doi:10.1016/j.jmb.2006.07.029 J. Mol. Biol. (2006) 362, 502–515

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Solution Structures of Human and Porcine β-Microseminoprotein

Houman Ghasriani[1], Kaare Teilum[1], Ylva Johnsson[2] Per Fernlund[2] and Torbjörn Drakenberg[1]*

[1]*Department of Biophysical Chemistry, Lund University P.O. Box 124, SE-221 00 Lund, Sweden*

[2]*Department of Clinical Chemistry, Lund University Wallenberg Laboratory, entr 46 Floor 5, Malmö University Hospital, SE-205 05 Malmö, Sweden*

**Corresponding author*

β-Microseminoprotein (MSP) is a small cysteine-rich protein (molecular mass about 10 kDa) first isolated from human seminal plasma and later identified in several other organisms. The function of MSP is not known, but a recent study has shown MSP to bind CRISP-3, a protein present in neutrophilic granulocytes. The amino acid sequence is highly variable between species raising the question of the evolutionary conservation of the 3D structure. Here we present NMR solution structures of both the human and the porcine MSP. The two proteins (sequence identity 51%) have a very similar 3D structure with the secondary structure elements well conserved and with most of the amino acid substitutions causing a change of charge localized to one side of the molecule. MSP is a β-sheet-rich protein with two distinct domains. The N-terminal domain is composed of a four-stranded β-sheet, with the strands arranged according to the Greek key-motif, and a less structured part. The C-terminal domain contains two two-stranded β-sheets with no resemblance to known structural motifs. The two domains, connected to each other by the peptide backbone, one disulfide bond, and interactions between the N and C termini, are oriented to give the molecule a rather extended structure. This global fold differs markedly from that of a previously published structure for porcine MSP, in which the two domains have an entirely different orientation to each other. The difference probably stems from a misinterpretation of ten specific inter-domain NOEs.

Keywords: cysteine-rich protein; prostate cancer; PSP94; seminal plasma; three-dimensional structure

Introduction

β-Microseminoprotein (MSP) is a hydrophilic, non-glycosylated, cysteine-rich protein with a molecular mass of around 10 kDa. It was first isolated from human seminal plasma.[1] Human MSP (hMSP) is secreted by the epithelial cells of the prostate gland[2] giving rise to the alternative name prostatic secretory protein of 94 amino acids (PSP94).[3] The expression of MSP is, however, not limited to the prostate gland[4] and its number of amino acid residues varies between species.

MSP proteins have been isolated and identified by amino acid sequencing from man,[5,6] chicken,[7] pig,[8] rat,[9] and ostrich.[10] Gene and/or cDNA sequences have been obtained from man,[11–14] rhesus monkey,[13,15] pig,[16] rat,[9] baboon,[17] cotton-top tamarin,[18] and mouse.[19] The amino acid sequences of the various MSPs do not suggest homology of MSP to any other protein in GenBank.

Many functions of MSP have been suggested[1,20–24] but none based on firm experimental evidence. The most promising observation that could lead to an understanding of the biology and function of MSP is the recent finding that hMSP binds to a protein in human blood (PSP94-binding protein)[25] and to CRISP-3 from human leukocytes.[26] Consequently, it is now possible to study the interaction of MSP with natural ligands motivating the determination of the three-dimensional structure of hMSP.

Abbreviations used: MSP, β-microseminoprotein; hMSP, human MSP; pMSP, porcine MSP; RDC, residual dipolar coupling; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TOCSY, total correlated spectroscopy; HSQC, heteronuclear single quantum correlation; RP-HPLC, reversed phase high pressure liquid chromatography.

