Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The three-dimensional structure of human S100A12

Olga V. Moroz,[a] Alfred A. Antson,[a] Garib N. Murshudov,[a] Norman J. Maitland,[b] G. Guy Dodson,[a,c] Keith S. Wilson,[a] Inge Skibshøj,[d] Eugene M. Lukanidin[d] and Igor B. Bronstein[a]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, England, [b]Yorkshire Cancer Research Unit, Department of Biology, University of York, York YO10 5DD, England, [c]National Institute of Medical Research, Mill Hill, London NW7 1AA, England, and [d]Department of Molecular Cancer Biology, Danish Cancer Society, Strandboulevarden 49, Copenhagen 2100, Denmark

Correspondence e-mail: igor@ysbl.york.ac.uk

Received 13 July 2000
Accepted 17 October 2000

PDB Reference: S100A12, 1e8a.

The crystal structure of human EF-hand calcium-binding protein S100A12 in its calcium-bound form has been determined to 1.95 Å resolution by molecular replacement using the structure of the S100B protein. The S100 family members are homologous to calmodulin and other related EF-hand calcium-binding proteins. Like the majority of S100 proteins, S100A12 is a dimer, with the interface between the two subunits being composed mostly of hydrophobic residues. The fold of S100A12 is similar to the other known crystal and solution structures of S100 proteins, except for the linker region between the two EF-hand motifs. Sequence and structure comparison between members of the S100 family suggests that the target-binding region in S100A12 is formed by the linker region and C-terminal residues of one subunit and the N-terminal residues of another subunit of the dimer. The N-terminal region of the target-binding site includes two glutamates that are conserved in most of the S100 sequences. The comparison also provided a better understanding of the role of the residues important for intra- and inter-subunit hydrophobic interactions. The precise role of S100A12 in cell behaviour is yet undefined, as is the case for the whole family, although it has been shown that the interaction of S100A12 with the RAGE receptor is implicated in inflammatory response.

1. Introduction

S100 proteins are small EF-hand calcium-binding proteins with a molecular weight in the range 10–14 kDa. Almost all S100 proteins tend to form dimers, considered to be essential for their target binding and function (Krebs *et al.*, 1995; Donato, 1999; Réty *et al.*, 2000). The exceptions found to date are the monomeric calbindin D_{9k}, which was suggested to have a calcium-buffering role (Szebenyj & Moffat, 1986), and p26olf, which is a single-chain molecule consisting of two almost identical S100-like domains (Miwa *et al.*, 1998; Tanaka *et al.*, 1999). In addition, two high molecular weight proteins, the intermediate filament-associated protein precursors profilaggrin (Markova *et al.*, 1993) and trichohyalin (Lee *et al.*, 1993), contain S100-like domains. There are also exceptions in terms of calcium binding: for example, S100A10 (Réty *et al.*, 1999) and the N-terminal loop of S100A7 (Brodersen *et al.*, 1998) do not bind calcium, while S100A3 was reported to have very low affinity for calcium (Fohr *et al.*, 1995). A number of the S100 proteins have been shown to bind zinc (reviewed in Brodersen *et al.*, 1999).

The name S100 was introduced because several proteins of this group were soluble in 100% ammonium sulfate (Moore, 1965). There is much confusion in the literature as some

members of the family have been assigned more then one name. Recently, a new nomenclature was introduced (Schäfer *et al.*, 1995) based on the arrangement of most of the known S100 proteins on human chromosome 1 (lq21). However, even with this more convenient nomenclature complete order has not been reached, since new proteins are being found between those already numbered. Such was the case with S100A12, located between S100A8 and S100A9 (Wicki *et al.*, 1996). S100A8, S100A9 and S100A12 all belong to the calgranulin subfamily (Kerkhoff *et al.*, 1998).

S100 proteins can in some cases form heterodimers. For instance, S100B from the brain exists in dimeric (80%) and heterodimeric (20% S100B + S100A1) forms (Isobe *et al.*, 1981). S100B has also been shown to form heterodimers with S100A6 (Yang *et al.*, 1999) and S100A1 with S100A4 (Wang *et al.*, 2000; Tarabykina *et al.*, 2000), while S100A8 forms heterodimers with S100A9 (Edgeworth *et al.*, 1991). In contrast, S100A12 does not form heterocomplexes with the two other calgranulins S100A8 and S100A9 (Vogl *et al.*, 1999).

There are no crystal structures of S100 proteins in the calcium-free state. NMR spectroscopy and Fourier transform ion cyclotron resonance electrospray ionization mass spectrometry have demonstrated that the Ca-free S100 proteins exist in solution in a number of different conformations (Potts *et al.*, 1995; Skelton *et al.*, 1995; Drohat *et al.*, 1996; Tarabykina *et al.*, 2000). This indicates that the proteins are very flexible without calcium. The only exception is S100A10, which cannot bind Ca^{2+} and is structurally similar to the 'calcium-loaded' conformation. However, the number of crystal structures of calcium-bound S100 proteins is growing rapidly. There are currently five S100 protein structures in the Protein Data Bank (Bernstein *et al.*, 1977): bovine calbindin D9K (Svensson *et al.*, 1992), bovine S100B (Matsumura *et al.*, 1998), human S100A7 (Brodersen *et al.*, 1999), human S100A10 (Réty *et al.*, 1999) and human S100A11 (Réty *et al.*, 2000). Recently, a paper on the structure of human S100A8 was published (Ishikawa *et al.*, 2000). The sequence identity between these proteins varies between 35 and 45%.

S100 proteins have two different calcium-binding EF-hand motifs. The N-terminal motif (EF-1) is composed of two α-helices (H_I and H_{II}) linked by a calcium-binding loop L_I; the second, C-terminal, motif (EF-2) is made up of helices H_{III} and H_{IV} linked by the calcium-binding loop L_{III}; the two EF motifs are connected by the linker loop L_{II}.

The N-terminal calcium-binding motif (EF-1) is a feature specific to S100 proteins. EF-1 has 31 residues in the consensus sequence (14 residues in the calcium-binding loop) and binds Ca^{2+} mainly through main-chain carbonyl O atoms (Nakayama *et al.*, 1992). EF-1 has low affinity for calcium compared with calmodulin and troponin C and indeed compared with S100 EF-2 (Kligman & Hilt, 1988).

In contrast, the C-terminal 'canonical' calcium-binding motif (EF-2) is closely similar to the EF hands in calmodulin (Babu *et al.*, 1985), troponin C (Herzberg & James, 1985) and parvalbumin (Moews & Kretsinger, 1975). Like them, it has a high affinity for Ca^{2+}. EF-2 is a 12-residue calcium-binding loop flanked by α-helices; its consensus sequence contains 29 residues (Moncrief *et al.*, 1990). In EF-2, calcium is liganded by five side-chain O atoms, one main-chain carbonyl O atom and a water molecule.

Although many functions have been proposed for the S100 proteins, their actual biological roles remain elusive. Numerous lines of evidence suggest that they participate in cell growth and differentiation, are involved in the assembly/disassembly of cytoskeletal proteins and modulate activity of various protein kinases (reviewed in Kligman & Hilt, 1988; Schäfer & Heizmann, 1996; Heizmann & Cox, 1998; Donato, 1999). A number of S100 targets have been identified. For example, S100B binds to p53 (Baudier *et al.*, 1992) tau-proteins (Baudier & Cole, 1988) and CapZ (Ivanenkov *et al.*, 1995), S100A10 binds to annexin II (Réty *et al.*, 1999) and S100A11 binds to annexin I (Réty *et al.*, 2000). These data suggest that S100 proteins may function in different ways, forming transient complexes with intracellular proteins as well as with extracellular receptors (Komada *et al.*, 1996; Hoffmann *et al.*, 1999). The interplay between S100 proteins and their targets may serve to modulate responses to calcium signals that control cell differentiation, motility and proliferation.

