stereo-view for the ensemble of 20 NMR structures with an averaged RMSD of 0.62 Å for the backbone atoms and 1.10 Å for the
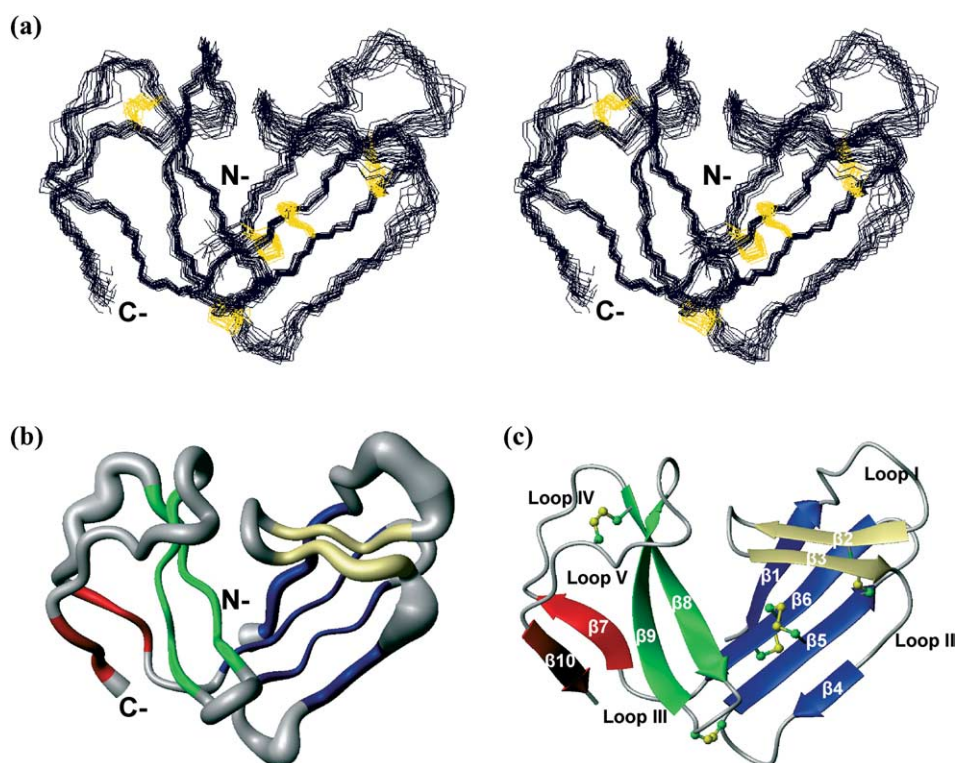

Figure 3. (a) Stereo-view of the 20 NMR structures superimposed on the secondary structure regions, with the disulfide pairings shown in yellow. (b) Sausage structure showing secondary structure elements as different colors. (c) Ribbon display of the averaged structure showing the secondary structures as different colors and disulfide bridges in yellow.

Table 1. Structural statistics on the final set of 20 simulated annealing structures of porcine β-microseminoprotein

A. *Constraints used*	
Intraresidue ($|i-j|=0$)	317
Sequential ($|i-j|=1$)	253
Medium range ($|i-j|\leqq4$)	60
Long range ($|i-j|>4$)	261
Total distance restraints	891
Hydrogen bonds	18×2
Dihedral angles	91
B. *Statistics for the final X-PLOR structures*	
Number of structures in the final set	20
X-PLOR energy (kcal mol^{-1})	
$\quad E_{NOE}$	4.07 ± 0.84
$\quad E_{cdih}$	0.20 ± 0.06
$\quad E_{bond}+E_{angle}+E_{improper}$	44.75 ± 1.25
$\quad E_{VDW}$	47.98 ± 2.16
NOE violation	
Number >0.3 Å	None
rms deviation (Å)	0.001
Deviation from idealized covalent geometry	
$\quad$ Bond angles (deg.)	0.30 ± 0.004
$\quad$ Impropers (deg.)	0.22 ± 0.004
$\quad$ Bond lengths (Å)	0.002
Mean global rms deviation (Å)	
Backbone (N, C$^\alpha$, C$'$	
Residues: 2–6, 14–17, 21–24, 28–30, 35–39, 43–48, 52–54, 61–66, 71–76, 88–90	0.62 ± 0.12
Heavy-atoms	
Residues: 2–6, 14–17, 21–24, 28–30, 35–39, 43–48, 52–54, 61–66, 71–76, 88–90	1.10 ± 0.13
Ramachandran data	
$\quad$ Residues in most favored regions (%)	67.3
$\quad$ Residues in allowed regions (%)	27.5
$\quad$ Residues in generously allowed regions (%)	4.8
$\quad$ Residues in disallowed regions (%)	0.4

heavy-atoms for residues in the secondary structure regions is shown in Figure 3(a). The sausage representation and the averaged ribbon structure are shown in Figure 3(b) and (c), respectively. A Ramanchandran plot (data not shown) based on 20 NMR structures showed that 67.3% of the residues were in the most favored regions, 27.5% in allowed regions, 4.8% in generously allowed regions, and 0.4% in the disallowed regions. Thus, the quality of NMR structure of porcine MSP was good, and the structural statistics on the final set of 20 structures are given in Table 1.

Structure description

The porcine MSP structure is clearly composed of two domains, an N-terminal domain (residues 1–50) composed of a four-stranded antiparallel β-sheet (β1 Cys[2]-Pro[6]; β4 Val[28]-Lys[30]; β5 Glu[35]-Cys[39]; and β6 Ala[43]-Asn[48]) and a double-stranded antiparallel β-sheet (β2 Glu[14]-Asp[17] and β3 Val[21]-Pro[24]) and a C-terminal domain (residues 51–91) composed of two double-stranded antiparallel β-sheets (β7 Ile[52]-Thr[54] and β10 Gly[88]-Val[90]; β8 Cys[61]-Asn[66] and β9 Thr[71]-Glu[76]). Among the four β-sheets, the double-stranded β-sheet at Glu[14]-Asp[17] and Val[21]-Pro[24] was not identified in our previous report.[10] The identification of this β-sheet is primarily due to the observations of the two characteristic NOEs of $d_{\alpha\alpha}(14, 24)$ and $d_{\alpha\alpha}(16, 22)$, which are very closed to diagonal peaks so that they were not assigned before. Moreover, the H-bond of Cys[15]·HN/His[23]·O was consistently detected in the

generated structures and the hydrogen/deuterium (H/D) exchange rate of Cys[15]·HN is not so fast, further supporting the existence of this β-sheet structure. On the basis of the φ and ψ backbone dihedral angles, we also identified that residues Asp[17]-Gly[20] form a type II' β-turn, and both Gly[40]-Ala[43] and Glu[76]-Asp[79] form type I β-turns. Four residues (Pro[53], Tyr[56], Val[74] and Val[86]) in the C-terminal domain form a hydrophobic core between the two antiparallel β-strands.

