Journal of Biomolecular NMR **29:** 399–402, 2004.

Letter to the Editor: NMR structure of the Apo-S100P protein

Yi-Chien Lee[a], David E. Volk[a], Varatharasa Thiviyanathan[a], Quinn Kleerekoper[a,*], Alexey V. Gribenko[b], Shanmin Zhang[a], David G. Gorenstein[a], George I. Makhatadze[b] & Bruce A. Luxon[a,**]

[a]*Sealy Center for Structural Biology and Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX 77555, U.S.A.;* [b]*Department of Biochemistry and Molecular Biology, Penn State University College of Medicine, Hershey, PA 17033, U.S.A.*

Received 19 February 2004; Accepted 16 March 2004

Key words: apo-S100P, NMR structure, homodimer

Biological context

Calcium ions play important roles in the regulation of cell growth (Heizmann and Cox, 1998). Calcium binding proteins control the amount of calcium required at different stages of cell development. The S100 proteins are small dimeric members of the EF-hand superfamily of Ca^{2+} binding proteins that mediate Ca^{2+} dependent signal transduction pathways involved in cell growth and differentiation, cell cycle regulation and metabolic control (Donato, 2001). S100 proteins have frequently been associated with a number of neurological diseases, neoplastic diseases, human cardiac diseases and tumor development.

Over twenty S100 proteins have been identified with distinct functions and tissue distribution (Donato, 2001). However, the three dimensional structures of only several dimeric S100 proteins have been determined by NMR (S100B (Drohat et al., 1999), S100A1 (Rustandi et al., 2002), S100A6 (Mäler et al., 1999)) method. Human S100P, a 95 amino acid residue protein first isolated in 1992 (Emoto et al., 1992), has 50% and 35% sequence identity with S100B and calcyclin (S100A6), respectively (Gribenko and Makhatadze, 1998). Protein expression studies have shown that the different amounts of S100P in androgen-dependent and androgen-independent prostate cell lines might mediate different pathways of prostate cancer development (Averboukh et al., 1996). Among the S100 family of proteins, S100P is the least studied protein probably due to its unknown function and the absence of a clear correlation with human diseases. Recently, it has also been reported that S100P overexpression is related with immortalization of human breast epithelial cells in vitro and in the early stages of breast cancer development in vivo (Guerreiro Da Silva et al., 2000). Several recent studies report over-expression S100P in several carcinoma, including pancreatic cancers (Jurcevic et al., 2003). These results suggest that S100P could be a useful and significant biomarker.

Methods and results

Details of the cDNA cloning, over-expression and isolation of the human S100P have been described previously (Gribenko et al., 1998). Briefly, the protein was expressed using a T7 expression system in *E. coli* strain BL21(DE3). A two-column purification scheme using a Fast Flow Q-Sepharose column (2 × 10 cm) and a Sephadex G-75 column (2.5 × 100 cm) was used. The ^{15}N-labeled and $^{15}N/^{13}C$-labeled samples for NMR experiments were prepared in a buffer containing 20 mM Tris-d_{11} (pH 6.8), 100 mM KCl, and 0.01 mM NaN_3 in 90% H_2O and 10% or 100% D_2O. The final protein concentration was 2.5–3.0 mM.

All NMR spectra were collected at 25 °C on Varian Unity Plus 750 or 600 MHz instruments processed with FELIX (MSI-Biosym). Sequence-specific backbone resonance assignments were determined using 3D HNCACB, HNCA and CBCA(CO)NH

*Present address: Department of Biochemistry & Molecular Biology, Structural Biology Research Center, University of Texas, Houston Medical School, Houston, TX 77030, U.S.A.

**To whom correspondence should be addressed. E-mail: bruce@nmr.utmb.edu

The coordinates for the structures have been deposited in the Protein Data Bank (PDB accession code 1OZO).

