doi:10.1006/jmbi.2000.4379 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 306, 25–35

JMB

Crystal Structure of the SMC Head Domain: An ABC ATPase with 900 Residues Antiparallel Coiled-coil Inserted

Jan Löwe*, Suzanne C. Cordell and Fusinita van den Ent

MRC, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

**Corresponding author*

SMC (structural maintenance of chromosomes) proteins are large coiled-coil proteins involved in chromosome condensation, sister chromatid cohesion, and DNA double-strand break processing. They share a conserved five-domain architecture with three globular domains separated by two long coiled-coil segments. The coiled-coil segments are antiparallel, bringing the N and C-terminal globular domains together. We have expressed a fusion protein of the N and C-terminal globular domains of *Thermotoga maritima* SMC in *Escherichia coli* by replacing the approximately 900 residue coiled-coil and hinge segment with a short peptide linker. The SMC head domain (SMChd) binds and condenses DNA in an ATP-dependent manner. Using selenomethionine-substituted protein and multiple anomalous dispersion phasing, we have solved the crystal structure of the SMChd to 3.1 Å resolution. In the monoclinic crystal form, six SMChd molecules form two turns of a helix. The fold of SMChd is closely related to the ATP-binding cassette (ABC) ATPase family of proteins and Rad50, a member of the SMC family involved in DNA double-strand break repair. In SMChd, the ABC ATPase fold is formed by the N and C-terminal domains with the 900 residue coiled-coil and hinge segment inserted in the middle of the fold. The crystal structure of an SMChd confirms that the coiled-coil segments in SMC proteins are antiparallel and shows how the N and C-terminal domains come together to form an ABC ATPase. Comparison to the structure of the MukB N-terminal domain demonstrates the close relationship between MukB and SMC proteins, and indicates a helix to strand conversion when N and C-terminal parts come together.

Keywords: chromosome segregation; structural maintenance of chromosomes; MukB; SMC; Rad50

Introduction

SMC (structural maintenance of chromosomes) proteins are large coiled-coil proteins, consisting of about 800 to 1500 residues, involved in sister chromatid cohesion, chromosome condensation, and DNA double-strand break processing (for recent reviews, see Cobbe & Heck, 2000; Hirano, 1998; Jessberger *et al*., 1998). They form a large family of proteins and are ubiquitous in eukaryotes and very common in eubacteria and archaea. SMC proteins consist of five domains. The globular N and C-terminal domains are connected by a long coiled-coil region to a globular hinge domain, resulting in a V-shaped or headphone-like structure. The N-terminal domain contains a Walker A motif, whereas the C-terminal domain contains a Walker B motif (Walker *et al*., 1982). SMC proteins bind ATP and have weak ATPase activity. SMC molecules probably form heterodimers in eukaryotes and are homodimers in eubacteria and archaea. They contain about 800 to 1500 residues per chain, producing molecules that, when fully stretched, are 1000-1500 Å long. Using electron microscopy, it has been shown that the coiled-coil segments are antiparallel and that the N and

Abbreviations used: SMC, structural maintenance of chromosomes; hd, head domain; cd, catalytic domain; ABC, ATP-binding cassette; MAD, multiple anomalous dispersion; DSBR, double-strand break repair; HisP, histidine permease.

E-mail address of the corresponding author: jyl@mrc-lmb.cam.ac.uk

0022-2836/01/010025–11 $35.00/0

C-terminal domains are located at both ends of the V-shaped molecule (Melby *et al*., 1998). The hinge domain is placed at the neck of the V-shaped molecule. Two arrangements of the dimers are possible, with one SMC chain going from one end of the V to the other, or one SMC chain forming only one leg of the V-shaped molecule, in both cases forming the same antiparallel coiled-coil. In both arrangements, the hinge domains dimerize, but the N and C termini form either intra- or intermolecular heterodimers. The crystal structure of the N-terminal domain of MukB, a distantly related SMC homologue in *Escherichia coli*, showed that the Walker A motif is exposed on the surface and indicated strongly that the N and C-terminal domains in SMC molecules have to come together in a single domain to create a functional ATPase (van den Ent *et al*., 1999). A recent crystal structure of the Rad50cd (catalytic domain) confirmed that the N and C-terminal domains of an SMC protein involved in DNA double-strand break repair form a single globular ABC (ATP-binding cassette) ATPase domain (Hopfner *et al*., 2000).

On the basis of their sequences and function, SMC proteins can be grouped into eight families (Cobbe & Heck, 2000). The first four families contain SMC proteins SMC1-SMC4, which are involved in chromosome dynamics in eukaryotes. Families 5 and 6 contain the Rad50 and Rad18 proteins that are involved in DNA repair. Bacterial SMC proteins and distantly related SMC proteins like *E. coli* MukB form SMC families 7 and 8, respectively.

In eukaryotes, SMC1 and SMC3 are part of 14 S cohesin, together with Scc1 and Scc3 (yeast proteins) (Koshland & Guacci, 2000; Nasmyth *et al*., 2000). During the separation process of sister chromatids, at the metaphase-to-anaphase transition, Scc1 is proteolytically cleaved by separin protein Esp1. A tentative model of the role of SMCs in this process is that Scc1/Scc3 bind two SMC heterodimers each with DNA bound to their head domains.

Eukaryotic SMC2 and SMC4 are part of the 13 S condensin complex, together with three other proteins (Hirano *et al*., 1997; Hirano & Mitchison, 1994). The condensin complex introduces (+) writhe into DNA by bending the DNA into coils that remodel the chromosome into a more compact structure (Kimura *et al*., 1999). A very similar role has been proposed for the MukBEF complex, which is thought to be the functional homologue of condensin in *E. coli* (Sawitzke & Austin, 2000). However, no cofactor has been reported for other bacterial SMC proteins so far.

A third role for SMC proteins is in DNA repair. The two double-strand break repair (DSBR) pathways, homologous recombination and non-homologous end joining depend on the Rad50/Mre11/Nbs1 complex (Haber, 1998). Rad50 is a member of a specialized subfamily of SMC proteins, together with SbcC from *E. coli* (Connelly *et al*., 1998) These SMC proteins carry a unique signature motif CXXC in the hinge domain, which is much smaller in these proteins. The precise role of Rad50 in DSBR is not known, but Rad50 mutants incapable of ATP binding have a null phenotype.

SMC proteins are highly conserved. All eukaryotes from yeast to humans seem to contain at least SMC1, 2, 3, and 4. Also in bacteria, SMCs are wide-spread, but no SMC or MukB homologues have been found for some organisms, such as *Rickettsia prowazekii* and *Methanobacterium thermoautotrophicum*. In general, bacteria contain both an SMC homologue and a Rad50/SbcC homologue.

Still enigmatic is the role of SMC proteins outside the cell. Bamacan, an extracellular proteoglycan, contains or is SMC3 protein. Overexpression of SMC3/bamacan in fibroplasts generated foci of transformation and SMC3-overexpressing fibroplasts acquired anchorage-independent growth (Ghiselli & Iozzo, 2000).

Here, we present the crystal structure of the head domain of SMC from *Thermotoga maritima* at 3.1 Å resolution as solved by multiple anomalous dispersion (MAD). The SMC head domain (SMChd) has been expressed in *E. coli* as a fusion protein replacing the coiled-coil segment with a short peptide linker segment.

