The Dimerization Interface of the Metastasis-associated Protein S100A4 (Mts1)

IN VIVO AND IN VITRO STUDIES[S]

Svetlana Tarabykina, David J. Scott, Pawel Herzyk, Tessa J. Hill, Jeremy R. H. Tame, Marina Kriajevska, Daniel Lafitte, Peter J. Derrick, Guy G. Dodson, Norman J. Maitland, Eugene M. Lukyanidin, and Igor B. Bronstein

From the "Department of Molecular Cancer Biology, Danish Cancer Society, Copenhagen DK-2100, Denmark," Department of Biology, University of York, Heslington, "York Structural Biology Centre, Department of Chemistry, University of York, York," YCR Cancer Research Unit, Department of Biology, University of York, York YO10 5DD, United Kingdom, the "Institute of Mass Spectrometry, Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom," and "National Institute for Medical Research, Mill Hill, London NW7 1AA, United Kingdom"

The S100 calcium-binding proteins are implicated in signal transduction, motility, and cytoskeletal dynamics. The three-dimensional structure of several S100 proteins revealed that the proteins form non-covalent dimers. However, the mechanism of the S100 dimerization is still obscure. In this study we characterized the dimerization of S100A4 (also named Mts1) in vitro and in vivo. Analytical ultracentrifugation revealed that apoS100A4 was present in solution as a mixture of monomers and dimers in a rapidly reversible equilibrium (Kd = 4 ± 2 μM). The binding of calcium promoted dimerization. Replacement of Tyr-75 by Phe resulted in the major part of the dimer interface in homologous S100 proteins. By using the yeast two-hybrid system we showed that only a few residues of helix IV, namely Phe-72, Tyr-75, Phe-78, and Leu-79, are essential for dimerization in vivo. A homology model demonstrated that these residues form a hydrophobic cluster on helix IV. Their role is to stabilize the structure of individual subunits rather than provide specific interactions across the dimerization surface. Our mutation data showed that the specificity at the dimerization surface is not particularly stringent, which is consistent with recent data indicating that S100 proteins can form heterodimers.

The S100 family consists of 19 members, which function as transducers of the calcium signal in a tissue-specific manner. They are involved in mediating a wide range of biological processes, including cell division and differentiation, cytoskeleton organization, and cell motility (1, 2). It was suggested that individual S100 proteins might have specific functional roles in the cell types where they are expressed (3).

The S100 proteins utilize two EF-hand motifs to bind calcium ions. Upon binding of calcium these proteins undergo conformational changes resulting in exposure of hydrophobic regions on their surface, which allows interaction with target molecules (4, 5). It is generally believed that the S100 proteins act as noncovalent dimers (with the exception of calbindin D9k) inside the cell (3), whereas covalently linked dimers were shown to have extracellular functions (1, 6). Recently, the three-dimensional structure of a few S100 proteins was resolved by NMR spectroscopy (7–10) and x-ray crystallography (4, 11–15). These studies revealed a unique antiparallel homodimeric fold, which was not found in other calcium-binding proteins. The major part of the dimer interface is formed by helices IV and IV′ of the S100 subunits and is stabilized by contacts between helices I and I′.

S100A4, a member of the S100 family, is believed to be directly involved in the control of tumor metastasis (16, 17). However, the precise mechanism of its action is not known. It interacts with cytoskeletal proteins including tropomyosin (18) and the heavy chain of non-muscle myosin (19, 20). We have previously shown that S100A4 inhibits phosphorylation of myosin by protein kinase C (21) and by CK2 (22) and causes depolymerization of myosin filaments in vitro (22), thereby possibly modulating the motility of metastatic tumor cells. Recently, we have applied the yeast two-hybrid system as an in vitro approach to search for new S100A4 interacting partners. The screening of a mouse mammary adenocarcinoma cDNA library revealed homodimerization of S100A4 as well as its interaction with another member of the S100 family, S100A1 (23). Interestingly, mutational analysis revealed that Cys-76 and, perhaps, Cys-81 of S100A4 are important for its interaction with S100A1 but not for the S100A4 homodimerization, suggesting that the mechanisms of homo- and heterodimerization of the S100 proteins could be different. The present study
addressed the issue of the S100A4 dimerization, both in vitro and in vivo. The obtained data showed that in solution the S100A4 protein exists in a monomer-dimer equilibrium that is influenced by the binding of calcium.

It has been suggested that the S100 dimerization plays an essential role in specific target recognition. This hypothesis was supported by the three-dimensional structures of S100A10 (7) and S100A11 (8), which revealed that these proteins bind their target peptides via both the hinge and C-terminal regions of one subunit and the N-terminal region of another subunit. The NMR structure of the complex of S100B with a p53 fragment showed that whereas the general location of the binding site is similar to the one of S100A10 and S100A11, the N-terminal residues do not take part in target binding (24). To date, very little is known about how the S100 dimers form in living cells. Here, we used the yeast two-hybrid approach to study the structural basis of the S100 dimerization in vivo. Since the three-dimensional structure of S100A4 is not yet established, we have applied molecular modeling based on the available structures of the S100 proteins (4, 12) to interpret the results of mutational analysis. We have identified the amino acid residues crucial for the S100 dimerization in vivo and in vitro. These residues are highly conserved among the S100 proteins, and therefore our results may have implications for other members of the S100 family in general.