Target binding by S100 proteins has been proposed to be analogous to that of calmodulin (Zhang *et al.*, 1995); *i.e.* binding of calcium leads to conformational changes in the protein that expose a mostly hydrophobic target-recognition surface (Chazin, 1995; Schäfer & Heizmann, 1996). Recent progress in the NMR structure determination of the S100 proteins is providing more and more evidence in favour of this hypothesis, although the nature of the conformational changes and the location of the target-binding surface are quite different from those of calmodulin. In spite of the fact that EF-1 appears to have a lower affinity for calcium, the latest model of calcium-induced structural changes in S100 proteins is that Ca^{2+} first binds to the EF-1 loop, which in terms of conformation is close to the calcium-loaded state from the beginning. The EF-2 loop in calcium-free S100-proteins is much more disordered. It was also proposed that upon binding of Ca^{2+} to EF-1, site–site communication causes significant reorganization of EF-2, making it also ready to bind calcium (Mäler *et al.*, 2000). When the calcium ion is bound, EF-2 loses its flexibility and as a result helix III is reoriented and fixed in the conformation that exposes the target-binding site, which is mostly hydrophobic but also has negative charges in several regions (Smith & Shaw, 1998; Drohat *et al.*, 1999; Mäler *et al.*, 1999).

Human S100A12 was first characterized by Guignard *et al.* (1995). Human S100A12 has homologues in pig granulocytes (Dell'Angelica *et al.*, 1994; Nonato *et al.*, 1997) and bovine amniotic fluid (Hitomi *et al.*, 1996), with which it shares a sequence identity of 70 and 78%, respectively. S100A12 together with S100A8 and S100A9 belong to the calgranulin subfamily; *i.e.* they are expressed exclusively in granulocytes and are relatively abundant in this type of leukocyte (Kerkhoff *et al.*, 1998). Recent data indicate that S100A12 and two other calgranulins play different roles in granulocyte function (Vogl *et al.*, 1999).

Table 1
Statistics for the X-ray data.

Resolution (Å)	No. of unique reflections	$I/\sigma(I)$	$I/\sigma(I) > 3$ (%)	Completeness (%)	$R_{merge}(I)$ (%)
25.00–4.20	1726	16.4	97.9	99.4	5.7
4.20–3.33	1743	13.2	96.9	99.7	6.5
3.33–2.91	1729	8.7	91.7	99.9	11.0
2.91–2.65	1724	7.23	84.4	99.9	18.1
2.65–2.46	1726	7.0	77.6	99.6	23.8
2.46–2.31	1728	7.58	73.9	98.8	10.8
2.31–2.20	1696	6.3	64.6	98.4	13.1
2.20–2.10	1750	5.24	58.8	99.3	13.7
2.10–2.02	1705	3.83	48.7	98.6	20.4
2.02–1.95	1721	2.68	35.7	99.1	25.8
All *hkl*	17248	10.6	73.0	99.3	8.6

† $R_{merge} = \sum|I - \langle I\rangle|/\sum I$, where I is the intensity of the reflection.

Table 2
Parameters of translation–libration–screw (TLS) after refinement.

B values in the crystal are calculated using $B_{crystal} = B_{ind} + \mathbf{T} + (\mathbf{r} - \mathbf{r}_0)^T\mathbf{L}(\mathbf{r} - \mathbf{r}_0) - (\mathbf{r} - \mathbf{r}_0)\mathbf{S} + [(\mathbf{r} - \mathbf{r}_0)\mathbf{S}]^T$, where B_{ind} is the value from the output coordinate file, $\mathbf{T}$ (Å^2) is the translation tensor, $\mathbf{L}$ ($^\circ{}^2$) is the libration tensor, $\mathbf{S}$ (Å$^\circ$) is the screw tensor or correlation between translation and libration, $\mathbf{r}$ (Å) is the position of the atom in the unit cell and $\mathbf{r}_0$ (Å) is the origin of the rigid body. In this case, the origin of rigid body was the centre of mass of the unit cell. Using TLS parameters one can make inference about relative stability of the rigid groups. Higher $\mathbf{L}$ values usually correspond to the less ordered rigid groups. In the case of S100A12 it can be seen that chain B is less ordered than the chain A, but the difference between these values is so small that this effect cannot be observed in the electron density.

	Chain *A*	Chain *B*
$\langle\mathbf{T}_{11}\rangle$	0.026	0.10
$\langle\mathbf{T}_{22}\rangle$	0.036	0.046
$\langle\mathbf{T}_{33}\rangle$	0.029	0.036
$\langle\mathbf{T}_{12}\rangle$	−0.019	−0.040
$\langle\mathbf{T}_{13}\rangle$	0.018	0.030
$\langle\mathbf{T}_{23}\rangle$	−0.017	−0.040
$\langle\mathbf{L}_{11}\rangle$	2.091	1.941
$\langle\mathbf{L}_{22}\rangle$	1.665	2.419
$\langle\mathbf{L}_{33}\rangle$	2.333	5.288
$\langle\mathbf{L}_{12}\rangle$	0.166	−0.241
$\langle\mathbf{L}_{13}\rangle$	−0.149	0.070
$\langle\mathbf{L}_{23}\rangle$	−1.211	−2.055
$\langle\mathbf{S}_{22} - \mathbf{S}_{11}\rangle$	−0.117	0.037
$\langle\mathbf{S}_{11} - \mathbf{S}_{33}\rangle$	0.084	−0.081
$\langle\mathbf{S}_{12}\rangle$	−0.157	0.201
$\langle\mathbf{S}_{13}\rangle$	0.174	−0.259
$\langle\mathbf{S}_{23}\rangle$	0.075	0.046
$\langle\mathbf{S}_{21}\rangle$	0.116	−0.342
$\langle\mathbf{S}_{31}\rangle$	−0.144	0.68
$\langle\mathbf{S}_{32}\rangle$	−0.179	−0.355

New insights into the possible biological significance and role of S100A12 were obtained by studying the activation of the receptor for advanced glycation end products (RAGE), for which involvement in tumour growth and spread has recently been reported (Taguchi *et al.*, 2000). It appears that S100A12 is necessary for the proper functioning of granulocytes; strong evidence for this is its abundance in cells and its very specific response for calcium signalling when the protein is moving from cytosol to cytoskeleton and membranes (Vogl *et al.*, 1999). It has also been reported that S100A12 binds to anti-allergic drugs (Shishibori *et al.*, 1999), indicating that it is involved in highly diverse events and signalling pathways. One such pathway is the activation of NF-κB complexes, which results in expression of multiple gene products contributing to the inflammatory response (Hoffmann *et al.*, 1999).

We report here the three-dimensional crystal structure of S100A12 determined with data extending to 1.95 Å resolution. To begin to identify the regions of structural and functional importance in S100A12, we compare it with the five previously known S100 X-ray structures listed above. The comparison identifies a putative target-binding site in S100A12 and suggests the structural basis of its function.