Results

Protein expression and structure determination

In order to obtain a functional SMChd, a covalent fusion of the N and C-terminal domains of *T. maritima* SMC was cloned and expressed in *E. coli*. From 12 l of culture, approximately 10 mg protein could be obtained. The protein is highly soluble and stable in 200 mM sodium chloride-containing buffer, and runs as a monomer under these conditions on size-exclusion columns (data not shown). It is not stable in low-salt buffer. After incubation of the protein with MgATP, it remains monomeric on size-exclusion columns in 200 mM sodium chloride (data not shown). SMChd is capable of ATP-dependent DNA aggregation, as demonstrated in Figure 1.

SMChd has been crystallized using sodium formate and PEG as precipitants to form trigonal and monoclinic crystals. These crystals diffract to 2.8 Å resolution and are hemihedrally twinned with twin fractions ranging from 0.3 to 0.4 as detected by Yeates' statistical method (Yeates, 1997). The twin operator is $h + k, -k, -l$ producing pseudo $P3_112$ symmetry. The monoclinic crystals diffract to 3.1 Å resolution and a selenomethionine-substituted monoclinic crystal was used in a three-wavelength MAD experiment to solve the phase problem (Table 1). All 30 sites were located using the anomalous differences of the PEAK wavelength and a noise-free electron density map at 3.5 Å resolution was obtained (Figure 2(a)).

The atomic model contains residues 1-50 and 61-146 of the N-terminal domain, and residues 1024-1045 and 1060-1164 of the C-terminal domain.

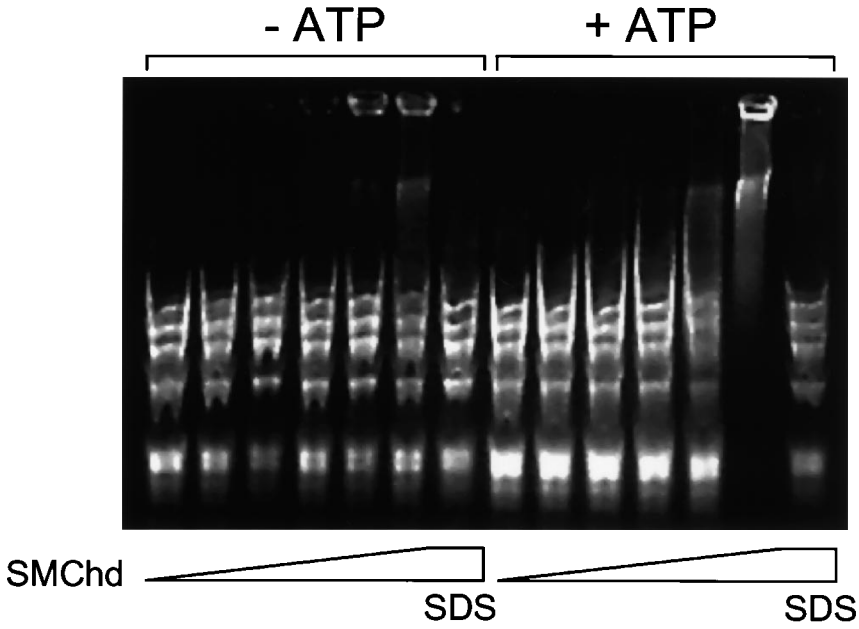

Figure 1 . SMC head domain (SMChd) DNA aggregation assay. Increasing amounts of SMChd were incubated with digested ΦX174 plasmid with and without ATP. All lanes: reactions in 100 mM Tris (pH 7.8), 100 mM NaCl, 10 mM magnesium acetate, 0.02 mg/ml ΦX174/*Hae*III-digested plasmid DNA. Lanes 8-14 contain 2 mM neutralized ATP, lanes 7 and 14 contain 1 % SDS in the reaction. Increasing amounts of SMChd (lanes 1, 8 0.0156 mg/ml 1:1 serial dilution to 1 mg/ml lanes 7, 14) were incubated at 37 °C for ten minutes and separated on a 1 % agarose gel in TB buffer.

A portion of the final $2F_o - F_c$ electron density map after refinement at 3.1 Å resolution is shown in Figure 2(b). The trigonal crystal form could be solved by molecular replacement using the refined SMChd model and hemihedrally detwinned data.

SMChd crystal structure

The N and the C-terminal parts of SMC from *T. maritima* form a single globular domain with the coiled-coil segment inserted between helices HD and HE (Figure 3). The structure contains three β-sheets with the central β-sheet containing elements from both the N and C-terminal parts of the molecule. The N-terminal part contributes the lower (Figure 3) antiparallel β-sheet, strand S3 to the central β-sheet, the P-loop between S3/HA, and S8 to the central β-sheet. It ends with helix HD, which, together with helix HE, points in the direction of the coiled-coil attached to the SMC head domain. The 14 amino acid residue linker used to create the SMChd fusion protein is not ordered in the crystal structure, suggesting it is flexible and designed long enough not to create structural tension between the N and C-terminal parts. The SMC C-terminal part contributes the upper small β-sheet, as well as helices HE, HF, and HG to SMChd (Figure 3). Helices HE, HF, and HG are sandwiched between the upper and central β-sheet.

Both the monoclinic and trigonal crystal forms contain a continuous helix of head-to-tail arranged

Table 1. Crystallographic data

Crystal	λ(Å)	Resol.(Å)	$I/\sigma I$[a]	R_m (%)[b]	Multipl.[c]	Compl. (%)[d]	Occupancies[e]	f'/f''[f]
PEAK	0.9787	3.5	5.3	0.058	6.4(3.2)	97.2	8.4(8.5)	−6/7
INFL	0.9790	3.5	4.4	0.075	7.4(3.7)	97.8	0.000(2.6)	−10/3
HREM	0.9137	3.5	3.9	0.079	7.4(3.7)	97.8	11.3(4.5)	−2/3
NATI	1.2440	3.1	3.0	0.071	2.3	97.3		

Space group $P2_1(4)$, $a = 134.00$ Å, $b = 49.19$ Å, $c = 233.88$ Å, $\beta = 94.63°$.

[a] Signal to noise ratio for the highest resolution shell of intensities.

[b] R_m: $\Sigma_h\Sigma_i|I(h,i) - I(h)|/\Sigma_h\Sigma_i I(h,i)$ where $I(h,i)$ are symmetry-related intensities and $I(h)$ is the mean intensity of the reflection with unique index h.

[c] Multiplicity for unique reflections, anomalous multiplicity in parentheses.

[d] Completeness for unique reflections, anomalous completeness is identical because inverse beam geometry was used.

[e] Occupancies refined in MLPHARE. Occupancies for MAD datasets given in units of scattering electrons.

[f] f'/f'' ratio as determined from fluorescence scan of the crystal.