**EXPERIMENTAL PROCEDURES**

**Yeast Strain and Media**—The Saccharomyces cerevisiae strain YRG2 (Matα ura3-52 his3-200 ade2-101 lys2-801 trpl-901 leu2-3 112 gal80-538 lys2::UASgal1::TATAgal1::HIS3 URA3::UASgala1::TATAΛYCl1-lacZ) (Stratagene) was used for all assays. Yeast cultures were grown at 30 °C in either YPD medium (2% peptone, 1% yeast extract, 2% glucose) or SD minimal medium (0.67% nitrogen base, 18.2% D-sorbitol, 2% glucose) supplemented with amino acids.

**GAL4 Domain Fusion Constructs**—The open reading frame of mouse mts1 was cloned in-frame with either the GAL4 DNA binding domain (BD) (7) or the GAL4 transcriptional activation domain (AD) in the yeast expression vectors pBD-GAL4 and pAD-GAL4 (Stratagene), respectively. The p271 plasmid containing the mts1 CDNA (25) was digested with NcoI, flushed with the Klenow fragment of DNA polymerase I, and then digested with SallI. The pBD-GAL4 and pBD-GAL4 plasmids were cut with SallI, flushed with the Klenow polymerase, digested with SallI, and ligated with the mts1 insert. The resulting constructs named pADMs1 and pBDMts1, respectively, were verified by dried colony terminal sequencing (26). Expression of fusion proteins in yeast was confirmed by Western blotting with a polyclonal anti-S100A4 antibody.

**Yeast Transformation and Phenotype Analysis**—Transformation was performed using the YRG2 Yeast Competent Cell kit (Stratagene) according to the manufacturer’s protocol. Yeast transformed with both pBDMts1 and pADMs1 were plated onto SD agar plates lacking Trp and Leu and grown for 3–4 days at 30 °C. Single colonies were streaked onto either a selective SD agar plate lacking Trp, Leu, and His or on a nitrocellulose membrane (Hybond-C, Amersham Pharmacia Biotech) placed on an SD agar plate without Leu and Trp and incubated for 3 or 1 day, respectively, at 30 °C. The filter β-galactosidase assay was performed according to the manual of Stratagene (HybriZAP Two-hybrid Vector Kit). The time required for color development ranged from 1 h to overnight.

**S100A4 Mutagenesis**—All mutations were introduced by PCR using either the pBDMts1 or the pADMs1 plasmid as a template. The sequences of the primers used for the S100A4 mutagenesis are given in the Supplemental Material. PCR amplification was performed using a 10.1 AmpliTaq (PerkinElmer Life Sciences/Pyruvate/Stratagene) polymerase mixture. PCR products were eluted from an agarose gel, flushed with the Klenow fragment of DNA polymerase I, phosphorylated with T4 polynucleotide kinase, and self-ligated. The open reading frames of the plasmids were verified by sequencing.

**Protein Expression and Purification**—For expression of the His-tagged wild type and mutant S100A4 proteins in Escherichia coli M15, the T5 RNA polymerase promoter-based expression vector pQE30 (Qiagen) was used. Proteins were purified as described (23). The non-fusion S100A4 protein was expressed in Spin detergent (Peprotech, Rocky Hill, NJ) using the S100A4 CDNA was cloned into the pVL1393 vector. The plasmid was co-transfected into insect cells along with the BaculoGold-labeled viral DNA (Pharmingen). The obtained recombinant virus was amplified and used for protein expression. The insect cells were lysed in the buffer containing 25 mM HEPES (pH 7.5), 130 mM NaCl, 1 mM EGTA, and 0.5% Nonidet P-40. The supernatant was concentrated on an Amicon membrane (UP10) and subjected to gel filtration chromatography on a Superdex 75 column (Amersham Pharmacia Biotech) in 20 mM Tris-HCl (pH 7.5) and 100 mM NaCl. Protein fractions containing the S100A4 protein were further purified on a 1-ml Resource Q column (Amersham Pharmacia Biotech). The bound protein was eluted in Tris-HCl buffer with a NaCl gradient.

**Fourier Transform Ion Cyclotron Resonance Electrospray Ionization Mass Spectrometry**—Mass spectrometry measurements were made using an FTICR mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a shielded 9.4T super-conducting magnet (Magnet Scientific Ltd., Abingdon, UK), a cylindrical “infinity” ICR cell with diameter d = 0.06 m, and an external ESI source (Analytica, Branford, CT), as has been described previously (27, 28). The ESI source was equipped with a capillary made of Pyrex and coated on both ends with nickel. The background pressure in the ICR analyzer cell was usually below 2 × 10⁻⁵ mbar. Solutions of S100A4 (15 μM) in 5 mM ammonium acetate buffer (pH 6.5) were sprayed at room temperature. Carbon dioxide was used as a drying gas in the electrospray source.

**Analytical Ultrafiltration**—Sedimentation equilibrium experiments were carried out using a Beckman XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) with an AN60 Ti 4-hole rotor. Protein solutions in the presence or absence of 1 mM calcium were loaded into ultrafiltration cells containing Epon six channel centerpieces with 12-mm path lengths and quartz windows. The loading concentrations of the wild type S100A4 or mutant proteins were 11, 7, and 5 μM. Samples were centrifuged to equilibrium at 10,000, 15,000, 22,000, 28,000, and 36,000 rpm at 20 °C, a process that typically took around 10–12 h. Absorbance traces at 280 nm taken 4 h apart were compared to ensure that equilibrium had been attained at each speed, and this was ascertained by the lack of systematic deviation across each cell for the difference scans. Data were excised from the traces using the software supplied with the ultracentrifuge. The data were then globally fit either to a single species fit as shown in Equation 1,