2. Methods

2.1. Data collection

S100A12 protein was isolated from human granulocytes. Crystallization and preliminary X-ray analysis have been reported previously (Moroz *et al.*, 2000). In short, the crystals were grown by hanging-drop vapour diffusion with a protein concentration of 5.0–8.0 mg ml^{-1}. The reservoir contained 0.1 *M* sodium cacodylate pH 6.5, 0.2 *M* $CaCl_2$ and 20–25% PEG 5K monomethyl ester (Brzozowski, 1993). Crystals belong to the space group *R*3, with unit-cell dimensions $a = b = 99.6$, $c = 64.2$ Å. The initial data were collected in-house to 2.5 Å resolution. The same crystal was used to extend the resolution to 1.95 Å using synchrotron radiation at the ESRF beamline 14-4 (wavelength 0.93 Å). Several reflections in the synchrotron data had intensities above the dynamic range of the detector; therefore, they were merged with the data collected in-house. The data were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). 48 386 individual reflections were reduced to 17 248 unique observations with an overall R_{merge} of 8.6% (Table 1). The data are 99.3% complete.

2.2. Structure determination and refinement

All crystallographic calculations were performed using the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994). The structure was solved by molecular replacement (MR) with *MOLREP* (Vagin & Teplyakov, 1997) using the structure of S100B with a sequence identity of 38% (Matsumura *et al.*, 1998; PDB code 1mho) as a starting model. *MOLREP* showed clearly that there were two molecules in the asymmetric unit.

The resulting model was used for maximum-likelihood refinement as implemented in the program *REFMAC* (Murshudov *et al.*, 1997). In the early stages, rigid-body refinement with each molecule as a single rigid group was employed. After convergence of rigid-body refinement, individual atomic refinement was performed. In the first cycles, non-crystallographic symmetry (NCS) averaging using *DM* (Cowtan & Main, 1993) was used to improve the phases. The resulting phases were used for phased refinement (Pannu *et al.*, 1998) with suitable blurring factors to optimize the fall of *R*

Table 3
Model refinement statistics.

Resolution limits (Å)	25–1.95
Total No. of reflections	16326
Percentage observed	99.98
Percentage of free reflections	5
R_{cryst}† (%)	17.8
Free R factor (%)	21.6
R.m.s. deviations from ideal geometry (target values are given in parentheses)	
Bond distances (Å)	0.018 (0.021)
Bond angles (°)	1.714 (1.931)
Chiral centres (Å^3)	0.15 (0.20)
Planar groups (Å)	0.007 (0.020)
Main-chain bond B values (Å^2)	0.42 (1.50)
Main-chain angle B values (Å^2)	0.79 (2.00)
Side-chain bond B values (Å^2)	1.48 (3.00)
Side-chain angle B values (Å^2)	2.49 (4.50)

† $R_{merge} = \sum \big||F_o| - |F_c|\big| / \sum |F_o|$.

and R_{free} values. To check if the refinement was proceeding well, omit maps were calculated by systematically removing part of the structure, refining the remaining model and checking if the resulting electron density confirmed the positions of the omitted atoms. After the check was finished, all atoms were included and refinement with phases and updating of the phases with the NCS was iterated. In the intermediate stages, manual rebuilding was performed using the *XFIT* option (Oldfield, 1994) of the program *QUANTA* (Molecular Simulations).

For further refinement, a bulk-solvent correction based on contrast value in the region not occupied by the described atoms as implemented in the program *REFMAC* was used and improved R and R_{free} values by around 2%. In the final stages, TLS refinement (Schomaker & Trueblood, 1968; Winn *et al.*, 2001) was used to account for overall anisotropic motion of the molecules. Firstly, all B values of atoms were set to 30; TLS refinement with each molecule as one rigid group was then performed and individual positional and isotropic B values were subsequently refined. When all B values were set to 30, R and R_{free} were 23.9 and 27.9%, respectively; TLS refinement gave a 3% drop in R and R_{free} (19.9 and 24.0%, respectively) and individual atomic refinement improved R and R_{free} further to 18.3 and 22.7%, respectively. The final TLS parameters are listed in Table 2. H atoms were added at their calculated positions and their contribution to structure factors was taken into account. Addition of H atoms is one way to evaluate the correctness of the atomic positions: if R and R_{free} drop when the H atoms are included, it means the atomic positions are correct. In the case of S100A12 structure, the R and R_{free} values dropped to 17.8 and 21.6% after the H atoms were included. The results of refinement are summarized in Table 3.

2.3. Sequence analysis

Multiple sequence-alignment and sequence-similarity analysis of 19 S100 proteins were performed using the *PILEUP* and *PLOTSIMILARITY* programs within the Wisconsin Package (GCG, Madison, Wisconsin, USA; Genetics Computer Group, 1994).

Table 4
Root-mean-square deviations (main-chain atoms) between S100A12 and the known X-ray structures of other S100 proteins.

For calbindin, S100B and S100A11 r.m.s.$_{xyz}$ was calculated between the residues 6–40 and 51–80 of S100A12 and corresponding regions of the other molecules; linker loop regions were excluded from calculations. In the case of S100A7, r.m.s.$_{xyz}$ was calculated for residues 6–18, 21–24, 26–40 and 51–80 of S100A12. For comparison with S100A10, the residues of S100A12 were 6–19 and 26–40; for comparison with calmodulin the residues were 6–18 and 25–40.

	R.m.s.$_{xyz}$ (Å)
S100A10 without peptide	0.81
S100A10 plus peptide	0.94
S100A11	0.92
S100B	1.04
Calbindin	1.17
S100A7	1.57
Calmodulin	3.01

3. Results and discussion

3.1. Quality of the final model

The structure of S100A12 has been refined to an R factor of 17.8% for all data without a σ cutoff in the resolution range 25.0–1.95 Å excluding 5% randomly distributed reflections assigned to calculate R_{free} (21.6%). There is a dimer of S100A12 in the asymmetric unit. The final electron-density map allowed the positioning of the first 87 of 91 residues for chain A and the first 88 of 91 residues for chain B. The model contains 1413 atoms including four calcium ions and 189 water molecules. The dispersion precision indicator (Cruickshank, 1996; Murshudov & Dodson, 1997) estimates an average root-mean-square (r.m.s.) coordinate error of 0.14 Å. The overall G factor calculated by *PROCHECK* (Laskowski *et al.*, 1993, 1994) as a measure of stereochemical quality of the model is 0.11, which is somewhat better than expected at this resolu-

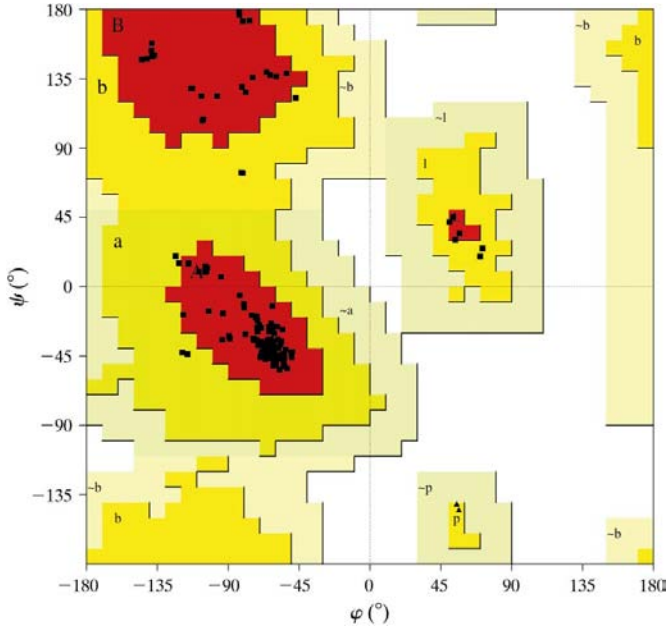

Figure 1
Ramachandran plot for both subunits of SA100A12 calculated using *PROCHECK* (Laskowski *et al.*, 1993).