In addition to the hydrogen bonds observed in the secondary structure regions, several other H-bonds, including Glu[35]·NH/Thr[31]·O$_{\gamma1}$, Asp[79]·NH/Glu[76]·O, Lys[82]·NH/Asp[79]·O, and Cys[84]·NH/Val[74]·O, were detected in the final structures. When comparing to the exchange rate, we found that these amide protons indeed are slowly exchanged. Based on the H/D exchange rate of the amide protons, the rigidity of the β-sheets is categorized into three levels, the four-stranded antiparallel β-sheet at the N terminus being the most rigid because the amide protons in this region have very slowly exchanging rates, followed by the double-stranded antiparallel β-sheet of Cys[61]-Asn[66] and Thr[71]-Glu[76], and the two double-stranded antiparallel β-sheets of Glu[14]-Asp[17] and Val[21]-Pro[24]; Ile[52]-Thr[54] and Gly[88]-Val[90] being the least stable.

The orientation of the two domains was derived primarily on the basis of the observed long-range NOEs within two domains such as $d_{N\gamma}(19, 75)$, $d_{N\gamma}(19, 77)$, $d_N\beta(20, 75)$, $d_{N\gamma}(20, 75)$, $d_N\beta(77, 19)$ and $d_N\beta(79, 19)$, as shown in Figure 4. To gain an insight
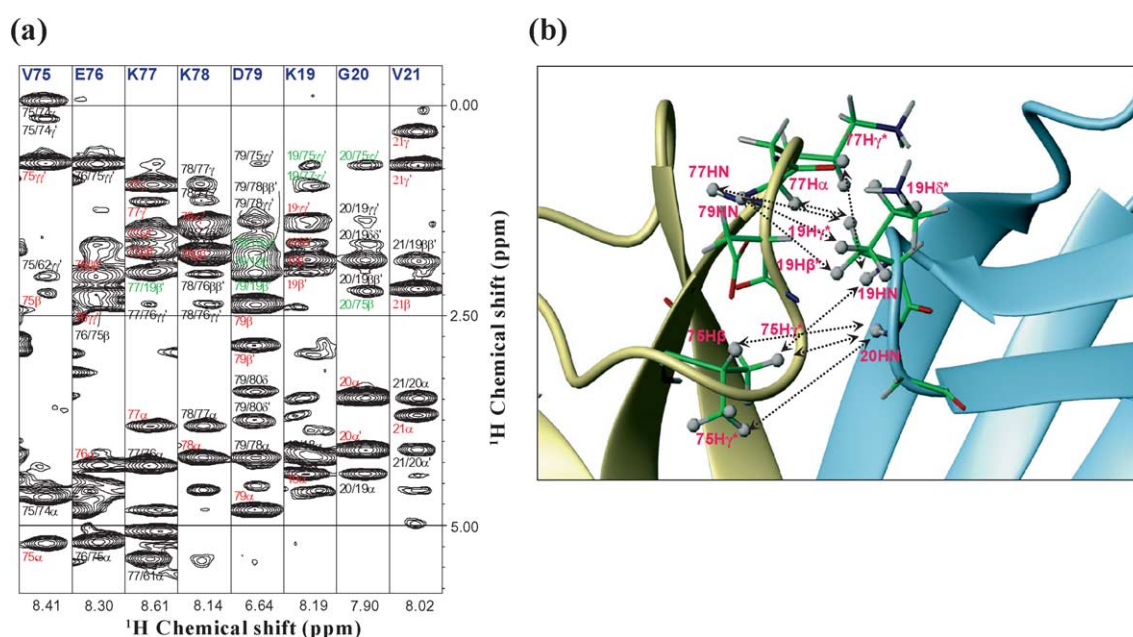
Figure 4. The critical long-range NOEs observed between the N and C-terminal domains of porcine MSP. (a) Strip plots of the ^{15}N-NOESY HSQC spectra for the residues involved in inter-domain interactions were selected; the long-range NOEs between domains are shown in green, and intra- and sequential NOEs in red and black, respectively. (b) The residues involved in inter-domain interactions are shown as ball-and-stick representations on the ribbon structure and the double arrows indicate the observed long-range NOEs between two domains.

into the interactions between the two domains, we checked for hydrophobic interaction and hydrogen bonds between the domains, but none was found. In contrast, the surface structure of porcine MSP (Figure 5) showed that all 21 charged residues were exposed to the aqueous phase. It showed that Lys[19] at the N-domain is close to Asp[79] at the C-domain and so does for Asp[17] and Lys[63] or Lys[67], indicating that the electrostatic force may play an important role for the domain interaction. To further confirm the orientation between the domains, the residual dipolar couplings (RDCs) of the N–H vector were measured and included in the structure calculation using the XPLOR-NIH program. Figure 6(a) shows