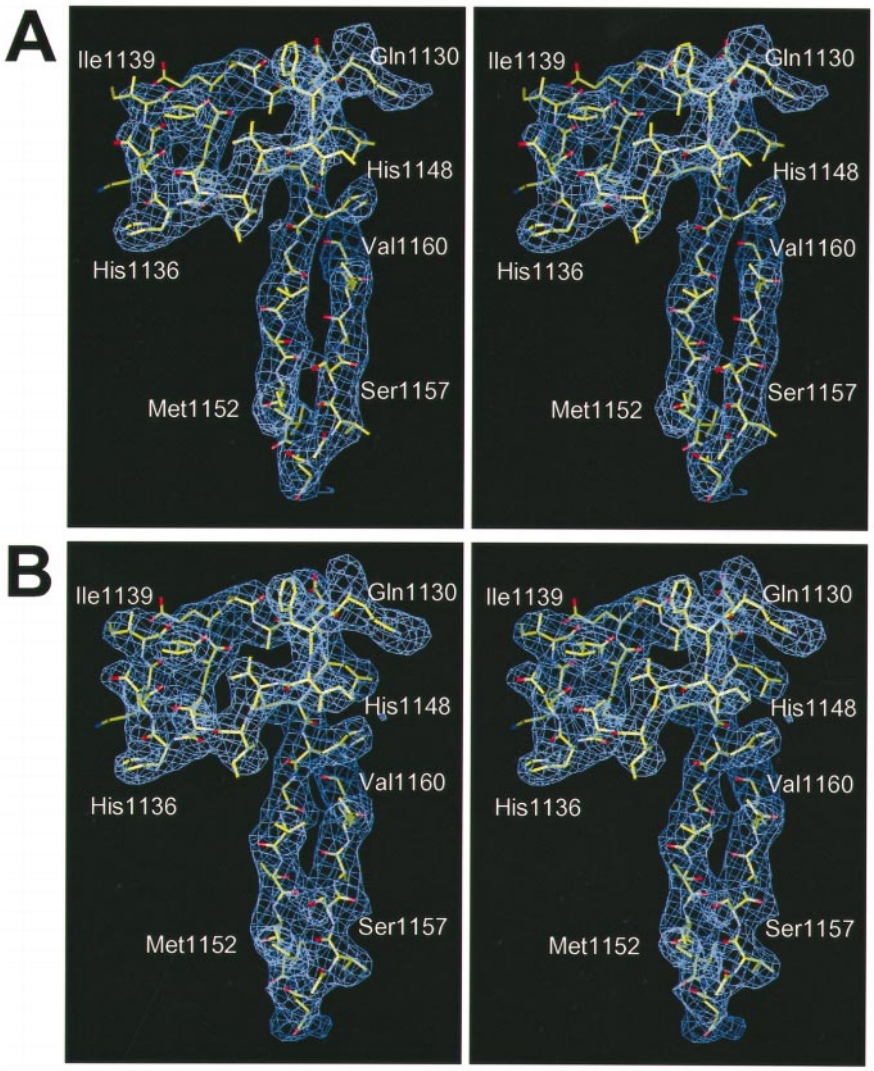

Figure 2. (a) Stereo drawing of the solvent-flattened MAD electron density map at 3.5 Å resolution (contoured at 1σ). Residues 1130-1160 of the C-terminal part of SMChd are shown. (b) Stereo drawing of the same stretch of residues with the final $2F_o - F_c$ electron density map at 3.1 Å resolution superimposed. The Figure was prepared with MAIN (Turk, 1992).

SMChd molecules (Figure 4(a)). The trigonal crystal form is made from six of these helices related by 2-fold axes (Figure 4(b)). The 3_1 screw axis of the crystals generates the helical repeat from six monomers. The monoclinic crystal form contains one helix per asymmetric unit with six non-crystallographically related SMChd molecules generating two turns of the same 3_1 helix as in the trigonal crystal form. Both crystal forms suffer from a lack of crystal contacts linking these helices together, giving rise to very anisotropic scattering. In the case of the monoclinic crystals, the helices are tilted by 45° relative to the crystal axes.

SMChd is an ABC ATPase

A DALI search (Holm & Sander, 1995) for related structures showed a very clear relationship of the fold of SMChd to the recent structure of the Rad50 catalytic domain (PDB ID 1F2T, chain A, B, Hopfner *et al.*, 2000), MukB N-terminal domain (PDB ID 1QHL, van den Ent *et al.*, 1999), and histidine permease ATP-binding domain (PDB ID 1B0U, Hung *et al.*, 1998), a prototypical ABC ATPase (ATP-binding cassette). For Rad50, 243 C^{α} atoms were superimposed with a RMS deviation of 3.0 Å. MukB could be superimposed with only 112 C^{α} atoms and an RMS deviation of 2.9 Å. Histidine permease (HisP) superimposes with 192 C^{α} atoms and a RMS deviation of 3.8 Å. A structure-based sequence alignment is shown in Figure 5. The four related structures with their common core are shown side-by-side in Figure 6.

Rad50 is a member of the SMC family involved in double-strand DNA repair and the catalytic domain consists of N and C-terminal parts very similar to SMChd. The two structures are very closely related, although the overall sequence identity

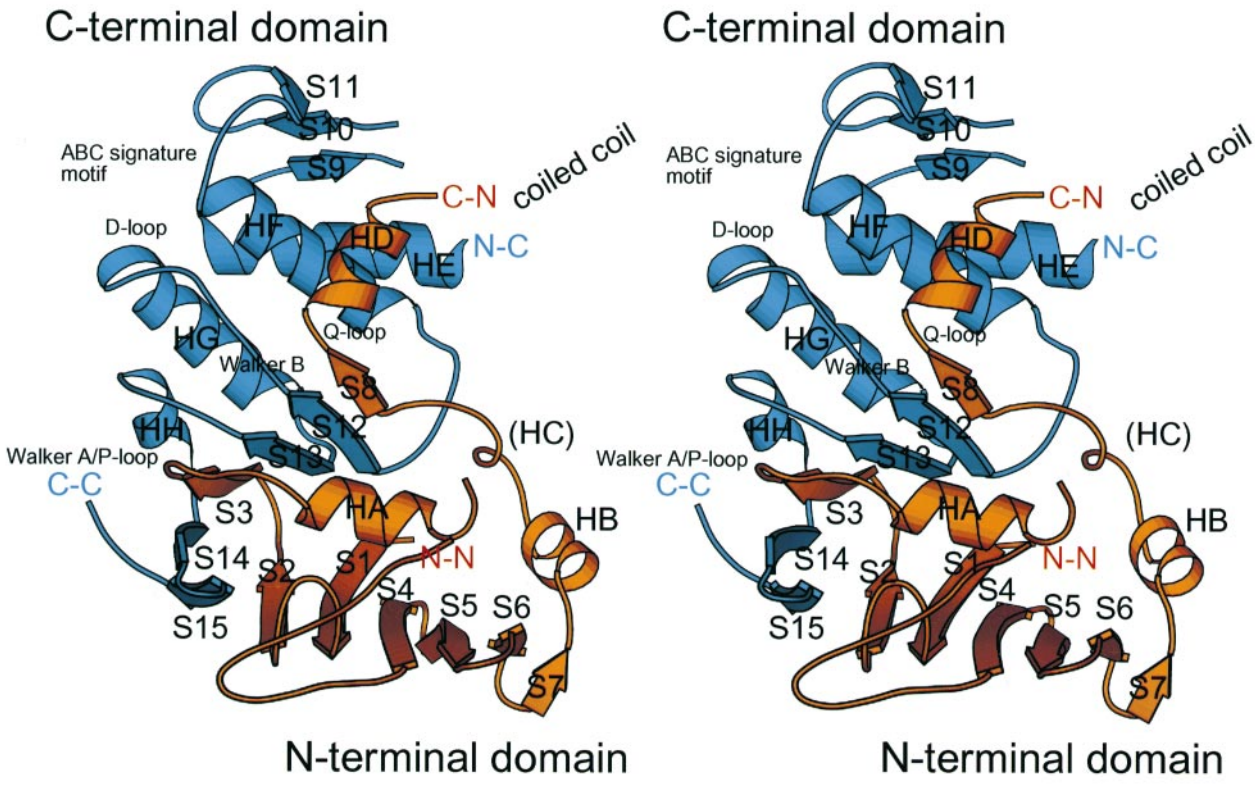

Figure 3. Ribbon representation of the SMChd crystal structure. N-N and C-N refer to the N and C termini of the N-terminal part, respectively. N-C and C-C refer to the C-terminal part accordingly. The deduced position of the coiled-coil is indicated. The N-terminal part of SMChd is colored in orange, the C-terminal part in blue. The Figure was prepared with MOLSCRIPT (Kraulis, 1991).