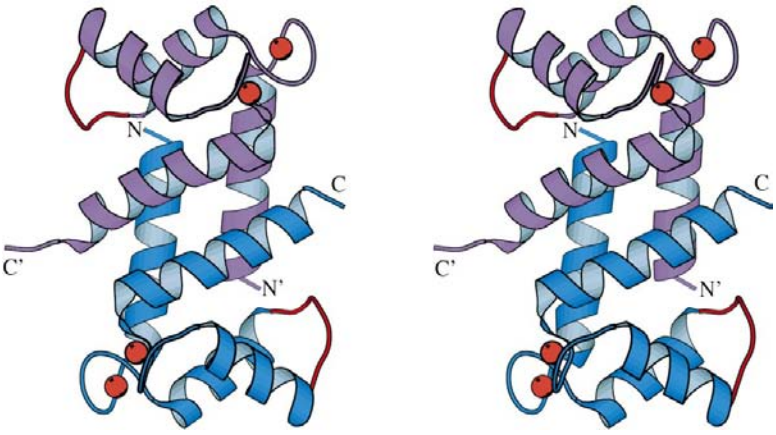

Figure 2
Stereoview of the S100A12 dimer. Ribbons of the two monomers are in dark cyan and purple. Calcium ions are shown in red and the flexible linker loop in orange. This figure and Figs. 3, 6, 7, 8 and 9 were generated using the program *MOLSCRIPT* (Kraulis, 1991).

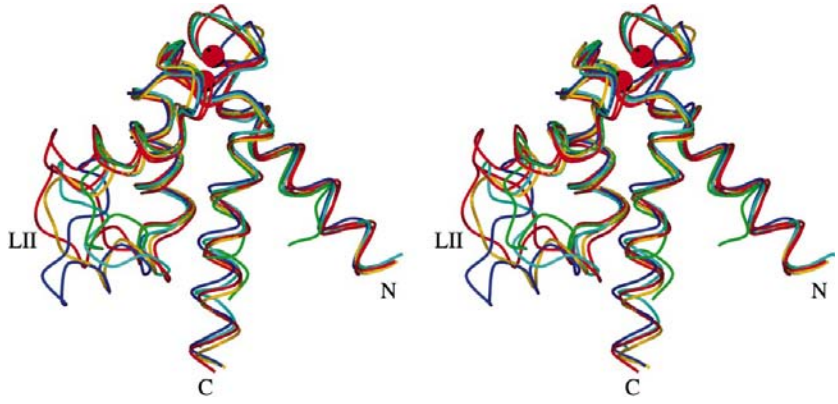

Figure 3
C^{α} superposition of the S100A12 monomer with the corresponding regions of the other S100 family members. S100A12 is shown in cyan, S100A7 (residues 3–89) in purple, S100A10 (residues 2–86) in yellow, S100A11 (residues 6–87) in orange, S100B (residues 2–88) in red and calbindin (residues 1–75) in green.

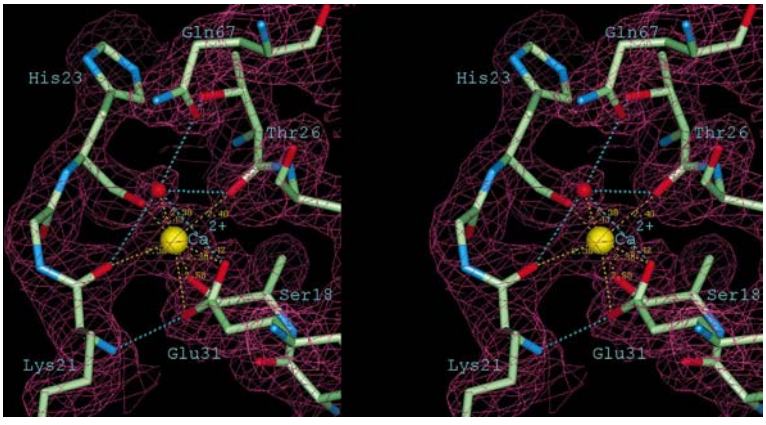

Figure 4
Stereoview of the electron density for the region of the N-terminal Ca-binding loop overlapped with the final model, the $2F_o - F_c$ electron-density map contoured at the 1σ level (0.43 e Å^{-3}). Ca^{2+} coordination is shown in yellow dashed lines and the hydrogen bonds are indicated as blue dashed lines. The calcium ion is yellow and the water molecule is red. The figure was generated using *QUANTA* (Molecular Simulations).

tion. Of the non-glycine residues, 94.5% fall in the most favoured regions of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) and 5.5% fall in additionally allowed regions (Fig. 1).

3.2. The S100A12 monomer

The overall fold of the S100 monomer consists of four α-helices and is similar to that of the other S100 proteins (Figs. 2 and 3). The first helix (H_I) is formed by residues 2–18, the second (H_{II}) by residues 29–39, the third (H_{III}) by residues 50–60 and the fourth (H_{IV}) by residues 70–86 (85 for subunit *B*). The pairs of helices H_I–H_{II} and H_{III}–H_{IV} are linked by the calcium-binding loops L_I and L_{III}, respectively.

The calcium-binding EF-hand motifs H_I–L_I–H_{II} and H_{III}–L_{III}–H_{IV} are linked by a short antiparallel β-bridge formed by the residues Leu27 and Asp69. Such a β-bridge is a common feature of the EF-hand proteins (Falke *et al.*, 1994). Another direct hydrogen bond between the two EF-hand motifs is formed between residues Thr26 and Glu67. In EF-1, the calcium ion is coordinated by the peptide carbonyl groups of residues Ser18, Lys21, His23, Thr26, a bidentate carboxylate group of Glu31 and a water molecule. This water molecule also forms one of the hydrogen bonds with the side chain of Gln67 of the EF-2 providing an additional link between the two calcium-binding motifs (Fig. 4). It could therefore play a role in the cross-talk between the EF-hand domains. In EF-2, the calcium ion is coordinated by the monodentate carboxylate groups of Asp61, Asn63, Asp65, the peptide carbonyl group of Gln67, the bidentate carboxylate group of Glu72 and a water molecule. For both EF hands, the geometry of the ligands is a classic pentagonal bipyramid.

It has been shown previously (Kligman & Hilt, 1988; Potts *et al.*, 1995) that the hydrophobic core packing is similar for all S100 proteins; Fig. 5 illustrates the high sequence similarity for the residues involved in buried hydrophobic interactions. The results of mutation studies (Kragelund *et al.*, 1998) imply that hydrophobic core residues have a significant influence on the calcium-binding properties of calbindin. Mutation studies of S100A4 have identified residues Phe72 and Tyr75 of S100A4 (Phe70 and Phe73 in S100A12) as crucial for protein integrity (Tarabykina *et al.*, unpublished results). The structure of S100A12 shows that Phe70 and Phe73 together with Phe14 and several other residues form a central hydro-

phobic core and suggests that replacing any of these three bulky residues would have drastic consequences for the proper folding and structural stability. These phenylalanines are highly conserved throughout the S100 proteins (Fig. 5) and are in the same conformation in all known X-ray structures of S100 proteins (Fig. 6).