a selected region of the IPAP ^{15}N-^{1}H HSQC spectra measured in isotropic solution and in the presence of 5.3 mg/ml of Pf1. The structures generated including the RDC restraints showed an orientation very similar to those generated without the RDC restraints, further verifying the correctness of the domain orientation. As expected, the structures generated including the RDC restraints were improved, with a backbone RMSD of 0.55 Å in secondary structure regions compared to 0.59 Å for the structure without the RDC. The superimposition of 20 NMR structures calculated with or without RDC restrains (ten structures for each category) is shown in Figure 6(b).
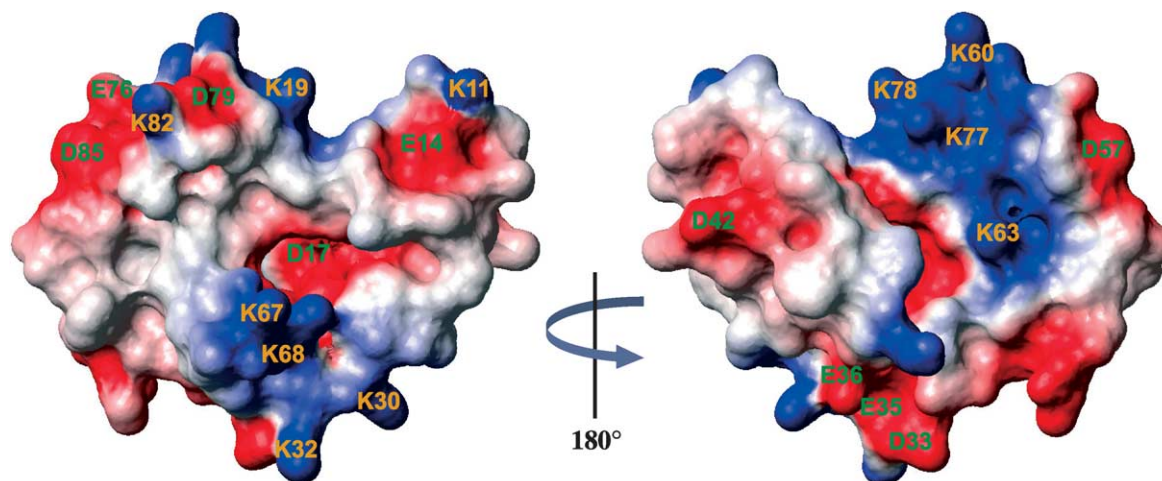

Figure 5. Surface structure of porcine MSP displayed as a 180° rotation with positive and negative charges shown in blue and red, respectively.
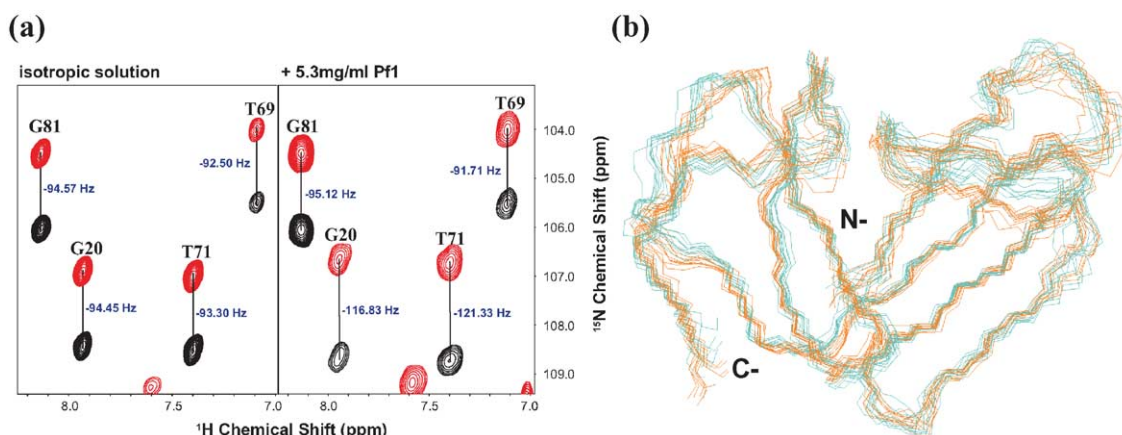
Figure 6. (a) A selected region of the IPAP ^{15}N-^{1}H HSQC spectra for porcine MSP measured at 310 K and pH 7.7 in 50 mM phosphate buffer in the absence (isotropic solution) and presence of 5.3 mg/ml of Pf1. (b) NMR structures generated with (ten structures in orange) and without (ten structures in turquoise) RDC restraints superimposed on secondary structure regions.

Determination of disulfide pairings based on structure calculation

In the previous study,[10] we used specific NOEs and structure calculation based on 90 restraints to determine the disulfide pairings of porcine MSP, and found that these disulfide pairings differed from those in ostrich MSP, for which the disulfide patterns were derived by MALDI-TOF mass analysis. In the present study, we repeated the structure calculation using a full set of restraints based on three different disulfide pairings, as previously described (P1 2-47, 15-39, 34-70, 37-46 and 61-84; P2 2-15, 34-70, 37-46, 39-47 and 61-84; P3 2-15, 34-70, 37-47, 39-46 and 61-84). The structural statistics (see Table S1 in Supplementary Material) showed that the P1 structure is the best based on the final energy and Ramanchandran plot analysis. For examples, many distance (>0.3 Å) and dihedral ($>3°$) violations were seen in P2 and P3 structures, but no such violations were found in P1. Therefore, the structural calculation further confirms the disulfide pairings for porcine MSP.

Backbone dynamics analysis

For the backbone dynamics study, relaxation data for 81 backbone nitrogen resonances, except for the five proline residues and the residues in which the cross-peaks were very weak or severely overlapped, were collected (Figure 7). The averaged values for the relaxation parameters were $1.88(\pm0.05)$ s^{-1} for R_1, $9.68(\pm0.88)$ s^{-1} for R_2, and $0.66(\pm0.05)$ for NOE. The resulting initial estimate of τ_m was $6.50(\pm0.29)$ ns. The models generated for 67 residues were selected using the FAST-Modelfree program and checked using a manual model selection strategy based on simulations of the distribution of Γi and Fi. After final optimization with the selected model for each residue, the τ_m value was $6.33(\pm0.25)$ ns. The S^2 value (order parameter) for each residue was determined