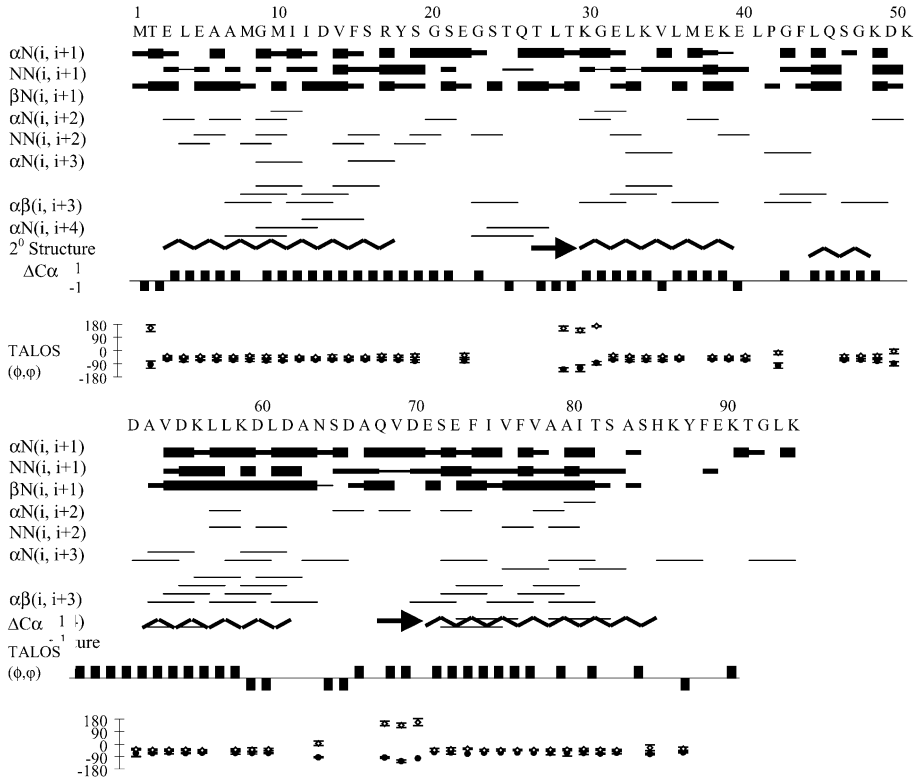

Figure 1. Diagram of sequential and short-range NOEs and secondary structure for apo-S100P. The protein sequence is listed on the top using the one-letter amino acid code. The height of the bar represents the intensity of NOE cross peaks classified as strong, medium and weak. $^{13}C_\alpha$ and $^{1}H_\alpha$ chemical shift deviation from random coil values were shown in the row labeled in ΔC_α and ΔH_α. An index of 1 indicates α-helical structure, an index of 0 indicates random coil structure and an index of −1 indicates β-sheet structure. The bar in the TALSO-predicted backbone torsion angles, ϕ (●) and φ (◊), indicates the standard deviation from the average dihedral angles of the ten residues from the chemical shift database.

experiments. The carbonyl carbon chemical shifts were determined using data from the HNCO and HN(CA)CO experiments. The side chain proton and carbon spin systems were determined using the 3D HCCH-TOCSY, C(CO)NH and HC(CO)NH experiments (Sattler et al., 1995; Grzesiek and Bax, 1993). The ^{1}H chemical shifts were referenced relative to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). The ^{1}H, ^{15}N and ^{13}C chemical shifts were referenced indirectly using the $^{1}H/X$ frequency ratios and have been deposited in the BioMagResBank (accession code BMRB-5103).