Helices H_{II} and H_{III} are linked by the hinge-region L_{II}. Sequence comparison between different S100 proteins shows that this is the region of least sequence similarity (Kligman & Hilt, 1988; Fig. 5). It was proposed that this region is responsible for the specificity in target recognition long before any three-dimensional structural information was obtained.

3.3. Dimer structure

The overall size of the S100A12 dimer (Fig. 2) is approximately 45 × 30 × 25 Å. The total area buried between the two subunits is 2750 Å^2. The structures of the crystallographically independent monomers *A* and B are essentially the same, except for several surface residues such as Asn12, His15, Phe24, Asp25, Glu55, Glu66 and several surface lysines, for which the electron density is sometimes ill defined. There is a significant difference in the conformations of the C-termini (residues 86–87 for monomer *A* and residues 86–88 for monomer *B*) which could be the result of different crystal packing. The root-mean-square deviation in the main-chain atom positions (r.m.s.$_{xyz}$) between the main-chain atoms of subunits *A* and *B* is 0.272 Å. The dimer is stabilized mainly by hydrophobic contacts between helices H_{IV} (and partly H_I) of different subunits which are antiparallel (Fig. 7). The EF-hand loops L_I and L_{III} and the L_{II} linker regions are positioned on the outer surfaces of the dimer.

The dimer is formed in a similar way to those of other S100 proteins, with two central hydrophobic clusters interacting with each other through the dimer interface (Fig. 7). There is sequence similarity in the interface region (Fig. 5), but most of the residues involved in the inter-subunit contacts are simultaneously involved in intra-monomer interactions and the sequence similarity is probably a consequence of these residues being essential for the monomer stability. This is supported by the observation that calbindin usually has the same residues in the same conformations even though it does not form dimers. The only residues involved exclusively in the inter-subunit interactions are

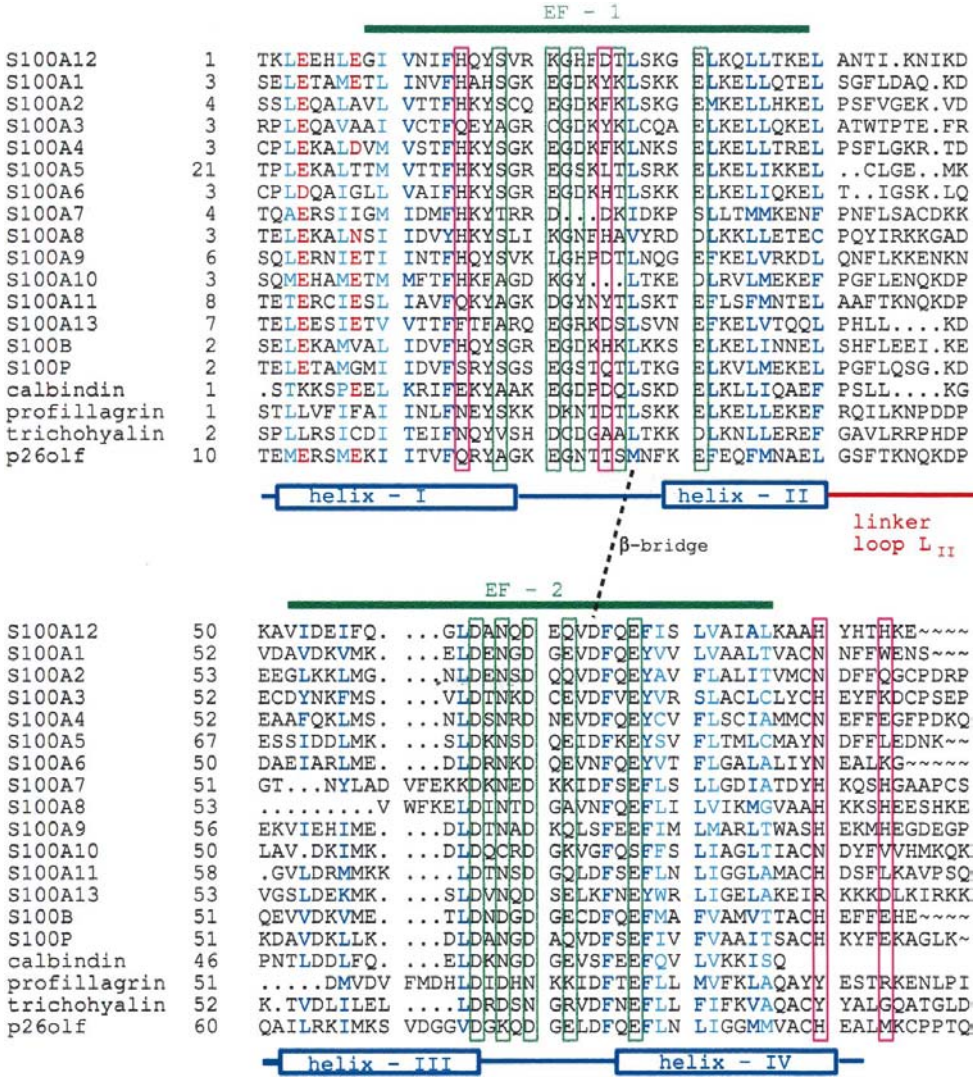

Figure 5
Sequence alignment of 19 known different members of the S100 family. All the sequences are shown for the human proteins except for p26olf, which is from the frog, as the human protein is not known. The secondary structure corresponds to S100A12. Helices are shown by blue boxes, calcium-binding loop regions are shown by blue lines and the linker loop connecting two EF-hand motifs is shown by a red line. Residues binding calcium are outlined with green boxes and those binding zinc in the S100A7 structure with magenta boxes. The EF-hand consensus motifs are shown by green lines above the alignment table. Conserved N-terminal glutamates that are proposed to be involved in target binding are in red. Hydrophobic core residues are in blue; those involved in dimer formation are in light blue. All sequences were obtained from the PIR protein sequence database; alignment was performed using *GCG* (Wisconsin package). Protein accession numbers in the PIR database are PIR2:JC4712 (S100A12), PIR1:BCHUIA (S100A1), PIR2:A41988 (S100A2), PIR2:C48219 (S100A3), PIR2:A48219 (S100A4), PIR2:B48219 (S100A5), PIR1:BCHUY (S100A6), PIR1:A54327 (S100A7), PIR1:BCHUCF (S100A8), PIR1:B31848 (S100A9), PIR1:JC1139 (S100A10), PIR1:I137080 (S100A11), PIR2:JC5064 (S100A13), PIR1:BCHUIB (S100B), PIR2:S24146 (S100P), PIR2:A48118 (profillagrin), PIR1:A45973 (trichohyalin), PIR2:JE0330 (p26olf), PIR1:JN0246 (calbindin).

Leu3 and His6 (Leu3 is highly conserved, His6 is not) located in the region absent in monomeric calbindin. Residues mostly involved in dimer formation, such as Ile74 and Leu81, have hydrophobic substitutions throughout the S100 sequences: this probably explains why not all S100 proteins are able to form heterodimers with one another.

Dimerization is supposed to be essential for target recognition by S100 proteins (Krebs *et al.*, 1995; Donato, 1999). There is even an S100 protein, p26olf from olfactory tissue of frog, that consists of one uninterrupted chain formed by two nearly identical S100-like domains (Miwa *et al.*, 1998). The results of molecular modelling show that it most probably adopts a folding pattern similar to a typical S100 protein (Tanaka *et al.*, 1999). The importance of dimerization for S100 function has been proved by the X-ray structures of S100A11 and S100A10, where the target-binding site is formed by both subunits. The biological consequences of S100 dimerization could be to bring together two protein targets or to disrupt interaction between two proteins. The heterodimerization phenomenon provides an interesting mechanism underlying the diversity of S100 function.