(Figure 7), showing that six residues had S^2 values lower than 0.7, 35 residues between 0.7 and 0.9, and 26 residues greater than 0.9. The distribution of S^2 values clearly showed that the secondary structure elements are rigid. The N and C-terminal residues have large S^2 values, in good agreement that these residues form β-sheet structures. Furthermore, the order parameters of the N-terminal four-stranded antiparallel β-sheets (β1, β4, β5, and β6) are larger than those in other β-sheets, indicating that this β-sheet is the most rigid. For the other three
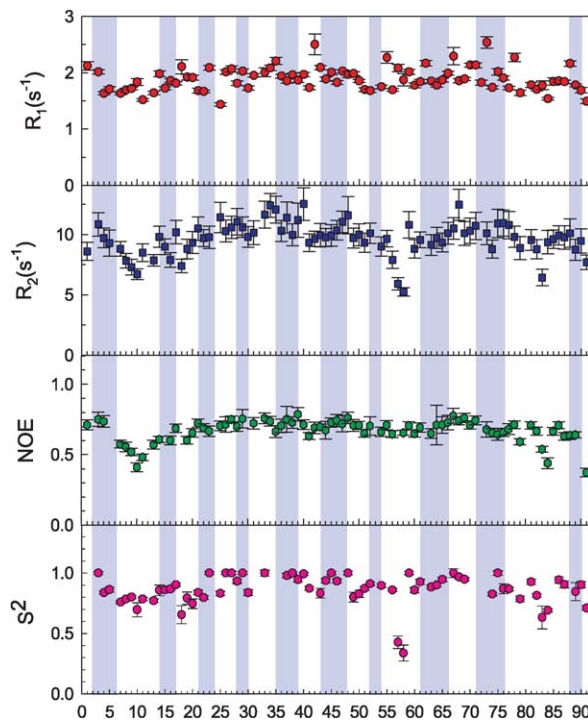

Figure 7. R_1, R_2, NOE, and S^2 (order parameter) values obtained at 310 K at pH 5.0 shown as a function of residue. Shaded ranges indicate the ten β-strands of porcine MSP.

double-stranded β-sheets, the β-sheet composed of β8 and β9 possesses higher order parameters than the other two β-sheets that are the least rigid among all of β-sheet structures. Therefore, the rigidity of β-sheets based on order parameters is in general agreement with the result based on exchange rate, as described above. As expected, the flexible parts are located mainly in the loop regions. For instances, the four residues of Asp[57], Thr[58], Thr[83], and Cys[84] that have lowest S^2 values are located in the loop IV and V regions, respectively.

Structural similarity search

Leaving out the double-stranded antiparallel β-sheet (β2 and β3), the fold of the N-terminal domain is similar to a Greek-key motif. In contrast, the C-terminal domain is an unusual motif. To check whether porcine MSP was a novel fold structure, we used DALI, CATH and CE methods to perform a structural comparison with available structures in the Protein Data Bank. Neither CATH nor CE found any structure similar to that of porcine MSP, but the DALI search showed two structures, fibronectin (1E8B) and integrin (1JV2), with Z-score value greater than 2.0. Structural comparison showed that the structural similarity between porcine MSP with fibronectin or integrin were particularly located in the N-terminal domain. Superimposition of the matched regions of

fibronectin (160 residues, Z-score = 2.9) with the N-terminal 44 residues of porcine MSP gave an RMSD of 8.85 Å for the backbone atoms (Figure 8), while the RMSD is 3.29 Å for the matched regions of integrin (927 residues, Z-score = 2.0) superimposed with the N-terminal 42 residues of porcine MSP (data not shown). Accordingly, the structural similarity between these two proteins and porcine MSP was quite poor, in terms of the overall structure fold, so that we concluded that porcine MSP is a novel fold structure, especially the unusual C-terminal domain.

Discussion

The CD spectra revealed that porcine MSP cross-linked with five disulfide bonds is composed mainly of β-strands and is thermally and chemically stable. The disulfide pairings of porcine MSP determined previously[10] were reconfirmed by structure calculation using a full set of restraints. The 3D structure showed that porcine MSP is a globular protein with two domains and that the major inter-domain interactions are probably electrostatic. The domain orientation in the structures calculated with or without RDC restraints is highly similar, confirming the accuracy of tertiary fold in porcine MSP. The β-methylene protons of Cys[37] and Cys[46] were found to have very up-field chemical shifts at 0.56/1.59 ppm and 1.72/2.18 ppm,
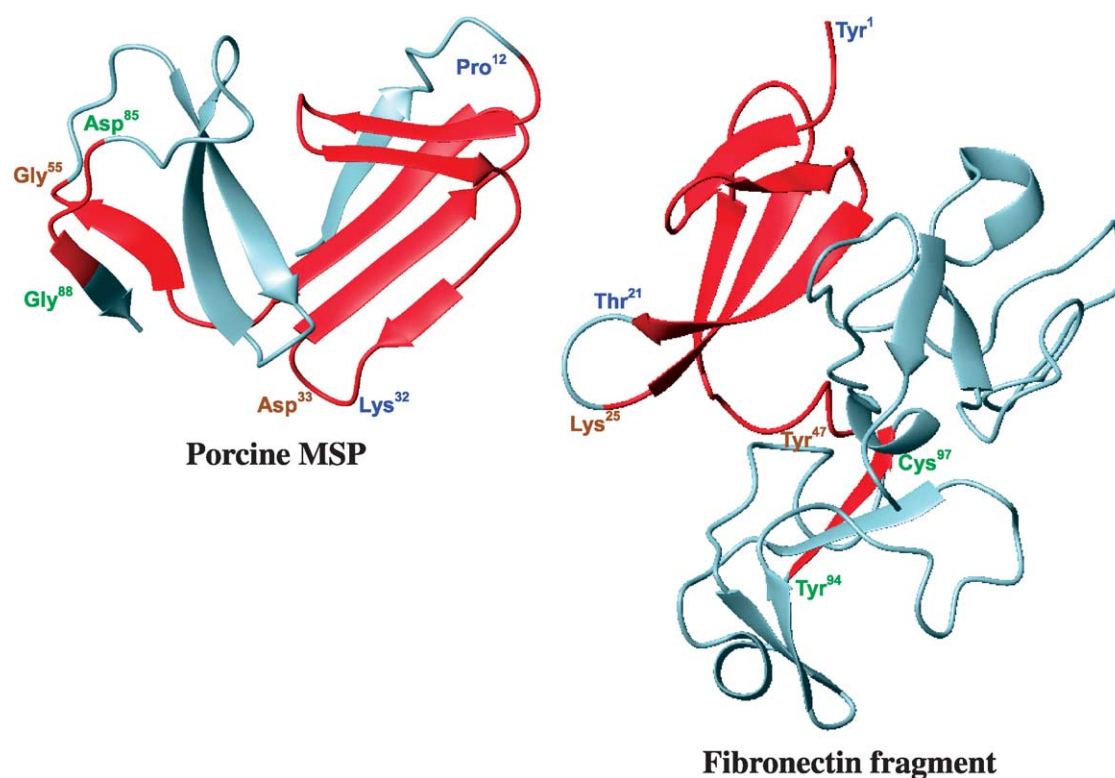