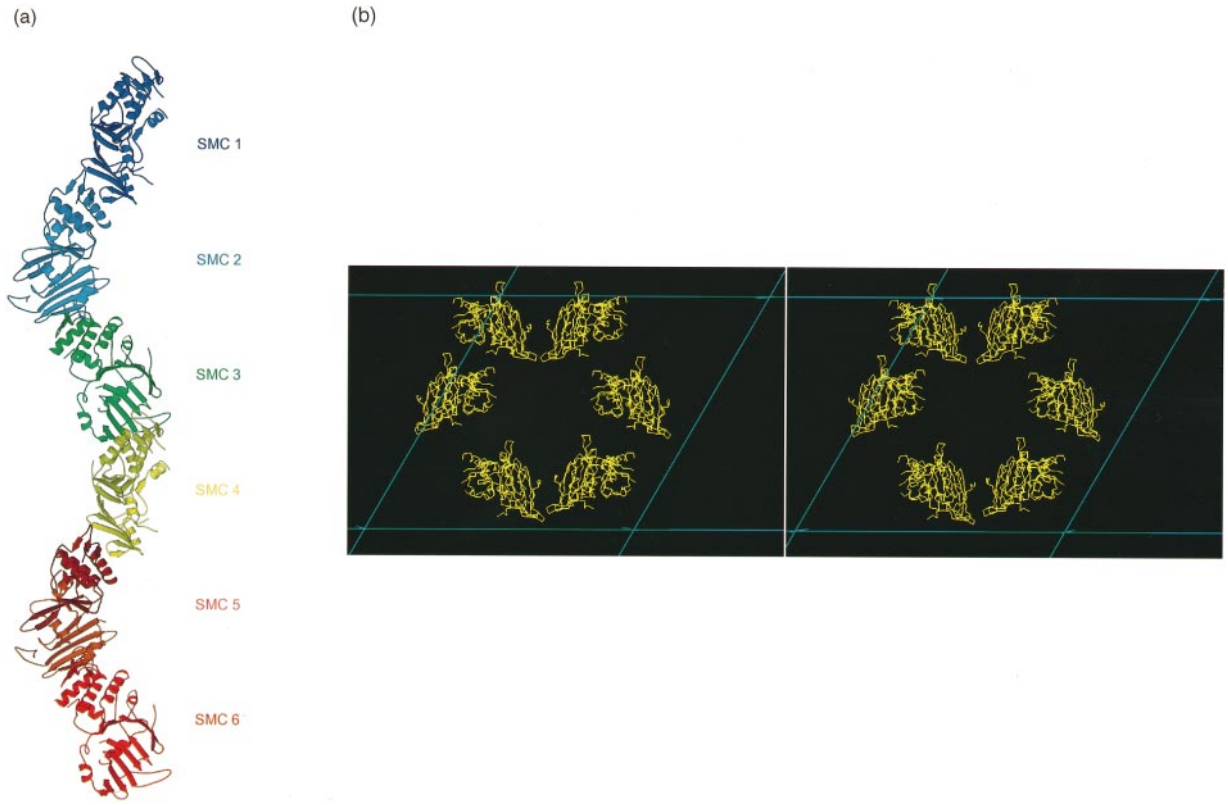

Figure 4. (a) Ribbon representation of the six non-crystallographically related molecules in the asymmetric unit of the monoclinic crystal form. Six SMChd molecules form two turns of a head-to-tail 3_1 helix (prepared with MOLSCRIPT, Kraulis, 1991). (b) Stereo representation of the six NCS-related molecules in the asymmetric unit of the trigonal space group. The same 3_1 SMChd helix as in the monoclinic crystals is formed by each of the six molecules by crystallographic $P3_1$ symmetry perpendicular to the paper plane (prepared with MAIN, Turk, 1992).

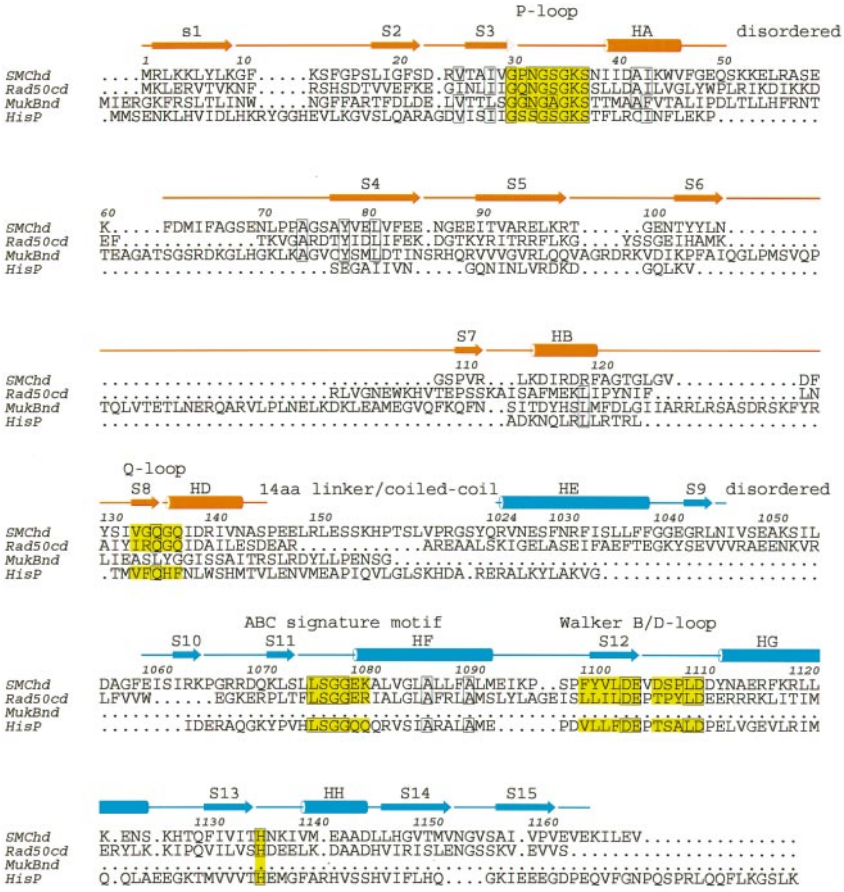

Figure 5. Structure-based sequence alignment of SMChd from *Thermotoga maritima* (PDB ID 1E69), Rad50 from *Pyrococcus furiosus* (PDB ID 1F2T chains A, B), the N-terminal domain of MukB from *Escherichia coli* (PDB ID 1QHL), and histidine permease protein HisP from *Salmonella typhimurium* (PDB ID 1B0U). A preliminary sequence alignment was prepared, adjusted using the DALI (Holm & Sander, 1995) results and checked for all six pairs manually in three dimensions. For clarity, non-overlapping regions in three dimensions have not been separated. The numbering corresponds to SMChd, as do the secondary structural elements. The Figure was prepared with ALSCRIPT (Barton, 1993).

is only 18 % based on the structure-based sequence alignment (Figure 5). SMChd has a longer loop between HA and S4 and the upper β-sheet is in a different conformation. The numbering of secondary structure elements in SMChd follows that of Rad50 (Hopfner *et al*., 2000), although S8 has not been assigned in Rad50 and β-strands from S8 onwards have different numbers. Helix HC in Rad50cd is only a helical loop in SMChd. Residues implicated in nucleotide binding in the Rad50cd dimer structure are highly conserved: Walker A (P-loop) and Walker B motifs, the Q-loop involved in nucleotide-dependent conformational changes, and the signature motif of the ABC ATPase family, which is involved in dimer formation (Figure 5). Also conserved are the D-loop and a histidine residue involved in activation of the attacking water molecule in the Rad50 dimer structure (Hopfner *et al*., 2000).