\[
c_{\text{total}} = c_{\text{monomer}} \exp \left[ \frac{\omega^2}{2 F H} M \left( 1 - \frac{1}{ip} r^2 - r^2 \right) \right] \quad (\text{Eq. 1})
\]

or to a self-associating dimerization model shown in Equation 2,

\[
c_{\text{total}} = c_{\text{monomer}} \exp \left[ \frac{\omega^2}{2 F H} M \left( 1 - \frac{1}{ip} r^2 - r^2 \right) \right] + K c_{\text{monomer}} \exp \left[ \frac{\omega^2}{2 F H} M \left( 1 - \frac{1}{ip} r^2 - r^2 \right) \right] \quad (\text{Eq. 2})
\]

where K is the association constant; c is the concentration; r is the radial position; \( r_i \) is the radial reference position; M is the molar mass of the smallest detectable species; \( \rho \) is the partial specific volume; \( \omega \) is the angular velocity; R is the universal gas constant, and T is the thermodynamic temperature. Best fit parameters were obtained by non-linear least squares regression using the program NONLIN (29).

Sedimentation velocity experiments were carried out using Epon 2-channel centerpieces. The set speed was 40,000 rpm, and scans were obtained at 230 nm every 5 min. Loading concentrations were 22, 11, and 5 μM, respectively. The extinction coefficient at 230 nm was calculated from a wavelength scan at a fixed radial position at 3000 rpm from the ratio of absorbance at 280 and 230 nm.

The heterogeneity of the solution was analyzed using the method of van Holde and Weischt (50, 51). For single sedimenting species molar masses and sedimentation coefficients were obtained using a direct fit of the Lammm equation (Equation 3) using the program SedFit (32),

\[
\frac{d c}{d r} = \frac{1}{r} \frac{d c}{r} - \frac{1}{r^2} \frac{d c}{d r} - \frac{1}{r^2} \frac{d c}{d r} = \frac{1}{r^2} \frac{d c}{d r} (\text{Eq. 3})
\]

where \( c \) is the weight concentration; \( r \) is the radial position; \( t \) is the time...
Dimerization of S100A4 in Vivo and in Vitro

**Table 1.** Sequence alignment used for homology modeling of the S100A4 protein generated with the QUANTA software (MSI, San Diego, CA). The upper two sequences, representing both structural templates annotated by their Protein Data Bank and Swiss-Prot codes, were aligned based on the structural overlap. The top line indicates regions of similarity between all three sequences. The bottom line shows positions of secondary structure elements (gray bars correspond to a-helices) of the final S100A4 model.

| Sequence | Helices | Link Region | Cα²⁺ Binding Loop 1 | Cα²⁺ Binding Loop 2 |
|----------|---------|-------------|----------------------|---------------------|
| S100A4  | III     | III         | IV                   | IV                  |

**Figure 1.** Sequence alignment used for homology modeling of the S100A4 protein generated with the QUANTA software (MSI, San Diego, CA). The upper two sequences, representing both structural templates annotated by their Protein Data Bank and Swiss-Prot codes, were aligned based on the structural overlap. The top line indicates regions of similarity between all three sequences. The bottom line shows positions of secondary structure elements (gray bars correspond to a-helices) of the final S100A4 model.

from beginning of sedimentation; D is the diffusion coefficient, and s is the sedimentation coefficient. The molar mass can be obtained via the Svedberg equation (Equation 4),

\[ s = \frac{M(1 - \frac{A}{s})}{DRT} \]  

(Eq. 4)

This program was also used to fit the reaction boundary of a monomer/dimer-interacting system (32).

Sedimentation coefficient distributions were obtained using the maximum entropy method of Schuck (33). Frictional coefficients were obtained by setting \( f_f \) to an initial value of 1.0, to obtain the sedimentation distribution, and then letting it float in a subsequent fitting routine. Robustness of the distributions was ascertained by 1000 runs of a Monte Carlo distribution. From this, the 68% and 95% confidence limits were obtained.

The solvent density and partial specific volume of S100A4 was determined by the method of Laue et al. (34) using the program SEDNTERP and was found to be 1.01233 and 0.7248  

**RESULTS**

**Two-hybrid Analysis of the S100A4 Dimerization**—We have recently shown by yeast two-hybrid screening of a mouse adenocarcinoma cDNA library that S100A4 forms homodimers and heterodimers with S100A1, and we confirmed this result by ESI-FTICR mass spectrometry and analytical ultracentrifugation (23). In the present study we aimed to characterize structural aspects of the S100A4 dimerization both in vivo and in vitro. The S100A4 cDNA was cloned in-frame with either the GAL4 BD or AD. The fusion proteins were co-expressed in the YRG-2 yeast strain harboring integrated lacZ and HIS3 reporter genes under the control of the GAL4 upstream-activating sequence (UAS). Expression of the fusion proteins in yeast was confirmed by immunoblotting (not shown). Interaction between two hybrid proteins was scored by the ability of the transformed yeast to grow on a selective medium lacking histidine and by \( \beta \)-galactosidase activity.

Since the structure of the S100A4 protein is not known, we used the structures of highly homologous S100 proteins to identify amino acid residues important for the S100A4 dimerization in vivo. Our study mainly addressed S100A4 helix IV, which is known to form a large portion of the dimer interface of other S100 proteins (7–15). In all available structures the helix IV begins at the residue corresponding to Phe-72 of S100A4 and can be up to 21 residues long, particularly for the holoproteins (4). Four groups of the S100A4 mutant proteins fused to GAL4 BD and in some cases also with GAL4 AD were expressed in yeast: (i) long C-terminal deletions; (ii) single and double amino
acid deletions; (iii) simultaneous substitutions of several amino acid residues; and (iv) single amino acid substitutions. All the mutant proteins were tested for interaction with wild type S100A4 or with the identical mutant proteins.