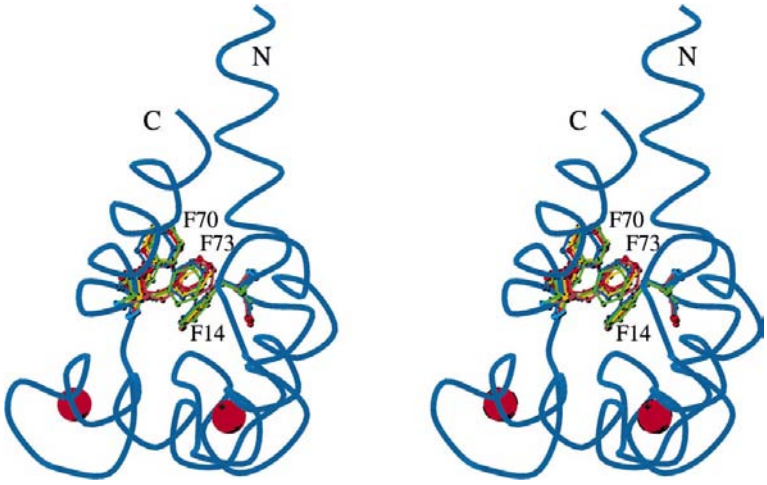

Figure 6
Stereo plot of the C^{α} trace of the S100A12 monomer in cyan with three conserved phenylanines contributing to the hydrophobic core. The phenylalanines from six different S100 X-ray structures are in ball-and-stick, in red for S100B, in gold for S100A10, in orange for S100A11, in purple for S100A7 and in green for calbindin.

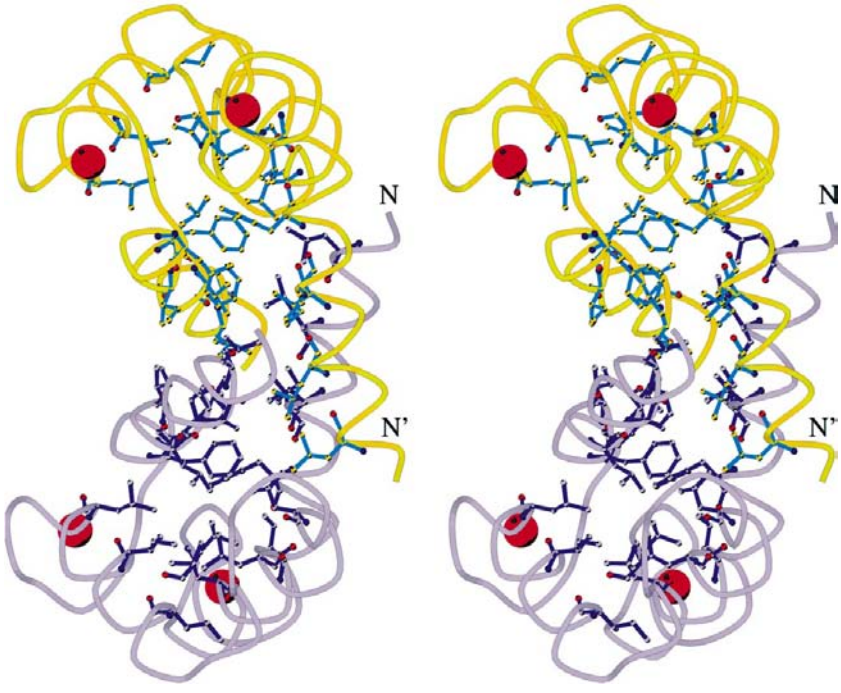

Figure 7
Stereoview of the S100A12 dimer (in grey and yellow) with conserved residues participating in hydrophobic interactions given in ball-and-stick representation; residues from one subunit are shown in dark blue and residues from another subunit are shown in light blue.

3.4. Comparison with the other known S100 structures

The overall structure of S100A12 is rather similar to the other known X-ray structures of S100 proteins (Table 4). The smallest r.m.s.$_{xyz}$ in the main-chain atom positions is with the structure of S100A10 without the target peptide. The calcium-binding loops in this case were excluded from the comparison because there are significant differences in amino-acid sequence in these regions of S100A10 relative to the majority of S100 proteins: these differences account for the inability of S100A10 to bind calcium. The r.m.s.$_{xyz}$ values for peptide-bound S100A10 and S100A11 are slightly larger, presumably because of small changes upon binding of the peptides.

Calmodulin was included in the comparison to illustrate the similarities and differences of the S100 EF-hands to those of calcium-binding proteins from another closely related subfamily (Chattopadhyaya *et al.*, 1992; Moncrief *et al.*, 1990).

3.5. Functional implications

Rapid progress in protein sequencing and X-ray structure determination of S100 proteins allows more comparisons than would have been possible a year ago. Fig. 5 shows a sequence alignment for 18 human S100 proteins and the N-terminal domain of p26olf from frog. Sequence alignment illustrates that the EF-hand regions have high sequence identity; the lowest identity occurs for residues 44–54, *i.e.* in the flexible linker L_{II}. The identity is also low for the C-terminus. It has been proposed previously (Kligman & Hilt, 1988) that these regions were involved in the specific target recogni-

tion. This hypothesis is supported by the way in which S100A10 (Réty *et al.*, 1999) and S100A11 (Réty *et al.*, 2000) bind their target peptides (annexin II and annexin I fragments, respectively) *via* the L_{II} loops and C-terminal regions of one subunit and the N-terminal regions of a second subunit.

The NMR structure of the complex of S100B with a p53 fragment shows that the general location of the binding site is similar to those of the binding sites of S100A10 and A11. As in S100A10 and A11, L_{II} loop and C-terminal residues are involved in interactions with the target, but the N-terminal residues do not take part in target binding and the peptide orientation in the case of S100B is different from that for the annexin fragments (Rustandi *et al.*, 2000).

Structural comparison of S100A12 with the other S100 proteins proves once more that the main differences lie in the region of the flexible loop L_{II} (Fig. 3). Superposition of the peptide-bound structures of S100A10 and S100A11 shows the region of S100A12 where the target is likely to be bound. While S100A10 and S100A11 form short α-helixes at the peptide-bound region of the loop L_{II}, there is no helix in the same region of S100A12 (Fig. 8). However, subunit *B* in the peptide-free structure of S100A10 does not form this short α-helix, while in the peptide-bound S100A10 the α-helices are formed in both subunits (Réty *et al.*, 1999), so there is reason to suppose that this helix might form upon target binding in the S100A12 structure. On the whole, there is sufficient sequence and structural similarity in the target-binding region to suppose that the hydrophobic residues involved in target binding are Val11, Leu40, Ala80 and Tyr86. In the structures of the S100A10–annexin II and S100A11–annexin I complexes, two N-terminal glutamates are involved in peptide recognition. These two residues are highly conserved, especially the former (corresponding to Glu4 in the S100A12 structure), suggesting that they might be taking part in target binding for some of the other S100 proteins (Fig. 5). Structural comparison shows that in S100A12 Glu4 and Glu8 are in the same conformation as in the S100A10 and S100A11 complexes (Fig. 9). Comparison of S100A10 structures with and without the target peptide shows that in the presence of the peptide the flexible loop L_{II} moves away to widen the target-binding cleft (region 42–44 shifts by 1.5–3.5 Å). It is possible that the equivalent region in A12 would also move upon target binding. The comparison also suggests that Ile44 would probably be involved in hydrophobic interactions with the target and that hydrogen-bonding interactions can be formed by Asn44, Gln45 or Lys46.