The N-terminal domain of MukB corresponds to the N-terminal part of SMChd with three major differences (Figure 6): the lower β-sheet contains a large insertion between S6 and S7, which in MukB contains two helices and a long loop that sit underneath the β-sheet. In SMChd, two strands from the N-terminal part form a β-sheet with four strands of the C-terminal part. These strands are absent from the MukB structure and corresponding sequence regions adopt different structures. Interestingly, in MukB S3 is a helix preceding the P-loop and S8 is a loop segment between two helices that probably correspond to helices HC and HD in SMChd. On the sequence level, the Q-loop is not conserved in MukB (Figure 5).

SMChd is an ABC ATPase, as can be seen from the convincing structural similarity to histidine permease nucleotide-binding domain (HisP, Figure 6, Hung *et al*., 1998). The lower β-sheet contains one strand less in HisP and in a slightly altered order but, on the sequence level, all functional motifs are highly conserved. The coiled-coil in SMChd and Rad50 extends from helices HD and HE, and these two helices are conserved in HisP. The coiled-coil domain in SMC proteins can therefore be regarded as a 900 residue insertion in the loop between these two helices.

A sequence alignment of 15 SMC proteins from eukaryotes, eubacteria, and archaea mapped onto the SMChd structure (Figure 7) shows the expected pattern of very high sequence conservation in functional motifs. This involves the P-loop, Walker B motif, Q-loop, and D-loop involved in nucleotide hydrolysis and the signature motif that has been implicated in dimer formation in Rad50. These functional motifs are all conserved between HisP and SMChd as well, highlighting the close relationship between SMC proteins and ABC ATPases.

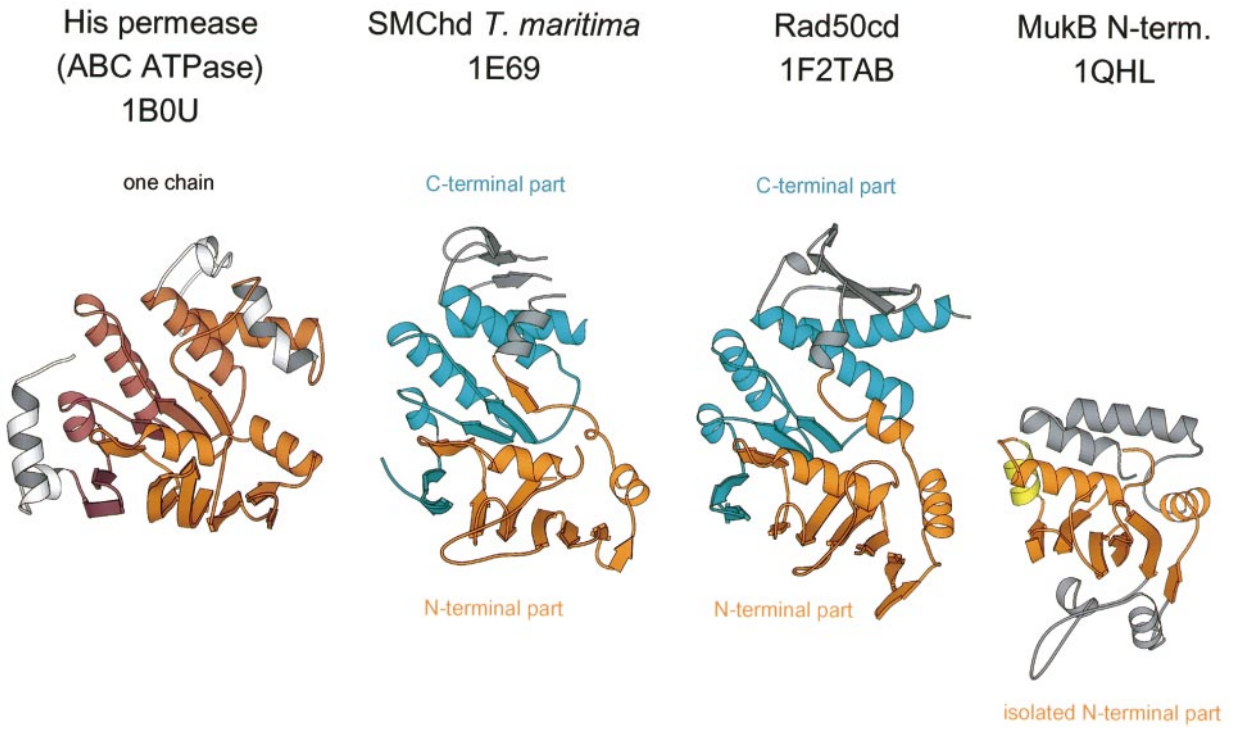

Figure 6. Structural alignment of histidine permease protein HisP from *Salmonella typhimurium* (PDB ID 1B0U), SMChd from *Thermotoga maritima* (PDB ID 1E69), Rad50 from *Pyrococcus furiosus* (PDB ID 1F2T chains A, B), and the N-terminal domain of MukB from *Escherichia coli* (PDB ID 1QHL). The structures have been aligned pairwise to SMChd in the orientation shown here and in Figures 3 and 7 using DALI (Holm & Sander, 1995). Regions not found in all structures are colored in grey. The C and N-terminal parts of SMChd, Rad50, and MukB have been colored in blue and orange, respectively. Strand S3 is a helix in MukB and this has been highlighted in yellow. The top helices in MukB may correspond to the segment HC-S8-HD in SMChd/Rad50. RMS deviations are listed in Materials and Methods. The Figure was prepared with MOLSCRIPT (Kraulis, 1991).