The region of the S100A4 protein essential for dimerization was identified by deleting portions of the C terminus (Table I). Analysis of the HIS3 and lacZ phenotype of the transfected yeast revealed that only ΔC1 and ΔC6 proteins, which span deletions between residues 81 and 90, can form dimers efficiently. Thus, residues 72–80 contain all the helix IV residues essential for dimerization.

Point mutations were then introduced into positions between Phe-72 and Phe-78 and also in positions 80 and 81 (Table II). Substitutions at several positions (Q73A, Y75F, C76S, V77G, F78L, S80K, and C81K) were found not to have any effect on the S100A4 dimerization (Table II). Since Y75F, C76S, and F78L can be regarded as conservative mutations, the contribution of positions 75, 76, and 78 to dimerization cannot be ruled out by this experiment, whereas positions 73, 77, 80, and 81 seem to be not important for dimerization. Multiple sequence alignment of the S100 proteins (Fig. 2) showed that the positions 77, 80, and 81 were highly tolerant. The position 73, largely preserving polar character, shows relatively high sterochemical diversity as it can accommodate Ser, Val, Asn, Gln, Glu, Lys, and Arg in other S100 proteins. The position 80 is relatively unimportant, since the S80K mutant protein is able to dimerize. Consequently, the stretch of potentially important residues finishes at position 79. The non-conservative mutations F72A, Y75K, and F78A abolished dimer formation suggesting that these positions contribute significantly to the S100A4 dimerization.

Next, a number of single amino acid residues at positions 72 and 75–79 were deleted (Table II). Deletion of one residue in an α-helix rotates the downstream residues by 100° about the helix axis, hence it is equivalent to a number of point mutations, where subsequent residues along the helix are displaced by one residue. The most interesting ones are those between the deleted residue position 75 and position 79, as the latter is the last potentially important one. All the point deletions presented in Table II abolished dimerization. Analysis of these mutations showed that all of them have the common consequence, replacement of Leu-79 with Ser. This therefore indicated the importance of the residue Leu-79 for the S100A4 dimerization.

To mimic the monomeric protein calbindin D9k, which has a more hydrophilic surface of helix IV (39), Ser-80 and Cys-81 were substituted by two lysine residues. These mutations did not affect dimerization. To assess the plasticity of complementary surfaces in the dimer interface, we substituted the region 76–81 (CVFLSC) of S100A4 with VVLVAA (mutant protein Ch-A, Table II), the equivalent sequence from S100A1, which was recently shown to interact with S100A4 (23, 40). This replacement did not change the dimerization ability of the protein, showing that the dimer interface of S100A4 is functionally flexible.

In Vitro Biophysical Analysis of S100A4 in Solution—The structural features of the recombinant mouse S100A4 protein as well as the Y75F mutant protein were also analyzed in vitro. The latest was particularly interesting for us, because in our preliminary crystallization trials this protein was suitable for crystallization, whereas the wild type protein had a tendency to aggregate at various conditions and therefore was not adequate to obtain good crystals.

First, a non-destructive mass spectrometry technique was employed to investigate the extent of the S100A4 oligomerization in solution. Fig. 3A shows the mass spectrum of wild type
S100A4 in ammonium acetate buffer (pH 6.5) obtained by Fourier transform ion cyclotron resonance (FTICR) mass spectrometry coupled with electrospray ionization (ESI). The spectrum revealed the presence of mainly monomeric and dimeric species but also indicated that traces of higher order oligomers were present. Monomeric species with six to nine charges were detected indicating an array of charged solution states, whereas the average mass was determined as 13,117.82 Da. This, however, differed from the mass calculated from the sequence of the His6-tagged recombinant S100A4 (13,119.96 Da). As the mass accuracy was expected to be less than 10 ppm for a protein of this size, errors of more than 0.15 Da are unlikely to occur (41, 42). The discrepancy between the experimental and theoretical average masses possibly indicated the presence of one intramolecular disulfide bridge within wild type S100A4, which suggests the importance of the amino acid residue at the position 75 for the dimerization process. The differences in monomer charge between wild type S100A4 and Y75F may indicate that a monomer conformation of Y75F is more favorable for dimerization.

To ascertain the changes in oligomeric state upon calcium binding, sedimentation studies were carried out using the analytical ultracentrifuge. Sedimentation velocity data were analyzed according to the boundary analysis method of van Holde and Weischat (see Refs. 30 and 31). The results are summarized in Fig. 4. A diagonal spread of sedimentation coefficients in the calcium-free state was shown for the wild type S100A4 protein. This is indicative of a system in rapidly reversible equilibrium (31). Upon calcium binding there was a shift toward a single faster sedimenting species. This implies that oligomerization occurred upon calcium binding. In contrast, the mutant Y75F protein was present as a single species both with and without calcium. Direct fitting the Lamm equation according to the method of Schuck (32) showed that each of the single species had a molecular weight close to that of a dimer. The results are summarized in Table III.