Figure 8. Comparison of the 3D structures of porcine MSP and a fibronectin fragment (PDB number 1E8B). Similar regions in the two structures are shown in red and the ranges of each region used for comparison are labeled using different color codes.
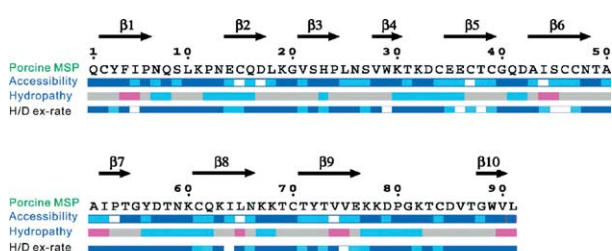
Figure 9. Comparison of the amide proton exchange rate with the accessibility and hydropathy values obtained using ENDscript. Relative accessibility is shown using different colors, with white indicating buried residues (<0.1), light-blue intermediate (0.1–0.4), blue accessible (>0.4), and blue with red borders highly exposed (>1). For hydropathy, pink indicates hydrophobic, gray intermediate, and light-blue hydrophilic. The amide proton exchange rates are colored in the same way as the accessibility data. Slowly exchanging residues are shown in white and rapidly exchanging residues in blue.

respectively. These unusual chemical shifts could be nicely explained by the ring-current effect from the tertiary structure that clearly shows Cys[37] is close to Trp[29] and Cys[46] has Tyr[3] nearby. Furthermore, the ring current shifts measured on the basis of the Johnson–Bovey model[16] with a distance cutoff of 5 Å using the program MOLMOL[17] indicated that the unusual shifts of Cys[37] and Cys[46] are indeed affected by the aromatic residues of Trp[29] and Tyr[3], respectively.

Comparison of the exchange rate with the accessibility and hydropathy properties of the averaged structure generated by ENDscript[18] (Figure 9) showed that these data are in good agreement with each other. It clearly indicated that there is no slowly exchanged amide proton and hydrophobic core in the inter-domain regions, evidence that both the H-bonds and hydrophobic core are not playing important role in the domain interaction. However, the exchange rate and accessibility seem to conflict with the high thermal and chemical stabilities for porcine MSP. Why the protein is thermally and chemically stable, although containing a lot of solvent-accessible regions, could be explained using the electrostatic interaction and disulfide bonds. As mentioned above, porcine MSP contains a number of charged residues, and the interaction between two domains is primarily due to the electrostatic force. It is understood that the electrostatic interaction is difficult to be completely disrupted at high temperatures and high denaturant concentrations except for changing the salts or pH values. For that reason, we applied CD experiments to check the conformational effect for porcine MSP under high NaCl concentrations and found more proteins are denatured under high-salt concentrations (see Figure S2 in Supplementary Material). This finding demonstrates that the electrostatic interaction is important for the stability

in the tertiary structure of MSP. In addition, the five disulfide pairings can contribute the energy for stabilizing the secondary conformation. Therefore, porcine MSP is very thermally and chemically stable, and the majority of the secondary structures stabilized by the electrostatic forces as well as the five disulfide bridges could still be seen at high temperatures and at high denaturant concentrations.

The structure of porcine MSP is composed of two domains that are covalently connected by the disulfide bond of Cys[34]-Cys[70] and by the linker at loop III (residues 49–51). Neither H-bonds nor hydrophobic interactions could be detected within two domains and hence one might consider that this linker should be flexible. However, compared the order parameters of the linker to those of other loops, loop III has higher order parameters and therefore is more rigid. This result indirectly provides evidence that the electrostatic interactions of inter-domains help to stabilize the conformation of porcine MSP.

Our collaborator found that FITC-labeled porcine MSP binds to the sperm membrane (unpublished results). To check whether porcine MSP interacts with the membrane or not, we applied CD and/or NMR experiments to study this interaction by using three membrane mimetic solutions, which are sodium dodecyl sulfate (SDS), dodecylphosphocholine (DPC) and 1,2-diheptanoyl-sn-glycero-3-phosphocholine (DHPC) micelles. Both CD and NMR spectra demonstrated that porcine MSP has a dramatic conformational change in SDS micelle, indicating that there is an interaction between them. However, the conformational change could not be seen in DPC and DHPC micelles (see Figure S3 in Supplementary Material), even at concentrations much higher than their CMC (critical micelle concentration). Taken together, we concluded that there is no hydrophobic interaction between porcine MSP and any of the SDS, DPC or DHPC micelles, and that the interaction between porcine MSP with SDS micelle might be due to electrostatic forces. Because the electrostatic forces play an important role in the stability of porcine MSP, it is likely that the negatively charged head group of SDS is able to disrupt the electrostatic interactions of porcine MSP, which subsequently causes the structural unfolding into a random coil. In contrast, both DPC and DHPC possess the same charged group, but this is different from that of SDS. It seems that the electrostatic interaction between DPC and DHPC with porcine MSP is too weak to cause any structural change in porcine MSP. It is therefore clear that SDS actually acts as a denaturing agent to disrupt the tertiary structure of porcine MSP. As a result, the direct interaction between porcine MSP with the membrane is unlikely to occur.