Randomized S100P subunit structures were generated using the program DIAMOD (Hanggi and Braun, 1994). These starting structures were refined using the SANDER module of AMBER6 (Case et al., 1999). NOE-based distances were determined from ^{15}N and ^{13}C-edited HSQC-NOESY experiments (Kay et al., 1992). The NOE cross peaks were first manually assigned. The NOE assignments were then further refined in an iterative molecular dynamics process incorporating the program SANE (Duggan et al., 2001) between MD calculations, with an S100P structure initially based on the published S100B protein structure (Drohat et al., 1999). Dihedral angle (ϕ and ψ) restraints were based on the chemical shift index (CSI) of the $^{1}H_\alpha$, $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts (Potts et al., 1996) or from the average values predicted by the TALOS program (Cornilescu et al., 1999). As the structure neared convergence, the X-ray structure of Ca^{2+}-bound S100P was released (Zhang et al., 2003) and this model was also used to check the final NOE assignments. A total 1552 distance restraints per monomer including 588 intra-residue, 425 sequential, 328 medium-range, 174 long-range and 36 intersubunit NOE based restraints, as well as 122 CSI- or 144 TALOS-based torsion angle constraints, were used to refine the structure (Table 1). The dimer starting structures were generated by placing two copies of each monomer structure 60 to 130 Å apart and vary-

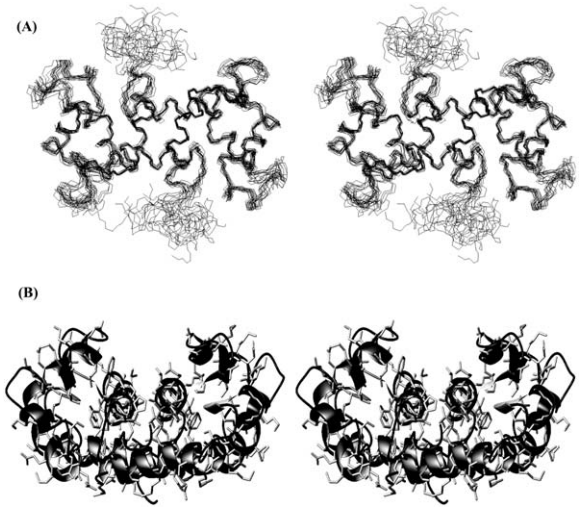

Figure 2. (A) Stereoview of the NMR structure of apo-S100P. The structures were superimposed using backbone atoms (N, C^{α} and C') of residues E3-S19, K30-E40, F44-G48, A53-A63 and E71-H86. (B) A ribbon diagram of apo-S100P with sidechain heavy atoms in the secondary structure is shown. View from the C-termini of the monomer is rotated 90° about the vertical axis.

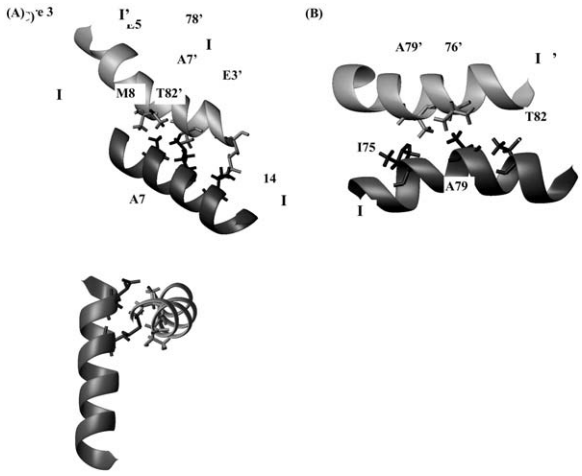

Figure 3. The dimer interface of apo-S100P. (A) The residues involved in the helix I/I′ interface, (B) the residues involved in the helix IV/IV′ interface and (C) the residues in the helix I/IV′ interface are shown.