Sedimentation equilibrium experiments were carried out in order to obtain the molecular weight distribution in solution in an absolute manner. Each protein was analyzed at three different loading concentrations and at three different speeds in the analytical ultracentrifuge. The results of the analysis are shown again in Table III. The data for wild type S100A4 fit a monomer/dimer rapidly reversible self-associating model. The dissociation constant was determined to be 4 (±2) μM. The Y75F protein was found to be dimeric both in the presence and in the absence of calcium, in agreement with both the sedimentation velocity results and the results of mass spectrometry.

One concern when studying metal-binding proteins is the influence of the hexahistidine tag used for protein purification. Therefore, a wild type S100A4 lacking the hexahistidine tag was expressed by using a baculovirus expression system and was purified from insect cells. The purified protein was then subjected to sedimentation studies. It was found that the distribution of sedimentation coefficients, as derived by the maximum entropy method of Schuck (33) for calcium-free protein, decreased with concentration again indicative of a self-associating system (Fig. 5A). The data from the two highest loading concentrations were globally fit to a monomer/dimer equilibrium model, and the fit yielded a dissociation constant of around 2 μM, in good agreement with the previous results (Table III). Fig. 5B shows the error profiles of the plots demonstrating that the determined dissociation constant represents the global minimum solution in the regression procedure. Upon calcium binding the protein appeared to sediment predominantly as a single species with a similar sedimentation coefficient to that of the His6-tagged S100A4 calcium bound protein, implying that the protein was predominately dimeric. There was, however, a small amount of oligomers sedimenting at around 5 S (Fig. 5C). It therefore appeared that the hexahistidine tag did not influence the self-association properties of the calcium-free protein. However, the tag had a small effect on the stability of the protein once it has bound calcium.

Homology Modeling—A homology model of the non-covalent homodimer of holoS100A4 was generated in order to rationalize the results of our mutagenesis experiments (see “Experimental Procedures”). As expected, the final S100A4 model was very similar to both the S100B and S100A10 templates. 100 and 98% of the residues were aligned with the 1MHO (S100B) and the 1A4P (S100A10) templates, respectively, and were within 3 Å r.m.s.d. on Ca atoms. Furthermore, the r.m.s.d. on the core residues were 0.8 and 1.0 Å, respectively.
model had no serious flaws detectable with the Protein Health module of the QUANTA software (MSI, San Diego, CA) and with PROCHECK (43).

The S100A4 model presented schematically in Fig. 6 demonstrates that the major part of the dimerization interface comes from the interdigititation of the "v"-shaped arrangements of helices I and IV in both subunits. The dimerization results in the total burial of 2830 Å² of the solvent-accessible surface, of which 80% is hydrophobic. The interface shown in Fig. 6, B, and C, joins two large hydrophobic cores present in both S100A4 subunits. The residues shown by the mutation experiments to be important for the S100A4 dimerization form a substantial part of this core.

Analysis of the S100A4 homology model as well as various S100 NMR and x-ray structures revealed the residues that contribute to the dimerization surface. These are Phe-72, Gln-73, Tyr-75, Cys-76, Leu-79, Ser-80, Ile-82, Ala-83, Met-84, Cys-86, Asn-87, Phe-90, and Glu-91. However, only four of them, namely Phe-72, Tyr-75, Phe-78, and Leu-79, were important for the S100A4 dimer formation in vivo as shown by our mutation experiments. Thus, the importance of a residue for the dimer assembly cannot be simply explained by its contribution to the dimerization surface.

In order to find out which residues of helix IV are important for the stability of the S100A4 subunits, we have calculated the intramolecular non-bonded energy per residue (data not shown). The results showed that the only significant contribution to this parameter comes from the N-terminal residues Tyr-75, Phe-72, and Phe-78. All three residues, identified by us as important for the S100A4 dimer formation, seem to stabilize strongly the arrangement of secondary structure elements and therefore contribute to the integrity of subunits. Their side chains form a very prominent ridge on the surface of helix IV. This ridge interacts with all other important parts of the sub-
unit, such as all remaining $\alpha$-helices, both calcium binding loops and the link region, as well as residues of the other subunit (Fig. 7A). The importance of the bulky character of this ridge is emphasized by the 96% conservation of Phe in position 72 (the remaining 4% are Tyr residues), 96% conservation of Phe/Tyr in the position 75, and full preservation of the big hydrophobic residue in position 78 (except for serine in two S100A3 sequences) as demonstrated by the multiple sequence alignment of 54 S100 proteins in Fig. 2.

The residues Phe-72, Tyr-75, and to lesser extend Leu-79, which were also proven to be important for dimer formation, interact with Phe-16, Val-13, and Ile-12 (these three residues make an interface between helices IV and I of the same subunit; Fig. 7). The specificity of this interaction is underlined by the conservation of residues 72 and 75 as described above, as well as 100% conservation of Phe/Tyr in position 16, and full preservation of the big hydrophobic residues in positions 12, 13, and 79 (except for position 13 in the non-dimerizing calbindin) as demonstrated in Fig. 2. However, the V13K S100A4 mutant protein was able to dimerize in the two-hybrid assay (data not shown). The residue Leu-79 makes a hydrophobic cluster with Ile-12, which constitutes a vertex point of the v-shaped dimerization surface (Fig. 7). This may play some role in the S100A4 self-association mechanism.