No structure similar to the whole tertiary fold of porcine MSP was found in the database, revealing that porcine MSP is a novel fold structure. Although the structure of the N-terminal domain

resembles the well known Greek-key motif, the fold of the C-terminal domain, which contains conserved disulfide bridges (Cys[34]-Cys[70] and Cys[61]-Cys[84]) and the hydrophobic core in all MSPs, is unusual. Because the N-terminal 61 residue domain of human MSP was suggested as a specific biological marker for the diagnosis of BPH,[8] it is reasonable to deduce the biological function for the N-terminal domain of porcine MSP based on the Greek-key motif. The Greek-key motif is usually involved in β-barrel or immunoglobulin-like β-sandwich structures and is one of the most common protein folds adopted by a variety of superfamilies, such as immunoglobulins,[19] cadherins,[20] fibronectin-type III domains,[21,22] and cupredoxins.[23] Most of these proteins belong to cell–cell adhesion molecules, which are related to several kinds of protein–protein interactions. These Greek-key motifs often pack face-to-face as double-layered β-structures in β-barrels or β-sandwiches[24] by means of hydrophobic cores or disulfide bonds connection.[25] In addition, the Greek-key motifs are observed in the structures defined as open-faced β-sheets[26] that possess diverse biological functions. Since porcine MSP exhibits the inhibitions of Na$^+$,K$^+$-ATPase activity[5] and of sperm motility,[27] it is probable that porcine MSP may belong to a member of peripheral membrane proteins that can interact with the integral membrane proteins, such as sodium pump (Na$^+$,K$^+$-ATPase). In addition, porcine MSP is likely to be involved in the protein–protein interactions to play a role in modulation of sperm motility.

In addition to having the conserved Cys residues between porcine and human MSPs, based on sequence alignment and the tertiary structure determined in this study, we further found that the residues in the secondary structure regions are highly conserved. Also, residues Asp[17]–Gly[20] and Glu[76]–Asp[79] located at the interaction sites between two domains in porcine MSP are identical with the corresponding residues in human MSP, and hydrophobic core residues (Pro[53], Tyr[56], Val[74] and Val[86]) are also identical between them. Moreover, GYD (Gly[55]–Asp[57]) in loop IV and KTC (Lys[82]–Cys[84]) in loop V, adjacent to the hydrophobic core, are also the same. Taken together, we suggested that human and porcine MSPs may have a similar structure topology and fold, even though the sequence identity is only 51%. Consequently, structural characteristics between porcine and human MSPs are very alike, and the structural information obtained in this work for porcine MSP may be applied for human MSP. Given the numerous putative biological functions suggested for MSPs and the highly conserved residues seen in the secondary structure regions between human and porcine, the tertiary structure of porcine MSP may provide valuable clues about the biological function of the MSP superfamily and also shed light on the role of MSP in prostate pathophysiology.

Materials and Methods

Preparation of porcine MSP

Native porcine MSP was obtained as previously described.[5] Recombinant porcine MSP was expressed at 37 °C in *Escherichia coli* M15 using the vector pQE30. To facilitate protein expression and purification, the recombinant MSP contained an extra 17 amino acid residues including a 6-His tag at the N terminus. Protein uniformly labeled with ^{15}N or ^{15}N/^{13}C was obtained by growing the *E. coli* cells in Mops medium supplemented with ^{15}NH$_4$Cl and/or [^{13}C]-glucose, expression being induced by addition of 2 mM isopropyl-β-thiogalactopyranoside (IPTG) when the $A_{600\,nm}$ reached 0.6. The bacteria were then disrupted using an M-110S microfluidizer (Microfluidics) and the supernatant recovered by centrifugation at 50,000 *g* for 30 minutes at 4 °C. The protein was applied to a nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography (QIAGEN), which was washed with washing buffer (5 mM imidazole in 50 mM sodium phosphate, 200 mM sodium chloride, pH 8.0) to remove contaminants, then eluted with washing buffer containing 120 mM imidazole. The purity of the eluted protein on a Coomassie blue-stained SDS/polyacrylamide gel was better than 90%. To further purify the protein, reversed-phase high-performance liquid chromatography (HPLC) was performed on a Vydac 214TP54 C4 reversed phase column (5 μm particle size, 4.6 mm × 250 mm) using a linear gradient of 17%–23% buffer B at a flow rate of 1 ml/minute over 20 minutes (buffer A, 0.1% aqueous trifluoroacetic acid (TFA); buffer B, 0.1% TFA in acetonitrile) monitoring at $A_{254\,nm}$. The purity of the eluted protein assessed by SDS-PAGE was greater than 95% and further dialyzed with 50 mM phosphate buffer at pH 5.0 for NMR study. For the amide proton exchange rates study, protein solutions were lyophilized into powder and then redissolved in ^{2}H$_2$O.

Circular dichroism (CD) experiment

CD experiments were performed at 25 °C using 20 μM MSP in 20 mM phosphate buffer in a 1 mm path-length cuvette in an Aviv 202 SF CD spectrometer (Lakewood, NJ, USA) previously calibrated with *d*-10-camphorsulfonic acid (CSA). The steady-state CD spectra at different temperatures and pH values were recorded from 180 nm to 260 nm at a scanning rate of 38 nm/minute with a wavelength step of 0.5 nm and a time constant of 100 ms. All CD data were obtained from an average of three scans with 1 nm bandwidth. After background subtraction and smoothing, all the CD data were converted from CD signal (millidegree) into mean residue ellipticity (deg cm^2 dmol^{-1}). Equilibrium thermal-denaturing experiments were performed by measuring the change of molar ellipticity at 231 nm. Data (231 nm) were collected at different temperatures over a range of 10–96 °C in 20 mM phosphate buffer, pH 5.0, using a scan rate of 2 deg. C/minute after three minutes equilibration at each point with a temperature controller. For the chemical denaturation experiment, guanidine hydrochloride (Gdn·HCl) was used and the change of molar ellipticity at 231 nm was monitored using 5 μM MSP in 20 mM phosphate buffer, pH 5.0 in a 1 cm path-length cuvette at 25 °C. The CD spectra of porcine MSP at different concentrations of micelles (SDS, DPC, and DHPC) were performed using 10 μM MSP in 20 mM phosphate buffer at pH 5.0 and 25 °C.