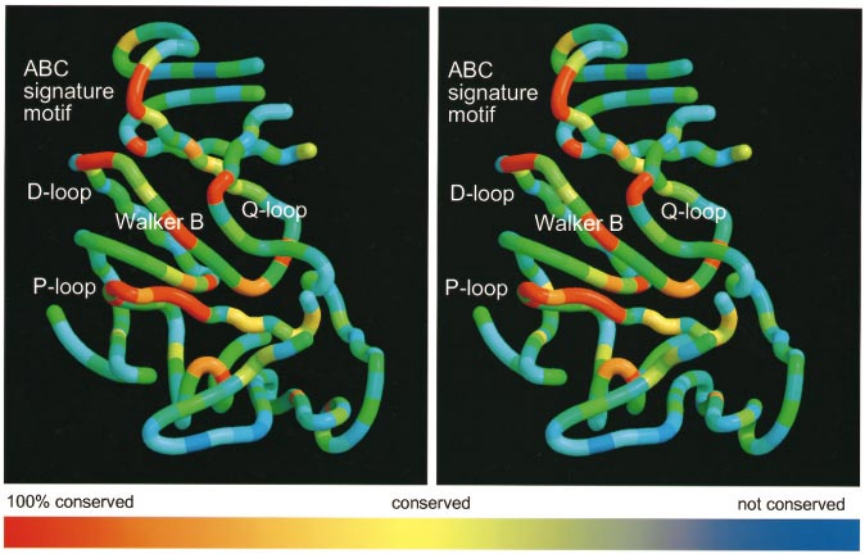

Figure 7. The 15 SMC sequences from SMC families SMC1-4 and the bacterial SMC family have been aligned using CLUSTALW (Thompson *et al.*, 1994) and the sequence conservation has been mapped onto a worm representation of SMChd. Red corresponds to 100 % conservation, blue to 0 % conservation. Loops showing very high sequence conservation correspond to the functional sequence motifs highlighted in Figure 5. The Figure was prepared with MOLSCRIPT (Kraulis, 1991).

Discussion

In this study we have produced functional SMC head domain (SMChd) from *T. maritima* by fusing the first 152 residues to the last 148 residues with a short, 14 residue, peptide linker. The construct is capable of DNA aggregation and the DNA aggregation is reversible with SDS, suggesting that a protein is involved (Figure 1). Addition of ATP does increase the efficiency of the reaction but this is not as pronounced as reported for Rad50cd (Hopfner *et al.*, 2000). Using a fusion protein is convenient because there is no danger of losing one chain. Replacing the coiled-coil in SMChd with a peptide linker connects two regions that are close together in the full-length protein. We were not able to show stable dimerization under any condition for this construct (data not shown). One reason for this is that the molecule forms high-order aggregates in low-salt buffer and the dimer may be not stable in the 200 mM sodium chloride needed to keep the protein monodisperse and folded. Neither the C-terminal His_6-tag nor the linker appear to hinder dimer formation, if compared to the Rad50 dimer structure (Hopfner *et al.*, 2000).

SMChd has been crystallized under high sodium formate conditions in space group $P3_1$, but these crystals are hemihedrally twinned and not suitable for structure determination. A monoclinic crystal form was produced and could be used for selenomethionine MAD structure determination. Although these crystals are very mosaic (1.8°) and diffract anisotropically to 3.1-3.5 Å resolution only, 30 sites could be located using Shake'n Bake (Miller *et al.*, 1994) and a clean electron density map could be obtained. The monoclinic crystals contain six molecules per asymmetric unit. The very anisotropic diffraction could immediately be explained by a helical arrangement of six SMChd subunits. Using the refined atomic model of SMChd as a search model, the structure of the twinned trigonal crystals could be solved by molecular replacement. Six molecules per asymmetric unit were positioned but in the trigonal crystals they form six non-crystallographic symmetry (NCS) related helices generated by crystal symmetry. The packing in these two crystal forms is therefore closely related and these two crystal forms appeared under many different conditions. Co-crystallization with different nucleotides in the reactions produced the same crystal forms with no nucleotide bound (data not shown). The interaction producing the helical filament of SMChd molecules is substantial (1157 $Å^2$, about 10 % of the surface of one SMChd molecule, Figure 4(a)). The small β-sheet formed by S9, S10, and S11 stacks to the large β-sheet of the N-terminal part from S2 to S7. We think it is unlikely this helical arrangement has biological relevance, since the residues involved in this interaction are among the least well conserved in the SMC family of proteins (Figure 7). However, the binding of SMC molecules to DNA is highly cooperative, indicating very close packing of the head domains on DNA; even filament formation has been suggested (Strunnikov, 1998).

The most striking feature of the SMChd crystal structure is the way the N and C-terminal parts come together to form a single globular domain. The middle β-sheet (Figure 3) contains four strands from the C-terminal part and two strands from the N-terminal part. This explains why co-expression (Rad50, Hopfner *et al.*, 2000) or fusion (SMChd) was necessary to get head domains expressed. It also questions some of the earlier experiments done with isolated N or C-terminal SMC constructs (Akhmedov *et al.*, 1998, 1999; Lockhart & Kendrick-Jones, 1998; Saleh *et al.*, 1996). Another important finding, reported with the Rad50cd crystal structure (Hopfner *et al.*, 2000), is the close relationship of SMC head domains to ABC ATPases (Figures 5 and 6). The structures superimpose very well and all loops involved in nucleotide binding and dimerization are conserved in three dimensions as well as in the primary sequences. The exceptionally long coiled-coil can therefore be regarded as an insertion between helices HD and HE of an ordinary ABC ATPase fold. The SMChd structure together with Rad50 confirm the existence of an antiparallel coiled-coil in all SMC proteins (Melby *et al.*, 1998; Saitoh *et al.*, 1995).

The N-terminal domain of MukB, a distant SMC homologue from *E. coli*, shows some surprising differences from SMChd. Most importantly, S3 is a helix in MukB (Figure 6). Looking at the sequence alignment (Figure 5), it becomes clear that this region is actually one of the most highly conserved between MukB and the ABC ATPases, and that there is no large insertion or deletion. On the basis of the alignment and the fact that the structure of the N-terminal domain of MukB was solved without the C-terminal part (van den Ent *et al.*, 1999), we propose that this helix will rearrange into a β-strand and ultimately form the middle sheet when translated as a complete head domain. A similar but less clear difference is S8, the second β-strand of the N-terminal part donated to the middle β-sheet. In MukB this region is a loop between two helices that probably correspond to helices HC and HD. This loop is the Q-loop in Rad50, containing an important residue for ATPase-dependent conformational changes (Hopfner *et al.*, 2000). The glutamate is not conserved in MukB. Taken together, the overall similarity of the N-terminal domain of MukB suggests that it is an SMC molecule capable of forming a head domain from its N and C-terminal parts. However, some remarkable conformational changes are necessary for the head domain formation in MukB, and it seems surprising that an isolated domain is stable on its own at all. Since proteins are translated from the N to the C terminus, it may be a built-in property of the N-terminal domains to be stable without the C-terminal domains, which are unstable when expressed alone (data not shown; and Andrew Lockhart, personal communication).