**DISCUSSION**

In the recent years the three-dimensional structures of S100B (4, 8, 10), S100A6 (7, 9), S100A7 (11), S100A8 (14), S100A10 (12), S100A11 (13), and S100A12 (15) have been elucidated. All of them revealed a homodimeric fold, which is
unique among calcium-binding proteins. The dimerization is
mediated by a number of contacts through helices designated
IV and IV of the S100 subunits. Additionally, the residues of
helices I and I and residues of the C-terminal extensions
contribute to the dimerization interface. However, it is still not
clear which of the identified contacts are crucial for the S100
dimerization in vivo. The yeast two-hybrid system was recently
applied as an in vivo assay to characterize structural aspects of
the S100A8/S100A9 (MRP8/MRP14) heterodimerization (44).
This study has shown that the heterodimer formed in vitro
upon calcium binding. Only a few residues in helix IV were
found to be essential for efficient dimerization in vivo. More-
over, at least one out of four residues previously identified by
NMR spectroscopy to be directly involved in dimerization was
not essential in vivo. Therefore, it seems to be reasonable to
combine NMR and x-ray crystallography studies with site-
directed mutagenesis and a functional in vivo test such as the
yeast two-hybrid system. In the present work we applied this
approach to identify stretches and individual amino acid resi-
dues, which are important for protein-protein interactions be-
tween the S100A4 subunits. Our results have demonstrated
that the residues 80–91 are not important for the self-assembly
of S100A4. Among these residues are Phe-89 and Phe-90,
which were previously shown to form additional contacts be-
tween dimers and, what is more important, to contribute to the
binding of target proteins in S100P (45) and S100A1 (46).

**TABLE III**

Analytical ultracentrifugation results

| Sample                  | Sedimentation coefficient | $M_r$        | $M_r^b$     | Dissociation constant |
|-------------------------|---------------------------|--------------|-------------|-----------------------|
| Y75F                    | 2.325 (2.308, 2.342)      | 24,486       | 27,831      | ND                    |
| Y75F + Ca$^{2+}$        | 2.426 (2.416, 2.434)      | 28,187       | 27,300      | ND                    |
| S100A4 (His$_6$) + Ca$^{2+}$ | 2.347 (2.338, 2.354)   | 28,397       | Fixed to 13,120 | 4 (±2) μM           |
| S100A4 + Ca$^{2+}$      | ND                        | ND           | Fixed to 11,721 | 2 μM                 |

$^a$ Molecular weight derived from direct fitting of the Lamm equation.

$^b$ Molecular weight derived from non-linear regression of the sedimentation equilibrium data. Three sets of loading concentrations, at three
different speeds, were used to derive the molecular weight. For interacting systems, the molecular weight was held constant, and the dissociation
constant was floated.

$^c$ Where applicable, a dissociation constant has been derived.

$^d$ Values were derived from the maximum entropy distributions.

**FIG. 5.** Sedimentation coefficient distributions for wild type S100A4
without a hexahistidine tag, ob-
tained by using the maximum en-
tropy method of Schuck (33). The
traces represent three loading concentra-
tions at relative ratios of 1:0.67:0.33. The
robustness of each derived distribution
was tested with 1000 runs of a Monte
Carlo distribution. The lines plotted are
as follows: black (mean distribution), yel-
low and green (68% confidence limit), blue
and red (95% confidence limit). A, in the
absence of calcium the protein sediments
slower upon dilution. This is indicative of
a self-associating system. Fitting the re-
action boundaries to a monomer/dimer
equilibrium yields a dissociation constant
of 2 μM, in agreement with the value of the
wild type His$_6$-tagged S100A4 pro-
tein. B, plot of root mean square error
versus dissociation constant for the mon-
omer/dimer equilibrium fit of the sedi-
mentation velocity data. As shown, the
determined $K_d$ lies in a global minimum.
C, in the presence of 1 mM CaCl$_2$ the pro-
tein forms species centered around 2.5–
2.6 S. This is close to that seen for the
dimer in previous experiments (see Fig.
4). A small amount of aggregate is pre-
ent, sedimenting at around 5 S.
and suggested that a similar situation occurred in other S100 proteins as well. Our in vivo study confirmed this suggestion and, moreover, revealed key residues responsible for the S100A4 dimerization. In contrast, the yeast two-hybrid experiments on the MRP8/MRP14 heterodimerization (44) demonstrated the importance of Phe-89 and Leu-95 for both homo- and heterodimer association. These residues correspond to Met-84 and Phe-90 in S100A4, and their deletion did not affect dimerization in our assays. To explain this contradiction one should take into account the unusually long C-terminal extensions of MRP8 and MRP14. This extension might create another interaction interface between dimers, and therefore the role Phe-89 and Leu-95 could be to provide a second putative association pathway by forming tetrameric or higher order intermediates. The presence of a small amount of aggregate sedimenting at 5 s in the presence of calcium, which was found in our ultracentrifugation experiments, points toward this. Additionally, small amounts of a trimer and tetramer were observed using FTICR. However, this may be due to the absence of salt needed for this analysis. Indeed, molecular weight av-

![Graphical presentation of the final model of the S100A4 homodimer.](image1)

**FIG. 6.** Graphical presentation of the final model of the S100A4 homodimer. A, model in ribbon presentation with two subunits coded with different colors and calcium ions in the space-filling representation, generated using QUANTA (MSI, San Diego, CA); B, hydrophobic core of S100A4; the side chains of hydrophobic residues are presented in space-filling representation in gray, and residues important for the dimer formation, Phe-72, Tyr-75, Phe-78, and Leu-79, are in yellow. Main chains are presented as tubes, color-coded according to the polarity of residues as follows: hydrophobic, gray; polar, red; and charged, white. The image is generated using the MOLVIEWER program (Dr. M. J. Hartshorn, personal communication). C, wall-eyed stereo view of the dimerization interface. Helices I and IV are in a ribbon representation, and molecular surfaces corresponding to the dimerization interface are shown with dots. The two subunits are color-coded with blue and red. Side chains of Phe-72, Tyr-75, Phe-78, and Leu-79 are presented as liquorice strands in yellow. The dotted surface becomes yellow whenever it corresponds to one of the above residues. The image is generated with QUANTA (MSI, San Diego, CA).