NMR experiment and resonance assignments

All NMR experiments were carried out at 310 K on a Bruker AVANCE 600 or AMX 500 NMR spectrometer (Bruker, Karlsruhe, Germany) equipped with a triple (^{1}H, ^{13}C and ^{15}N) resonance probe, including a shielded z-gradient. The sample (0.35 ml of 1.5 mM protein) was prepared in 50 mM phosphate buffer in 90% H_2O/10% ^2H_2O at pH 5.0 and placed in a Shigemi NMR tube (Allison Park, PA, USA). Two-dimensional (2D) ^{1}H NMR, COSY,[28] TOCSY,[29] and NOESY[30] spectra were collected for native and recombinant porcine MSP. All heteronuclear NMR experiments on recombinant MSP were carried out as described.[31] Assignment of the main-chain H^N, N^H, C^α, $C\beta$, and C' chemical shifts was based on NHCACB, CBCA(CO)NH, HNCO, and HN(CA)CO experiments. Assignment of side-chain resonances was based on the ^{1}H-^{15}N TOCSY-HSQC, ^{1}H-^{13}C HCCH-TOCSY, HC(CO)NH, and C(CO)NH spectra. 2D ^{1}H-^{15}N HSQC and 3D NOESY-HSQC experiments were used to assign side-chain amide resonances of Asn and Gln residues. Aromatic resonances were assigned using 2D ^{1}H-^{13}C HSQC, TOCSY, and NOESY spectra. H/D exchange rates of amide protons were determined by recording several 2D TOCSY spectra and a series of ^{1}H-^{15}N HSQC spectra at different times after the protein was dissolved in ^2H_2O at 310 K, pH 5.0. The amide proton exchange rates were categorized into the three classes of fast, medium, and slow. Hydrogen bond formation or solvent exclusion was assumed for amide protons with slow or medium exchange rates. The 2D ^{15}N-HSQC experiments were also carried out for the recombinant ^{15}N-labeled MSP in the presence of SDS and DPC micelles with the concentration ratio of 1:100 in 20 mM phosphate buffer at pH 5.0. 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as the external chemical shift standard at 0.00 ppm. The ^{15}N and ^{13}C chemical shifts were indirectly referenced using the consensus Ξ ratios of the zero-point frequencies at 310 K.[32] All spectra were processed using XWIN-NMR and analyzed using AURELIA on an SGI O_2 workstation. Linear prediction was used in the ^{13}C and ^{15}N dimensions to improve the digital resolution: 5.3 mg/ml of Pf1 filamentous phage[33] (*Asla*, Riga, Latvia) was added to ^{13}C,^{15}N-labeled porcine MSP in 50 mM phosphate buffer in 90% H_2O/10% ^2H_2O at pH 7.7 to partially align the Z-axis of the protein in the magnetic field. The $^1J_{NH}$ splittings were obtained for the free and aligned samples by IPAP ^{15}N-^{1}H HSQC experiments.[34] RDC values were measured by taking the difference between the correlated J-coupling observed in the aligned and isotropic solvents. The anisotropy and rhombicity estimated from the distribution of the observed RDC values[35] were 29.9 Hz and 0.5, respectively.

Structure calculation and analysis

NOE distance restraints for porcine MSP were derived based on 3D ^{15}N-edited NOESY-HSQC in aqueous solution and ^{13}C-edited NOESY-HSQC in 100% ^2H_2O recorded at 310 K, pH 5.0. Peak intensities were classified as large, medium, small, and very small, corresponding, respectively, to upper bound interproton distance restraints of 2.5 Å, 3.5 Å, 4.5 Å, and 6.0 Å. An additional correction of 1.0 Å was added for methylene and methyl groups. The backbone ϕ,ψ torsion angles were calculated using the TALOS program,[36] and those in good agreement with the NOE correlations were used for structure generations. The structure determination was performed using 891 distance restraints, 36 hydrogen bonds, and 91 torsion angle restraints. All minimization and dynamical simulated annealing calculations were carried out using the program XPLOR-NIH[37] on a Linux workstation. The force constants and molecular parameters were set to the default values as in the original sa_new.inp protocol in XPLOR-NIH. A total of 49 RDC restraints were used for the structure refinement with an error range of ± 1.0 Hz. The final force constant for RDC was 0.5 kcal mol^{-1} Hz^{-1}, while other force constants were the same as the default values. The INSIGHT II (Molecular Simulation Inc., San Diego, CA) and MOLMOL programs were used to display sets of structures and to generate the electrostatic surface potential of the final three-dimensional models. The distributions of the backbone dihedral angles of the final converged structures were evaluated by the Ramachandran dihedral pattern using PROCHECK-NMR[38] and MOLMOL.

Relaxation data analysis

All relaxation data were acquired at 310 K on a Bruker AVANCE 600 NMR spectrometer. Standard pulse sequences were employed to measure ^{15}N-T_1, ^{15}N-T_2[39] and steady state heteronuclear (^{1}H-^{15}N) NOE.[40] The relaxation delays in the T_1 experiments were 15, 45, 90, 150, 300, 480, 680, 900, 1500 and 2200 ms, and those in the T_2 experiments were 16.48, 48.44, 82.4, 131.84, 197.76, 296.64, 385.52, 494.4, 659.2 and 824 ms. ^{1}H-^{15}N steady-state heteronuclear NOE values were measured by recording spectra with or without a ^{1}H saturation period of three seconds. These spectra were processed using NMRPipe and analyzing by NMRView.[41] To determine the R_1 (spin–lattice) and R_2 (spin–spin) relaxation rates, resonance intensities were extracted and fitted to a non-linear, least-squares curve as a function of the relaxation delay time using the Levenberg–Marquardt non-linear fitting routines in NMRview. The uncertainties in the measured peak-height were set equal to the root-mean-square base-line noise in the spectra, and the errors in NOE were the standard deviation of two sets of repeated NOE experiments. After collecting the R_1, R_2, and NOE relaxation data, an axially symmetric diffusion tensor was employed and the atomic coordinates were obtained from the final averaged NMR structure followed by translation to the center of mass with the program pdbinertia†. The program r2r1_diffusion[42] was then used to estimate an axially symmetric diffusion tensor and orient the coordinate system of the PDB file of porcine MSP to the principal axis of the diffusion tensor. This final PDB file was then used as input to the FAST-Modelfree program.[43] The NH bond length and the ^{15}N-CSA values were identical with the defaults and the cutoff value of Γi was set to 0.80. The selected models and generated order parameter for each residue were checked and compared using a manual model selection strategy.[44]

Data Bank accession number

The chemical shifts of recombinant porcine MSP at pH 5.0 and 310 K have been updated and redeposited to BioMagResBank under accession number BMRB-5565. The 20 NMR structures and the complete list of restraints have been deposited in the RCSB Protein Data Bank under accession number 1XHH.