Table 1. Statistics for the ensemble of the 16 Apo-S100P NMR structures

Distance violations > 0.5 Å	4 ± 2^{d}	4 ± 2^{e}
Torsion angle violations > 5^0	0 ± 1	2 ± 2
rms deviation from mean structure (Å)		
Secondary structure[a] (backbone)	0.69	0.88
Secondary structure (heavy)	1.20	1.58
Backbone[b,c]	1.35 (0.89)	1.38 (1.04)
Heavy atoms	2.01 (1.39)	2.10 (1.55)
Backbone dihedral angles		
In most favored region	69%	70%
In additional allowed	22%	22%
In generously allowed region	5%	4%
Disallowed region	4%	4%

[a]Residues 3–19, 25–40, 44–46, 53–63, and 71–86.
[b]Residues 3–90.
[c]Numbers in parentheses calculated based on the backbone of residues 3–86.
[c]Structures calculated based on CSI torsion angle constraints.
[d]Structures calculated based on TALOS torsion angle constraints.

ing their relative orientation. The dimeric structures were then refined using the same rMD protocol for the monomer at a lower temperature. The inter-subunit restraints were used in this simulation. The rmsd values of the ensemble structures (Figure 2) based on residues 3–90 and 3–86 were 1.35 Å and 0.89 Å for the backbone atoms and 2.01 Å and 1.39 Å for all heavy atoms, respectively, calculated from the average structure. The regions of the secondary structures were well defined with low rmsd values (0.69 Å for the backbone and 1.20 Å for the heavy atoms) (Table 1). However, the four Ca^{2+}-binding loops are less well defined than the helical regions (the rmsd of the ensemble relative to the average structure for loop regions is ∼1.31 Å). The rmsd value between the two mean monomers for the backbone atoms is low (0.57 Å), indicating high symmetry between these two monomers. Structural figures of merit of 16 apo-S100P are summarized in Table 1. The final structures have been deposited in the Protein Data Bank (PDB ID 1OZO).

Discussion and conclusions

The secondary structure of apo-S100P was predicted based on CSI and TALOS analysis, as shown in Figure 1. Both predictions indicate that S100P has a high α-helix content with four regular α-helical segments, E3-S19, K30-E40, A53-A63 and E71-H86 and one additional α-helical segment (F44-G48) in the linker region. The Ca^{2+}-binding loop I connects helices I (E3-S19) and II (K30-E40), forming the N-terminal EF-hand, while helices III (A53-A63) and IV (F71-H86) are joined by the Ca^{2+}-binding loop II, forming the C-terminal EF-hand. These two EF-hands are connected by a loop, called the linker region (L41-V54). Each Ca^{2+}-binding loop contains a short β strand (T25-T29 for strand I; Q68-D70 for strand II). The orientation of helices I and II does not change significantly with the addition of Ca^{2+} in S100 proteins

(Smith and Shaw 1998). The most dramatic change is the relative position of loop I to helices I and II. This loop moves closer to both helices and coordinates with Ca^{2+} in the bound state. This absence of a conformational change suggests that the N-terminus might not participate in the interaction with the target proteins (Webber et al., 2000). A short α-helix (F44-G48) exists in the linkage (L41-D52) between two Ca^{2+}-binding domains not observed in the linkage region of most S100 proteins except in bovine apo-S100B (residues 41–45) (Kilby et al., 1996). In the Ca^{2+}-S100P X-ray structure, the coordinates of residues 46–51 were not reported, which suggests that these residues are very dynamic, even in the crystal state, and that they are involved with target binding (Zhang et al., 2003). The conformation of the second helix-loop-helix Ca^{2+}-binding domain of the S100 family has obvious differences in the apo-state for the position of helix III relative to helices II and V (Mäler et al., 2002), such as that observed for the rat S100B, and has less homology compared to the N-terminal domain in S100 proteins (Gribenko et al., 2002). But these large changes were not observed in S100P. This small interhelical angle change is also observed in S100A6 and calbindin D_{9k} (Mäler et al., 2002). Loop II between helices III and IV adopts a wide range to coordinate with Ca^{2+}. Helix IV in S100P is six residues longer in the Ca^{2+}-bound state than that in the free sate. This suggests that the extra segment of helix IV is involved in target recognition. These changes in the orientation of helix III and loop II and the hinge region may participate in target protein binding.