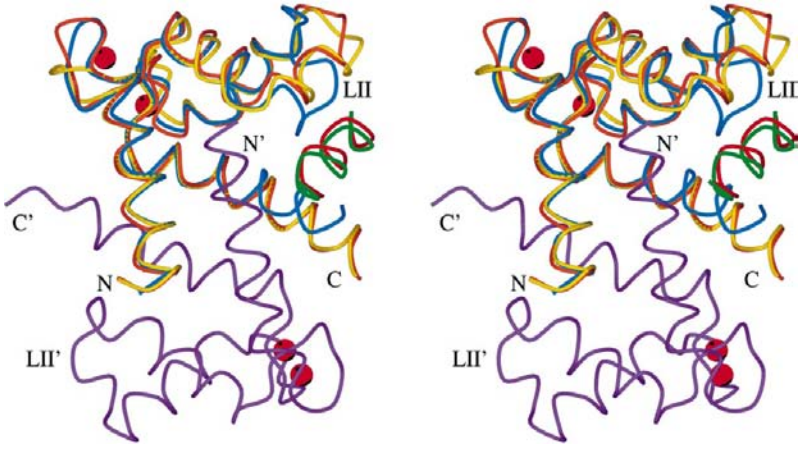

Figure 8
Stereoview of the C^{α} traces of the monomers of S100A10 (residues 1–87) and S100A11 (residues 5–94) bound to their targets (fragments of annexin II and annexin I, respectively) overlapped with the C^{α} trace of the S100A12 dimer. S100A10 is shown in yellow, S100A11 in orange, annexin II fragment in green, annexin I fragment in red and S100A12 monomers in cyan and purple.

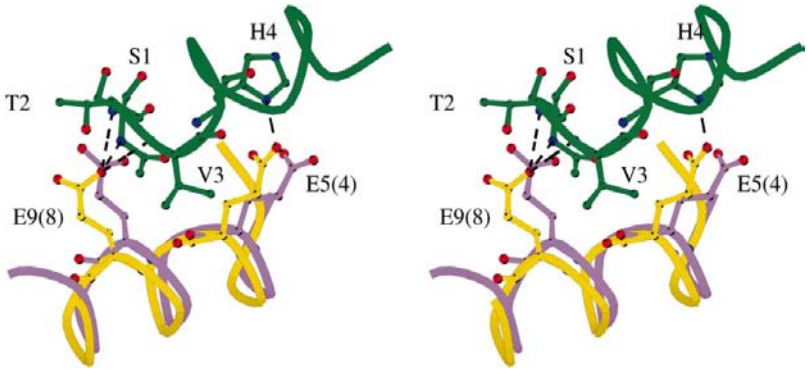

Figure 9
Superposition of the N-terminal target-binding region of S100A10 (residues 1–12) bound to the annexin II peptide on the corresponding region of S100A12 (residues 1–12). Glutamates forming hydrogen bonds to the target and the corresponding residues of the target peptide are shown in ball-and-stick. The corresponding glutamates in the S100A12 structure are shown in ball-and-stick. S100A12 is shown in purple, S100A10 in yellow and annexin II peptide in green.

There is a number of other conserved residues in S100 structures (Fig. 5). Hydrophobic residues presumably have a structural role, with Phe14, Phe70 and Phe73 being the most important for the hydrophobic core stability (Figs. 5 and 7). The role of the conserved Tyr17 and Glu39 is less well understood. These two residues have their side chains partially exposed to the surface of the protein, forming a hydrogen-bond interaction between each other. There is a set of conserved residues involved in Ca^{2+} binding (Fig. 5). Finally, there is a set of less conserved residues, His15, Asp25, His85 and His89, which are supposedly involved in Zn binding (Brodersen *et al.*, 1999).

The existing structural evidence is of course not sufficient to draw final conclusions on the exact model of Ca^{2+}-induced target binding by S100A12. We are currently trying to co-crystallize and determine the structure of the S100A12 complex with the RAGE receptor fragment, which will add substantially to the growing library of information of S100 structure–function relationships.

This work was funded by the Yorkshire Cancer Research, Danish Cancer Society, the Danish Medical Research Council, the Novo Foundation and the Danish Cancer Research Foundation. The York Structural Biology Laboratory is funded by an Infrastructure grant from the BBSRC No. 87/SB09829 (UK). We thank the staff of the ESRF for provision of data-collection facilities and financial assistance for data collection.

References

Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. K. (1985). *Nature (London)*, **315**, 37–40.

Baudier, J. & Cole, R. D. (1988). *J. Biol. Chem.* **263**, 5876–5883.

Baudier, J., Delphin, C., Grunwald, D., Khochbin, S. & Lawrence, J. J. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 11627–11631.

Bernstein, F. G., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. K., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brodersen, D. E., Etzerodt, M., Madsen, P., Celis, J. E., Thogersen, H. C., Nyborg, J. & Kjeldgaard, M. (1998). *Structure*, **6**, 477–489.

Brodersen, D. E., Nyborg, J. & Kjeldgaard, M. (1999). *Biochemistry*, **38**, 1695–1704.

Brzozowski, A. M. (1993). *Acta Cryst.* D**49**, 352–354.

Chattopadhyaya, R., Meador, W. E., Means, A. R. & Quiocho, F. A. (1992). *J. Mol. Biol.* **228**, 1177–1192.

Chazin, W. J. (1995). *Nature Struct. Biol.* **2**, 707–710.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Cruickshank, D. W. J. (1996). *Proceedings of the CCP4 Study Weekend, Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 11–22, Warrington: Daresbury Laboratory.

Dell'Angelica, E. C., Schleicher, C. H. & Santome, J. A. (1994). *J. Biol. Chem.* **269**, 28929–28936.

Donato, R. (1999). *Biochim. Biophys. Acta*, **1450**, 191–231.

Drohat, A. C., Amburgey, J. C., Abilgaard, F. & Starich, M. R. (1996). *Biochemistry*, **35**, 11577–11588.

Drohat, A. C., Tjandra, N., Baldisseri, D. M. & Weber, D. J. (1999). *J. Protein Sci.* **8**, 800–809.

Edgeworth, J., Gorman, M., Bennet, R., Freemont, P. & Hogg, N. (1991). *J. Biol. Chem.* **266**, 7706–7713.

Falke, J. J., Drake, S. K., Hazard, A. L. & Peersen, O. B. (1994). *Quart. Rev. Biophys.* **27**, 219–290.

Fohr, U. G., Heizmann, C. W., Engelkamp, D., Schäfer, B. W. & Cox, J. S. (1995). *J. Biol. Chem.* **270**, 21056–21061.

Genetics Computer Group (1994). *Program Manual for the Wisconsin Package, Version 8*. Genetics Computer Group, Madison WI, USA.

Guignard, F., Mauel, J. & Markett, M. (1995). *Biochem. J.* **309**, 395–401.

Heizmann, C. W. & Cox, J. A. (1998). *Biometals*, **11**, 383–397.

Herzberg, O. & James, M. N. G. (1985). *Nature (London)*, **313**, 653–659.

Hitomi, J., Maruyama, K., Kikuchi, Y., Nagasaki, K. & Yamaguchi, K. (1996). *Biochem. Biophys. Res. Commun.* **228**, 757–763.