Comparison of SMChd to Rad50cd reveals that they are almost identical (Figure 6). This can be seen also from the structure-based sequence alignment. Although the overall sequence identity in this alignment is only 18 %, the structures superimpose very well and all residues involved in ATP binding or dimerization in Rad50cd are conserved. This raises the important question of how a molecule involved in DNA double-strand break repair (Rad50) can be so similar to a protein involved in chromosome condensation (SMChd). The answer to this question probably lies in the hinge domain, which is different between Rad50 proteins and SMC proteins. SMC proteins have larger hinge domains and are V-shaped, whereas Rad50 is a more headphone-shaped molecule and has a CXXC signature motif in the hinge domain (Sharples & Leach, 1995). Other cofactors, like cohesin or condensin subunits, are therefore most likely to bind to the hinge domain. The SMC head domain's role would then be limited to DNA binding and probably dimerization. An important finding of the Rad50 structure is dimerization-dependent ATP binding (Hopfner *et al.*, 2000). Two head domains have to come together to bring Walker A and Walker B motifs from two subunits in close proximity to form two ATP-binding sites at the interface of the two head domains. Although we do not see dimer formation for SMChd in the crystals or in gel-filtration experiments, ATP binding of SMChd will require dimerization, as Rad50 and SMChd are structurally very similar. An explanation would be the salt concentrations used in our experiments that are required to keep the protein soluble, but may interfere with dimer formation. The ATP-dependent dimer formation of Rad50, which is in line with its place in the ABC ATPase family of proteins, suggested a mechanism for SMC proteins in which two heads form a coiled-coil linked dimer (Hopfner *et al.*, 2000). The coiled-coil would ensure an increased probability for an intramolecular dimer and ATP hydrolysis could switch the molecule from dimer to the monomer state. This model requires additional factors, since SMC's work in a stoichiometric or non-catalytic way (Kimura & Hirano, 1997) and the dimer state must not switch to the monomer state on its own. No such additional factor has been isolated so far, and the Rad50 active site seems to be complete when compared to other exchange factors requiring NTPases.

A molecular understanding of SMC function would greatly benefit from visualization of DNA binding both at the molecular and the ultrastructural level.

Materials and Methods

Cloning, expression, and crystallization

SMChd is the fusion of the N-terminal 152 residues of *T. maritima* SMC (SWALL:Q9X0R4, WWW.TIGR.ORG: TM1182) with its C-terminal 148 residues. These domains are linked with a 14 residue linker peptide (ESSKHPTSLVPRGS), containing a thrombin cleavage site, which has not been used in this study. The construct has a His_6-tag attached to the C terminus (GSHHHHHH). Genomic DNA purified from living *T. maritima* cells (DSMZ, Braunschweig, DSM 3109) was PCR amplified using pfu DNA polymerase and the following primers: TGACTACCATATGAGACTGAAAAAACTCTACTTA, TGACATCACTAGTAGGGTGCTTACTAGATTCCAATCTCAGTTCTTCAGGTGAGGC for the N-terminal domain and TGACTGCACTAGTCTGGTTCCGCGTGGGTCGTACCAGAGGGTGAACGAGAGTTTC, AGTCTACGGATCCCACCTCCAGTATTTTCTCCACCTC for the C-terminal domain, resulting in DNA fragments of 499 bp and 490 bp, respectively. These fragments were digested with *Spe*I, ligated and re-amplified using the outermost primers, resulting in a single DNA fragment that was digested with *Nde*I/*Bam*HI and ligated into cleaved pHis17 (Bruno Miroux, personal communication), resulting in vector pHis17-SMChd. C41(DE3) cells (Miroux & Walker, 1996) were transformed and produced SMChd after induction with IPTG. The expressed protein contains 322 residues and has a molecular mass of 35.8 kDa. For large-scale expression, 12 l of 2 × TY medium containing 100 μg/ml ampicillin was inoculated with a 1:100 diluted overnight culture and grown until A_{600} of 0.2-0.3 at 37 °C, before induction with IPTG. Cells were harvested and frozen in liquid nitrogen. The cells were opened by sonication after the addition of lysozyme in 50 mM Tris (pH 8.0). After centrifugation, the lysate was loaded onto a 12 ml Ni^{2+}-NTA column (QIAGEN). After an extensive wash with 300 mM NaCl, 20 mM imidazole, then 50 mM and 100 mM imidazole, all in 50 mM Tris (pH 6.0), the protein was eluted with 300 mM imidazole, 50 mM Tris (pH 6.0). Peak fractions were pooled and concentrated before loading onto a Sephacryl S200 16/60 size-exclusion column (Amersham-Pharmacia), equilibrated in 20 mM Tris (pH 7.5), 200 mM NaCl, 1 mM EDTA, 1 mM sodium azide. The protein eluted as a single peak and can be stored for several months at 4 °C. Selenomethionine-substituted protein was produced as described (van den Ent *et al.*, 1999; van Duyne *et al.*, 1993) with 5 mM DTT in the final buffer for size-exclusion chromatography and storage. Electrospray mass spectrometry measurements of the non-substituted protein and the selenomethionine-containing protein was used to check selenomethionine incorporation (SMChd observed 35836.8 Da, calculated 35840.1 Da; SeMetSMChd, observed 36072.0 Da, calculated 36074.6 Da).

Monoclinic crystals were grown using sitting-drop vapor diffusion. The reservoir solution contained 0.1 M sodium citrate (pH 5.6), 0.1 M ammonium sulfate, 22 % PEG6000, and 13 μM urea. Drops were composed of 2 μl reservoir and 2 μl protein solution at 10 mg/ml and were left for one week to equilibrate. Selenomethionine-substituted protein was crystallized under the same conditions. Trigonal crystals were grown using sitting-drop vapor diffusion and reservoir solution containing 0.1 M sodium acetate (pH 5.0), 1.6 M sodium formate, and trace amounts of dimethylsulfoxide (DMSO).

Data collection, structure determination, and refinement

Trigonal SMChd crystals were frozen in liquid nitrogen using cryoprotectant containing 0.1 M sodium acetate (pH 5.0), 2 M sodium formate, 20 % (w/v) glycerol, 200 mM NaCl, and 0.004 % (v/v) DMSO. Monoclinic

Table 2. Refinement statistics

Model	N-terminal domain: 1-50, 61-146 C-terminal domain: 1024-1045, 1060-1164 Six monomers/asymmetric unit, no water molecules
Diffraction data	NATI, 3.1 Å, all data
R-factor, *R*-free[a]	0.250, 0.272
B average/bonded (Å^2)[b]	73.9, 1.5
Geometry[c]	
Bonds (Å)	0.009
Angles (deg.)	1.395
Ramachandran (%)[d]	84.0/0.0
Restrained NCS (Å)[e]	0.05
PDB ID[f]	1E69, R1E69SF

[a] 5% of reflections were randomly selected for determination of the free *R*-factor, prior to any refinement.
[b] Temperature factors averaged for all atoms and RMS deviation of temperature factors between bonded atoms.
[c] RMS deviations from ideal geometry for bond lengths and restraint angles.
[d] Percentage of residues in the most favoured region of the Ramachandran plot and percentage of outliers (PROCHECK, Laskowski *et al*., 1993).
[e] RMS deviation between coordinates of atoms related by non-crystallographic symmetry.
[f] Protein Data Bank identifiers for coordinates and structure factors, respectively.

crystals were frozen in liquid nitrogen using cryoprotectant containing 0.1 M sodium citrate (pH 5.6), 0.1 M ammonium sulfate, 22% PEG6000, 13 μM urea, 200 mM NaCl, and 15% glycerol. The trigonal crystals turned out to be hemihedrally twinned. A three-wavelength MAD experiment using monoclinic selenomethionine-substituted SMChd crystals was performed to solve the phase problem. All data were indexed and integrated with the MOSFLM package (Leslie, 1991) and further processed using the CCP4 (1994) suite of programs. The MAD experiment was perfomed on beamline ID29 at the ESRF, Grenoble, using an MAR345 imageplate detector (MARresearch, Hamburg). Because the crystals suffer from severe anisotropy, 0.5° per image was recorded. The selenomethionine monoclinic SMChd crystals have a mosaic spread of 1.6-2.0°: 19 selenium atoms could be located in dataset PEAK using Shake'n Bake 2.1 (Miller *et al*., 1994). Initial phases were calculated using MLPHARE and SOLOMON assuming six monomers per asymmetric unit (Matthews coefficient 3.6, 66% (v/v) solvent). The map showed all features of the protein and the remaining 11 sites were located in an electron density map based on anomalous Fourier coefficients. A first model was build at 3.5 Å resolution using MAIN (Turk, 1992). Crystallographic refinement was carried out using CNS (Brünger *et al*., 1998) with Engh and Huber paramaters (Engh & Huber, 1991) against a maximum-likelihood target and tight non-crystallographic restraints. The native dataset was collected on beamline 14.1 at the SRS, Daresbury. Parameters of the final model are summarized in Table 2. The trigonal crystal form was solved by detwinning the data using CNS (Yeates, 1997) (twin fraction 0.29, twin operator $h + k, -k, -l$) and molecular replacement using CNS.