E-mail address of the corresponding author: torbjorn.drakenberg@bpc.lu.se

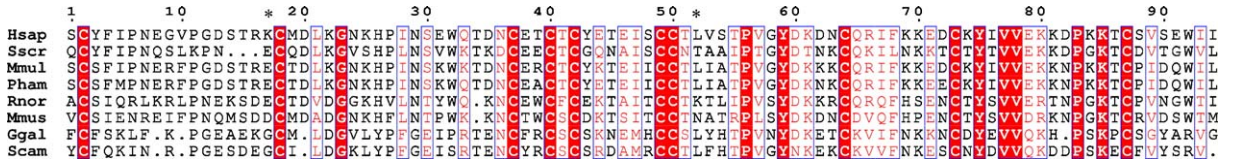

Figure 1. The amino acid sequences of mature MSP from human (Hsap), pig (Sscr), rhesus monkey (Mmul), baboon (Pham), rat (Rnor), mouse (Mmus), chicken (Ggal), and ostrich (Scam). Numbering is according to the human sequence. Stars indicate the locations of the exon 2/exon 3 and exon 3/exon 4 boundaries, respectively, for the human protein. Gaps have been inserted in the pig, rat, mouse, chicken, and ostrich sequences to allow proper alignment of the sequences. Totally conserved residues are white on red background and those with similarity are in light red. The alignment was done with the Espript program[54] at http://espript.ibcp.fr/ESPript/ using the default values.

A notable characteristic of MSP is that its amino acid sequence is highly variable between species (Figure 1). Except for the ten cysteine residues, only very few amino acid residues are totally conserved in all species that have been investigated.[10] It would, therefore, be interesting to see how conserved the three-dimensional structure of MSP is between different species. Therefore, we have determined the structures of both human and porcine MSP. Porcine MSP (pMSP) was first isolated as a sperm motility inhibitor[27] and later identified as MSP[21] after its amino acid sequence became available.[8] The porcine MSP is three amino acids shorter than hMSP and the sequence identity between the two proteins is 51%.

During our work, a solution structure of pMSP was published.[28] Although the secondary structure elements were practically identical to what we had found, the published structure showed a domain orientation of pMSP that differed considerably from that of our structure. The domain orientation that we observe for pMSP is almost identical to the one we observe for hMSP.

Here we report our experimental NMR data and our calculated structures of human and porcine MSP together with a detailed comparison of our structures of the two MSPs. We also provide a probable explanation for the difference of our pMSP structure from that previously reported.

Results

Protein production

The dominating form (P1) of pMSP isolated from porcine seminal plasma,[8] which is the native protein used in these studies, has an Asp residue in position 42 (GenBank accession no. S41663), whereas the only published cDNA sequence[16] for pMSP (GenBank accession no. S80724) codes for an Asn residue. However, cloning of cDNA from porcine prostate tissue revealed two different cDNA sequences. One was identical to the published cDNA sequence while the second coded for the same protein but with Asp in position 42. Consequently, the second cDNA was used to construct the vector for the expression of recombinant pMSP.

The recombinant MSPs were both expressed with a spacer of three extra glycine residues between the factor Xa cleavage site of the poly-His/S-protein tag and the N-terminal residue of the mature protein. This facilitated the efficient removal of the tag by factor Xa digestion. Preliminary experiments had shown that, at least for the porcine protein, the tag could not be removed when an extension was not present.

In typical expressions, 27 and 49 mg of poly-His/S-protein-tagged protein was obtained from one liter of culture for pMSP and hMSP, respectively. More than 50% of the protein was correctly folded according to analysis by RP-HPLC. Digestion of this material with activated factor X followed by RP-HPLC purification yielded 7 and 15 mg of recombinant pMSP and hMSP, respectively, with more than 95% purity.

Resonance assignment

The sequential backbone assignment of recombinant MSP was made according to standard procedures using three-dimensional hetero-nuclear NMR spectra (HNCA, HN(CO)CA, HNCO, and HN(CO)CACB). Thanks to good signal dispersion and narrow lines almost complete backbone assignment was obtained for both proteins. The side-chain assignment was based on 3D ^{15}N HSQC-TOCSY spectra in $H_2O/^2H_2O$ (90/10) as well as 2D TOCSY in 100% ^2H_2O. Also the side-chain assignment was close to complete. In total, 98% of the shifts were assigned for both proteins. The resonance assignment of native MSP was based on 2D-TOCSY and 2D-NOESY spectra, resulting in a 95% and 93% completeness of the backbone proton assignments for pMSP and hMSP, respectively.

Structure calculation

Initially, we used the program CNS and manually assigned NOESY cross-peaks that were integrated with the Sparky program and converted into upper distance restraints using some fixed distances as reference. An r^{-6} dependence of intensity on distance was assumed. In the final CNS calculations we used ca 1000 NOE restraints, 59 φ-angle restraints based on $^3J_{HN\text{-}H\alpha}$ coupling constants and 20 hydrogen bonds. We also used the S-S bond arrangement (2-47, 15-39, 34-70, 37-46 and 61-84) reported for pMSP by Wang *et al.*[29] Our structures deviate in one obvious way from the structure of