Hoffmann, M. A., Drury, S., Fu, C., Qu, W., Taguchi, A., Lu, Y., Avila, C., Kambham, N., Bierhaus, A., Nawroth, P., Neurath, M. F., Slattery, T., Beach, D., McClary, J., Nagashima, M., Merser, J., Stern, D. & Schmidt, A. M. (1999). *Cell*, **97**, 889–907.

Ishikawa, K., Nakagawa, A., Tanaka, I., Suzuki, M. & Nishihira, J. (2000). *Acta Cryst.* D**56**, 559–566.

Isobe, T., Ishioka, N. & Okuyama, T. (1981). *Eur. J. Biochem.* **115**, 469–474.

Ivanenkov, V. V., Jamieson, G. A. Jr, Gruenstein, E. & Dimlich, R. V. (1995). *J. Biol. Chem.* **270**, 14651–14658.

Kerkhoff, C., Klempt, M. & Sorg, C. (1998). *Biochem. Biophys. Acta*, **1448**, 200–211.

Kligman, D. & Hilt, D. C. (1988). *Trends. Biochem. Sci.* **13**, 437–442.

Komada, T., Araki, R., Nakatani, K., Yada, I., Naka, M. & Tanaka, T. (1996). *Biochem. Biophys. Res. Commun.* **220**, 871–874.

Kragelund, B. B., Jöhnsson, M., Bifulco, G., Chazin, W. J., Nilsson, H., Finn, B. E. & Linse, S. (1998). *Biochemistry*, **37**, 8926–8937.

Kraulis, P. (1991). *J. Appl. Cryst.* **24**, 946–950.

Krebs, J., Quadroni, M. & Van Eldik, L. J. (1995). *Nature Struct. Biol.* **2**, 711–714.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. (1994). *Proceedings of the CCP4 Study Weekend*, *From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 149–159. Warrington: Daresbury Laboratory.

Lee, S. C., Kim, I. G., Marekov, L. N., O'Keffee, E. J., Parry, D. A. D. & Steinert, P. M. (1993). *J. Biol. Chem.* **268**, 12164–12176.

Mäler, L., Blankenship, J., Rance, M. & Chazin, W. J. (2000). *Nature Struct. Biol.* **7**, 245–250.

Mäler, L., Potts, B. C. M. & Chazin, W. J. (1999). *J. Biomol. NMR*, **13**, 233–247.

Markova, N. G., Marekov, L. N., Chipev, C. C., Gan, S. Q., Idler, W. W. & Steinert, P. M. (1993). *Mol. Cell Biol.* **13**, 613–625.

Matsumura, C. H., Shiba, T., Inoue, T., Harada, S. & Kai, Y. (1998). *Structure*, **6**, 233–241.

Miwa, N., Kobayashi, M., Takamatsu, K. & Kawamura, S. (1998). *Biochem. Biophys. Res. Commun.* **251**, 860–867.

Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.

Moncrief, N. D., Kretsinger, R. H. & Goodman, M. (1990). *J. Mol. Evol.* **30**, 522–562.

Moore, B. W. (1965). *Biochem. Biophys. Res. Commun.* **19**, 739–744.

Moroz, O. V., Antson, A. A., Dodson, G. G., Wilson, K. S., Skibshoj, I., Lukanidin, E. M. & Bronstein, I. B. (2000). *Acta Cryst.* D**56**, 189–191.

Murshudov, G. N. & Dodson, E. J. (1997). *CCP4 Newslett. Protein Crystallogr.* **33**, 31–19.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Nakayama, S., Moncrief, N. D. & Kretsinger, R. H. (1992). *J. Mol. Evol.* **34**, 416–448.

Nonato, M. C., Garrat, R. C., Schleicher, C. H., Santome, J. A. & Oliva, G. (1997). *Acta Cryst.* D**53**, 200–202.

Oldfield, T. J. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Walker, pp 15–18. Warrington: Daresbury Laboratory.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.

Potts, B. C. M., Smith, J., Akke, M., Macke, T. J., Okazaki, K., Hidaka, H., Case, D. A. & Chazin, W. J. (1995). *Nature Struct. Biol.* **2**, 791–796.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Réty, S., Osterloh, D., Arie, J.-P., Tabaries, S., Seeman, J., Russo-Marie, F., Gerke, V. & Lewit-Bentley, A. (2000). *Structure*, **8**, 175–184.

Réty, S., Sopkova, J., Renouard, M., Osterloh, D., Gerke, V., Tabaries, S., Russo-Marie, F. & Lewit-Bentley, A. (1999). *Nature Struct. Biol.* **6**, 89–95.

Rustandi, R. R., Baldisseri, D. M. & Weber, D. (2000). *Nature Struct. Biol.* **7**, 570–574.

Schäfer, B. W. & Heizmann, C. W. (1996). *Trends Biochem. Sci.* **21**, 134–140.

Schäfer, B. W., Wicki, R., Engelkamp, D., Mattei, M.-G. & Heizman, C. W. (1995). *Genomics*, **25**, 638–643.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
Shishibori, T., Oyama, Y., Matsushita, O., Yamashita, K., Furuichi, H., Okabe, A., Maeta, H., Hata, Y. & Kobayashi, R. (1999). *Biochem. J.* **338**, 583–589.
Skelton, N. J., Kordel, J. & Chazin, W. J. (1995). *J. Mol. Biol.* **249**, 441–462.
Smith, S. P. & Shaw, G. S. (1998). *Biochem. Cell Biol.* **76**, 324–333.
Svensson, L. A., Thulin, E. & Forsen, S. (1992). *J. Mol. Biol.* **223**, 601–606.
Szebenyj, D. M. E. & Moffat, K. (1986). *J. Biol. Chem.* **261**, 8761–8777.
Taguchi, A., Blood, D. C., del Toro, G., Canet, A., Lee, D. C., Qu, W., Tanji, N., Lu, Y., Lalla, E., Fu, C., Hofmann, M. A., Kislinger, T., Ingram, M., Lu, A., Tanaka, H., Hori, O., Ogawa, S., Stern, D. M. & Schmidt, A. M. (2000). *Nature (London)*, **405**, 354–360.
Tanaka, T., Miwa, N., Kawamura, S., Sohma, K., Nitta, K. & Matsushima, N. (1999). *Protein Eng.* **12**, 395–405.
Tarabykina, S., Kriajevska, M., Scott, D. J., Hill, T. J., Lafitte, D., Derrick, P. J., Dodson, G. G., Lukanidin, E. & Bronstein, I. B. (2000). *FEBS Lett.* **475**, 187–191.
Vagin, A. A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vogl, T., Propper, C., Hartmann, M., Strey, A., Strupat, K., van den Bos, C., Sorg, C. & Roth, J. (1999). *J. Biol. Chem.* **274**, 25291–25296.
Wang, G., Rudland, P. S., White, M. R. & Barraclough, R. (2000). *J. Biol. Chem.* **275**, 11141–11146.
Wicki, R., Marenholz, I., Mischke, D., Schafer, B. W. & Heizmann, C. W. (1996). *Cell Calcium*, **20**, 459–464.
Winn, M. D., Isupov, M. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Yang, Q., O'Hanlon, D., Heizmann, C. W. & Marks, A. (1999). *Exp. Cell. Res.* **246**, 501–509.
Zhang, M., Tanaka, T. & Ikura, M. (1995). *Nature Struct. Biol.* **2**, 758–766.