---

† http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software

Acknowledgements

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2004.12.029

References

1. Makinen, M., Valtonen-Andre, C. & Lundwall, A. (1999). New World, but not Old World, monkeys carry several genes encoding beta-microseminoprotein. *Eur. J. Biochem.* **264**, 407–414.

2. Xuan, J. W., Kwong, J., Chan, F. L., Ricci, M., Imasato, Y., Sakai, H. *et al*. (1999). cDNA, genomic cloning, and gene expression analysis of mouse PSP94 (prostate secretory protein of 94 amino acids). *DNA Cell Biol.* **18**, 11–26.

3. Xuan, J. W., Wu, D., Guo, Y., Garde, S., Shum, D. T., Mbikay, M. *et al*. (1997). Molecular cloning and gene expression analysis of PSP94 (prostate secretory protein of 94 amino acids) in primates. *DNA Cell Biol.* **16**, 627–638.

4. Nolet, S., St-Louis, D., Mbikay, M. & Chretien, M. (1991). Rapid evolution of prostatic protein PSP94 suggested by sequence divergence between rhesus monkey and human cDNAs. *Genomics*, **9**, 775–777.

5. Chao, C. F., Chiou, S. T., Jeng, H. & Chang, W. C. (1996). The porcine sperm motility inhibitor is identical to beta-microseminoprotein and is a competitive inhibitor of Na$^+$,K$^+$-ATPase. *Biochem. Biophys. Res. Commun.* **218**, 623–628.

6. Mundle, S. D. & Sheth, N. A. (1993). Suppression of DNA synthesis and induction of apoptosis in rat prostate by human seminal plasma inhibin (HSPI). *Cell Biol. Int.* **17**, 587–594.

7. Shukeir, N., Arakelian, A., Kadhim, S., Garde, S. & Rabbani, S. A. (2003). Prostate secretory protein PSP94 decreases tumor growth and hypercalcemia of malignancy in a syngenic *in vivo* model of prostate cancer. *Cancer Res.* **63**, 2072–2078.

8. Xu, K., Wang, X., Ling, M. T., Lee, D. T., Fan, T., Chan, F. L. *et al*. (2003). Identification of a specifically expressed modified form of novel PSP94 protein in the secretion of benign prostatic hyperplasia. *Electrophoresis*, **24**, 1311–1318.

9. Lazure, C., Villemure, M., Gauthier, D., Naude, R. J. & Mbikay, M. (2001). Characterization of ostrich (Struthio camelus) beta-microseminoprotein (MSP): identification of homologous sequences in EST databases and analysis of their evolution during speciation. *Protein Sci.* **10**, 2207–2218.

10. Wang, I., Yu, T. A., Wu, S. H., Chang, W. C. & Chen, C. (2003). Disulfide pairings and secondary structure of porcine beta-microseminoprotein. *FEBS. Letters*, **541**, 80–84.

11. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595–603.

12. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). CATH–a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.

13. Shindyalov, I. N. & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747.

14. Sreerama, N. & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260.

15. Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR*, **4**, 171–180.

16. Case, D. A. (1995). Calibration of ring-current effects in proteins and nucleic acids. *J. Biomol. NMR*, **6**, 341–346.

17. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

18. Gouet, P. & Courcelle, E. (2002). ENDscript: a workflow to display sequence and structure information. *Bioinformatics*, **18**, 767–768.

19. Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320.

20. Shapiro, L., Kwong, P. D., Fannon, A. M., Colman, D. R. & Hendrickson, W. A. (1995). Considerations on the folding topology and evolutionary origin of cadherin domains. *Proc. Natl Acad. Sci. USA*, **92**, 6793–6797.

21. Hamill, S. J., Steward, A. & Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165–178.

22. Cota, E., Steward, A., Fowler, S. B. & Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J. Mol. Biol.* **305**, 1185–1194.

23. Arnesano, F., Banci, L., Bertini, I. & Thompsett, A. R. (2002). Solution structure of CopC: a cupredoxin-like protein involved in copper homeostasis. *Structure (Camb.)*, **10**, 1337–1347.

24. Zhang, C. & Kim, S. H. (2000). A comprehensive analysis of the Greek key motifs in protein beta-barrels and beta-sandwiches. *Proteins: Struct. Funct. Genet.* **40**, 409–419.

25. Feige, M. J., Walter, S. & Buchner, J. (2004). Folding mechanism of the CH2 antibody domain. *J. Mol. Biol.* **344**, 107–118.

26. Zhang, C. & Kim, S. H. (2000). The anatomy of protein beta-sheet topology. *J. Mol. Biol.* **299**, 1075–1089.

27. Jeng, H., Liu, K. M. & Chang, W. C. (1993). Purification and characterization of reversible sperm motility inhibitors from porcine seminal plasma. *Biochem. Biophys. Res. Commun.* **191**, 435–440.

28. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wuthrich, K. (1983). Improved spectral resolution in COSY ^{1}H NMR spectra of proteins *via* double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485.

29. Bax, A. D. & Davis, G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355–360.

30. Kumar, A., Ernst, R. R. & Wuthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.

31. Kay, L. E. (1995). Pulsed field gradient multi-dimensional NMR methods for the study of protein structure and dynamics in solution. *Prog. Biophys. Mol. Biol.* **63**, 277–299.

32. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E. et al. (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.

33. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065–1074.

34. Ottiger, M., Delaglio, F., Marquardt, J. L., Tjandra, N. & Bax, A. (1998). Measurement of dipolar couplings for methylene and methyl sites in weakly oriented macromolecules and their use in structure determination. *J. Magn. Reson.* **134**, 365–369.

35. Clore, G. M., Gronenborn, A. M. & Bax, A. (1998). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133**, 216–221.

36. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.

37. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Marius Clore, G. (2003). The XPLOR-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73.

38. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.

39. Kay, L. E., Nicholson, L., Delaglio, F., Bax, A. & Torchia, D. (1992). Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T_1 and T_2 values in proteins. *J. Magn. Reson.* **97**, 359–375.

40. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992). Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, **31**, 5269–5278.

41. Johnson, B. A. & Blevins, R. A. (1994). NMRVIEW: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

42. Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562–12566.

43. Cole, R. & Loria, J. P. (2003). FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR*, **26**, 203–213.

44. Mandel, A. M., Akke, M. & Palmer, A. G., III (1995). Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144–163.