Apo-S100P was found to exist as a homodimer in solution containing two symmetric apo-S100P subunits. The dimer interface of apo-S100P is maintained by packing helices I, I′, V and V′, as shown in Figure 2B, which is evident from numerous NOEs between helices I and I′, and V and V′ (Figure 3). Residues (A7, I11, V14, F71, I75 and A79) in these four helices (I, I′, IV and IV′) are involved in hydrophobic contacts (Figure 3). The interhelical angles of helices I and I′, and IV and IV′ are similar to those of apo-S100B and have no obvious change upon the addition of Ca^{2+}. Other hydrophobic contacts were also found between helix II′ to I (E40-L4). The charged residues (E40 and K39) in helix II, have been found to interact with residues in the N-terminus (E40-E5, E40-T2 and K39-T2). No ion-pair interactions were found in the dimer interface.

By comparison of the structures of the apo-S100P to other S100 proteins, the conformation of the N-terminal domain was not greatly changed in the absence and presence of the calcium ion. Few residues in the C-terminal domain play an important role in the conformation of apo-proteins. The structures of S100P and S100B were shown to be similar in the Ca^{2+}-binding state even though the most obvious differences in the apo-state were in the C-terminal domain. This indicates that the C-terminal domain may have similar functions for the S100P and S100B protein for recognition of the target proteins.

Acknowledgements

We wish to thank Drs Neuman Oezguen and Yuan Xu for assistance with the DIAMOD program. This research was supported by the NIH (AI27744 and ES06676), the Welch Foundation (H-1296), the Research Corporation Grant (RI0045), and the Sealy and Smith Foundation. Building funds for the UTMB NMR facility were provided by the NIH (1CO6CA59098).

References

Averboukh, L. et al. (1996) *Prostate*, **29**, 350–354.
Case, D.A. et al. (1999) *AMBER6*, University of California.
Cornilescu, G. et al. (1999) *J Biomol. NMR*, **4**, 289–302.
Donato, R. (2001) *Int. J. Biochem. Cell Biol.*, **33**, 637–668.
Drohat, A.C. et al. (1999) *Protein Sci.*, **8**, 800–809.
Duggan, B.M. et al. (2001) *J. Biomol. NMR*, **19**, 321–329.
Emoto, Y. et al. (1992) *Biochem. Biophys. Res. Commun.*, **182**, 1246–1253.
Gribenko, A.V. and Makhatadze, G.I. (1998) *J. Mol. Biol.*, **283**, 679–694.
Gribenko, A.V. et al. (2002) *Protein Sci.*, **11**, 1367–1375.
Gribenko, A.V. et al. (1998) *Protein Sci.*, **7**, 211–215.
Hanggi, G. and Braun, W. (1994) *FEBS Lett.*, **344**, 147–153.
Heizmann, C.W. and Cox, J.A. (1998) *Biometals*, **11**, 383–397.
Jurcevic, C. et al. (2003) *J. Pathol.*, **201**, 63–74.
Kay, L.E. et al. (1992) *J. Am. Chem. Soc.*, **114**, 10663–10665.
Mäler, L. et al. (2002) *J. Mol. Biol.*, **317**, 279–290.
Potts, B.C.M. et al. (1996) *Protein Sci.*, **5**, 2162–2174.
Rustandi, R.R. et al. (2002) *Biochemistry*, **41**, 788–796.
Smith, S.P. and Shaw, G.S. (1998) *Curr. Biol.*, **6**, 211–222.
Webber, D.J. et al. (2000) *Interaction of Dimeric S100B(ββ) with the Tumor Suppressor Protein: A Model for Ca-Dependent S100-Target Protein Interactions*, Kluwer Academic Publishers, Dordrecht.
Zhang, H. et al. (2003) *J. Mol. Biol.* **325**, 785–794.