![Structural role of the residues critical for dimer formation, Phe-72, Tyr-75, Phe-78 and Leu-79, in stabilization of the core of the S100A4 monomeric subunit.](image2)

**FIG. 7.** Structural role of the residues critical for dimer formation, Phe-72, Tyr-75, Phe-78 and Leu-79, in stabilization of the core of the S100A4 monomeric subunit. Color coding is as follows: helices I to IV are in blue, green, magenta, and red, respectively, and loops are in white. A, stereo view of the S100A4 monomeric subunit “as seen” from the other monomer, and the side chains of the above mentioned four residues and the ones interacting with them are shown as liquorice strands. The larger labels correspond to the residues of helix IV and I, and calcium ions are in space-filling representation. B, stereo view of the helix I to helix IV interface in a monomeric subunit of the S100 proteins, represented by structural overlap of 13 known structures with the following Protein Data Bank codes: 3ICB, 1UVO, 1SYM, 1B4C, 1QLK, 1CFP, 1MHO, 1CNP, 2CNP, 1A03, 1A4P, 1BT6, and 1PSR. Side chains of interacting residues are shown as liquorice strands.
The analysis of the S100A4 dimerization performed in this work demonstrated that the residues of helix IV, which are important for the dimer formation, contribute substantially to the stabilization of the monomeric units, in particular the fork-shaped arrangement of helices I and IV. The dimerization is achieved by non-covalent interactions between hydrophobic residues lining the internal surface of this arrangement in two subunits (Fig. 7). The yeast two-hybrid assay identified the key residues from helix IV as Phe-72, Tyr-75, Phe-78, and Leu-79. The structural model reinforced with the multiple sequence alignment of the S100 proteins identified complementary residues from helix I, namely Ile-12, Val-13, and Phe-16, which are believed to be important for the S100 dimer formation (39). All these residues, except for Phe-78, make a hydrophobic cluster at the interface between helices I and II. The residues Phe-78 and Tyr-75 link this cluster with the hydrophobic, conserved residues from helix II (Leu-34, Leu-37, and Leu-38) and helix III (Leu-62), both calcium binding loops (Leu-29 and Val-70), and the link region (Leu-42 and Leu-46), as can be seen from Fig. 7A. The vertex point of the dimerization surface made by helices I and IV is built by Leu-79 and Ile-12, which may contribute to the dimerization mechanism.

Calbindin D$_{9k}$ is the only S100 protein whose structure is known in the monomeric state. Although calbindin does not form dimers, it has strong sequence and structural homologies to other S100 proteins. Sequence alignment and comparison of the bovine holo-calbindin D$_{9k}$ crystal structure (Swiss-Prot code S10D_BOVIN; Protein Data Bank code 3ICB (48)) with the S100A4 model revealed sequence identity of 39% upon the structural overlap; 88% of the overlapped residues showed the main chain r.m.s.d. below 3 Å. The best structural overlap between these two structures consists of the following structural elements of S100A4: C-terminal halves of helix I and III, both calcium binding loops, helix II, and the N-terminal half of helix IV. Thus, the observations and conclusions concerning the role of some amino acid residues in the monomer stability of S100A4 can be compared with the effect of point mutations on the stability and calcium binding of calbindin D$_{9k}$ (49, 50). These studies convincingly demonstrated that replacement of the bulky residues equivalent to S100A4 Tyr-75, Phe-16, Val-70, Leu-29, and Leu-34 with small alanines or glycines significantly decreased the stability toward urea unfolding, largely reduced the Ca$^{2+}$ affinity, and substantially increased the Ca$^{2+}$ dissociation rate. Analysis of the S100A4 model showed that all these residues are clustered together and are positioned in the direct vicinity to the bulky ridge made by the side chains of the residues Phe-72, Tyr-75, and Phe-78 proven to be important for the S100A4 dimer formation (Fig. 7A).

Our conclusions concerning the structural role of the helix IV residues important for the S100A4 dimerization are consistent with the existing structures of the S100 proteins. Fig. 7B presents overlapped monomeric subunits from 13 S100 structures available in the Protein Data Bank data base and highlights well converged side chains of the helix IV key residues and complementary residues from helix I. The residues equivalent to Phe-78 in S100A4 are less converged mainly due to the different position of neighboring helix III in the apo- and holo-conformations. The best structural alignment of the protein backbone is achieved for the C-terminal half of helix I comprising residues 10–18, a short β-sheet between two calcium binding loops (residues 28–29 and 69–71), and the N-terminal half of helix IV (residues 72–81). Thus, the most invariant core of an S100 subunit contains all the helix IV key residues, a substantial part of the dimerization interface corresponding to the area close to the vertex between helices I and IV, as well as residues stabilizing the arrangement of helices I and IV.

The mutagenesis results presented in this work revealed that the part of the S100A4 dimerization surface corresponding to helix IV has very little sequence specificity. It seems that as long as the integrity of a monomeric subunit and a proper arrangement of helices I and IV are achieved, the protein forms a dimer in order to bury the vast hydrophobic area between these helices and to create a large hydrophobic core inside a dimer. The structural basis of these interactions suggests also a possibility for the heterodimerization between the S100 family members.