DNA aggregation assay

Increasing amounts of SMChd (1 mg/ml, twofold serial dilution) in 100 mM Tris (pH 7.8), 100 mM NaCl, 10 mM magnesium acetate were incubated with or without 2 mM ATP, with or without 1% (w/v) SDS, and 0.02 mg/ml ΦX174/*Hae*III-digested plasmid DNA at 37°C for ten minutes. Samples were separated on a 1% (w/v) agarose gel in TB buffer. At high protein concentrations, monomeric DNA fragments disappear and form high molecular mass aggregates that partly stay in the loading well under the conditions used.

Protein Data Bank accession numbers

The coordinates and structure factors for the monoclinic crystal form have been deposited with the Protein Data Bank (PDB ID codes 1E69 and R1E69SF, respectively).

Acknowledgements

We thank Andy Thompson at beamline ID29, ESRF (Grenoble) for assistance with MAD data collection and Ian Fearnley (MRC Dunn Human Nutrition Unit, Cambridge) for performing mass spectrometry.

References

Akhmedov, A. T., Frei, C., TsaiPflugfelder, M., Kemper, B., Gasser, S. M. & Jessberger, R. (1998). Structural maintenance of chromosomes protein C-terminal domains bind preferentially to DNA with secondary structure. *J. Biol. Chem.* **273**, 24088-24094.

Akhmedov, A. T., Gross, B. & Jessberger, R. (1999). Mammalian SMC3 C-terminal and coiled-coil protein domains specifically bind palindromic DNA, do not block DNA ends, and prevent DNA bending. *J. Biol. Chem.* **274**, 38216-38224.

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.

Cobbe, N. & Heck, M. M. S. (2000). SMCs in the world of chromosome biology - from prokaryotes to higher eukaryotes. *J. Struct. Biol.* **129**, 123-143.

Collaborative Computing Project No 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Connelly, J. C., Kirkham, L. A. & Leach, D. R. F. (1998). The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc. Natl Acad. Sci. USA,* **95**, 7969-7974.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Chrystallog. sect. A,* **47**, 392-400.

Ghiselli, G. & Iozzo, R. V. (2000). Overexpression of bamacan/SMC3 causes transformation. *J. Biol. Chem.* **275**, 20235-20238.

Haber, J. E. (1998). The many interfaces of Mre11. *Cell,* **95**, 583-586.
Hirano, T. (1998). SMC protein complexes and higher-order chromosome dynamics. *Curr. Opin. Cell Biol.* **10**, 317-322.
Hirano, T. & Mitchison, T. J. (1994). A heterodimeric coiled-coil protein required for mitotic chromosome condensation *in-vitro*. *Cell,* **79**, 449-458.
Hirano, T., Kobayashi, R. & Hirano, M. (1997). Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a *Xenopus* homolog of the drosophila barren protein. *Cell,* **89**, 511-521.
Holm, L. & Sander, C. (1995). DALI - a network tool for protein-structure comparison. *Trends Biochem. Sci.* **20**, 478-480.
Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P. & Tainer, J. A. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell,* **101**, 789-800.
Hung, L. W., Wang, I. X. Y., Nikaido, K., Liu, P. Q., Ames, G. F. L. & Kim, S. H. (1998). Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature,* **396**, 703-707.
Jessberger, R., Frei, C. & Gasser, S. M. (1998). Chromosome dynamics: the SMC protein family. *Curr. Opin. Genet. Dev.* **8**, 254-259.
Kimura, K. & Hirano, T. (1997). ATP-Dependent positive supercoiling of DNA by 13 S condensin: a biochemical implication for chromosome condensation. *Cell,* **90**, 625-634.
Kimura, K., Rybenkov, V. V., Crisona, N. J., Hirano, T. & Cozzarelli, N. R. (1999). 13 S condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation. *Cell,* **98**, 239-248.
Koshland, D. E. & Guacci, V. (2000). Sister chromatid cohesion: the beginning of a long and beautiful relationship. *Curr. Opin. Cell Biol.* **12**, 297-301.
Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
Leslie, A. G. W. (1991). Recent changes to the MOSFLM package for processing film and image plate data. CCP4 and ESF-EACMB Newsletters on Protein Crystallography, SERC Laboratory, Daresbury, Warrington WA44AD, UK.
Lockhart, A. & Kendrick-Jones, J. (1998). Nucleotide-dependent interaction of the N-terminal domain of MukB with microtubules. *J. Struct. Biol.* **124**, 303-310.
Melby, T. E., Ciampaglio, C. N., Briscoe, G. & Erickson, H. P. (1998). The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *J. Cell Biol.* **142**, 1595-1604.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). SnB - crystal-structure determination via shake-and-bake. *J. Appl. Crystallog.* **27**, 613-621.
Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289-298.
Nasmyth, K., Peters, L. M. & Uhlmann, F. (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. *Science,* **288**, 1379-1384.
Saitoh, N., Goldberg, I. & Earnshaw, W. C. (1995). The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *Bioessays,* **17**, 759-766.
Saleh, A. Z. M., Yamanaka, K., Niki, H., Ogura, T., Yamazoe, M. & Hiraga, S. (1996). Carboxyl terminal region of the MukB protein in *Escherichia coli* is essential for DNA binding activity. *FEMS Microbiol. Letters,* **143**, 211-216.
Sawitzke, J. A. & Austin, S. (2000). Suppression of chromosome segregation defects of *Escherichia coli* muk mutants by mutations in topoisomerase I. *Proc. Natl Acad. Sci. USA,* **97**, 1671-1676.
Sharples, G. J. & Leach, D. R. F. (1995). Structural and functional similarities between the SbcCD proteins of *Escherichia coli* and the Rad50 and Mre11 (Rad32) recombination and repair proteins of yeast. *Mol. Microbiol.* **17**, 1215-1217.
Strunnikov, A. V. (1998). SMC proteins and chromosome structure. *Trends Cell Biol.* **8**, 454-459.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position- specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.
Turk, D. (1992). Weiterentwicklung eines Programms für Molekülgrafik und Elektrondichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. PhD thesis, Technische Universität München.
van den Ent, F., Lockhart, A., Kendrick-Jones, J. & Löwe, J. (1999). Crystal structure of the N-terminal domain of MukB: a protein involved in chromosome partitioning. *Struct. Fold. Des.* **7**, 1181-1187.
van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105-124.
Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945-951.
Yeates, T. O. (1997). Detecting and overcoming crystal twinning. *Methods Enzymol.* **276**, 344-358.

Edited by R. Huber

(Received 4 September 2000; received in revised form 6 December 2000; accepted 6 December 2000)