The in vitro biophysical analysis of the Y75F mutant protein revealed that this protein was much more stable in the dimeric form. Analysis of the S100A4 model demonstrated that the hydroxyl group of Tyr-75 of the S100A4 subunit lies in the hydrophobic cavity, but in contrast to the dimeric form it is exposed to the solvent so that potential hydrogen bonding to the bulk solvent is possible. The lack of this group in the monomeric subunit of Y75F results in the exposure of a fully hydrophobic cavity to the water. This situation is energetically disadvantageous, and therefore it favors dimerization over free monomers. It should be noted that although the monomeric calbindin has phenylalanine in this position (equivalent to Tyr-75), it blocks the entrance to the cavity along with the hydrophobic side chains of Leu-6, Phe-36, and Ile-73, corresponding to Ile-12, Leu-42, and Ile-82 of S100A4, respectively. This hydrophobic patch is then covered by a layer of charged residues (Lys-1 and Glu-35) directly exposed to the solvent. Such an arrangement of the residues Ile-73, Lys-1, and (to some extent) Leu-6 is possible only because helices I and IV are significantly shorter in calbindin, which permits backbone distortion in this region. The fully hydrophobic environment of Tyr-75 in the dimeric form is achieved by the cavity containing the tyrosine hydroxyl group being filled with the side chain of Leu-5' from the other subunit (Fig. 8). The hydrophobic character of this residue is preserved throughout the S100 proteins except for calbindin and S100A11. In summary, the structural homology analysis of the S100A4 model has provided the ex-
plation why the Y75F mutant protein in the dimeric form could be more stable than wild type S100A4, whereas the opposite would occur in the monomeric form. This would explain the increased dimerization ability of Y75F seen in both the FTICR-ESI mass spectrometry and the analytical ultracentrifugation studies.

Although the functions of S100A4 inside and outside the cell are not fully understood, it seems that local calcium concentration may determine the localization of the S100 proteins (51) as well as their reactivity and specificity for cellular targets (46). Our data show that S100A4 undergoes calcium-dependent dimerization. Mass spectrometry revealed the presence of multiple monomer and dimer charge states, which can point toward multiple conformers. This may provide the structural basis for the diversity of the S100A4 functions. We have recently shown that the oligomeric but not the dimeric forms of wild type S100A4 stimulated neurite outgrowth in hypcampal neurons, whereas the dimer Y75F mutant protein did not (52). However, Y75F was able to interact with the intracellular S100A4 target, the heavy chain of non-muscle myosin IIA, as strongly as the wild type S100A4 protein. The diversity of the S100A4 functions therefore appears to reflect the innate structural plasticity of the protein.

Acknowledgments—We thank Dr. Walter J. Chazin for kindly providing the coordinates of the calbindin 1A03 and 2CNP structures (52). However, Y75F was able to interact with the intracellular S100A4 target, the heavy chain of non-muscle myosin IIA, as strongly as the wild type S100A4 protein. The diversity of the S100A4 functions therefore appears to reflect the innate structural plasticity of the protein.

REFERENCES

1. Zimmer, D. B., Cornwall, E. H., Landar, A., and Song, W. (1995) Brain Res. Bull. 37, 417–429
2. Schafer, B. W., and Heizmann, C. W. (1996) Trends Biochem. Sci. 21, 134–140
3. Donato, R. (1999) Biochim. Biophys. Acta 1450, 191–231
4. Matsumura, H., Shiba, T., Inoue, T., Harada, S., and Kai, Y. (1998) Structure 6, 233–241
5. Heizmann, C. W., and Cox, J. A. (1998) Biometals 11, 383–387
6. Winningham-Majer, F., Staecker, J. L., Barger, S. W., Coats, S., and Van Eldik, L. J. (1989) J. Cell Biol. 109, 3036–3070
7. Potts, B. C., Smith, J., Akke, M., Macke, T. J., Okazaki, K., Hidaka, H., Case, D. A., and Chazin, W. J. (1995) Nat. Struct. Biol. 2, 790–796
8. Gilby, P. M., Van Eldik, L. J., and Roberts, G. C. (1996) Structure 4, 1041–1052
9. Maler, L., Potts, B. C. M., and Chazin, W. J. (1999) J. Biomol. NMR 15, 233–247
10. Smith, S. P., and Shaw, G. W. (1998) Structure 6, 211–222
11. Brodersen, D. E., Etzerodt, M., Madsen, P., Celis, J. E., Thogersen, H. C., Nyborg, J., and Kjeldgaard, M. (1998) Structure 6, 477–489
12. Réty, S., Sopkova, J., Renouard, M., Osterloh, D., Gerke, V., Tabarias, S., Russo-Marie, F., and Lewit-Bentley, A. (1999) Nat. Struct. Biol. 6, 89–95
13. Réty, S., Osterloh, D., Arce, J. P., Tabarias, S., Seeman, J., Russo-Marie, F., Gerke, V., and Lewit-Bentley, A. (2000) Structure 8, 175–184
14. Ishikawa, K., Nakagawa, A., Tanaka, I., Suzuki, M., and Nishihira, J. (2000) Acta Crystallogr. Sect. D. Biol. Crystallogr. 56, 559–566
15. Moroz, O., Antson, A., Murshudov, G., Maitland, N., Dodson, G., Wilson, K., Skibbe, J., Lukanidin, E., and Bronstein, L. (2001) Acta Crystallogr. Sect. D. Biol. Crystallogr. 57, 20–29

3 Tarabykina, unpublished data.