

Research

The cohesin complex: sequence homologies, interaction networks and shared motifs

Susan Jones and John Sgouros

Address: Computational Genome Analysis Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London, WC2A 3PX, UK.

Correspondence: Susan Jones. E-mail: Susan.Jones@icrf.icnet.uk

Published: 6 March 2001

Genome Biology 2001, **2(3)**:research0009.1-0009.12

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/3/research/0009>

© 2001 Jones and Sgouros, licensee BioMed Central Ltd
(Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 14 November 2000

Revised: 23 January 2001

Accepted: 24 January 2001

Abstract

Background: Cohesin is a macromolecular complex that links sister chromatids together at the metaphase plate during mitosis. The links are formed during DNA replication and destroyed during the metaphase-to-anaphase transition. In budding yeast, the 14S cohesin complex comprises at least two classes of SMC (structural maintenance of chromosomes) proteins - Smc1 and Smc3 - and two SCC (sister-chromatid cohesion) proteins - Scc1 and Scc3. The exact function of these proteins is unknown.

Results: Searches of protein sequence databases have revealed new homologs of cohesin proteins. In mouse, Mmip1 (Mad member interacting protein 1) and Smc3 share 99% sequence identity and are products of the same gene. A phylogenetic tree of SMC homologs reveals five families: Smc1, Smc2, Smc3, Smc4 and an ancestral family that includes the sequences from the Archaea and Eubacteria. This ancestral family also includes sequences from eukaryotes. A cohesion interaction network, comprising 17 proteins, has been constructed using two proteomic databases. Genes encoding six proteins in the cohesion network share a common upstream region that includes the Mlul cell-cycle box (MCB) element. Pairs of the proteins in this network share common sequence motifs that could represent common structural features such as binding sites. Scc2 shares a motif with Chk1 (kinase checkpoint protein), that comprises part of the serine/threonine protein kinase motif, including the active-site residue.

Conclusions: We have combined genomic and proteomic data into a comprehensive network of information to reach a better understanding of the function of the cohesin complex. We have identified new SMC homologs, created a new SMC phylogeny and identified shared DNA and protein motifs. The potential for Scc2 to function as a kinase - a hypothesis that needs to be verified experimentally - could provide further evidence for the regulation of sister-chromatid cohesion by phosphorylation mechanisms, which are currently poorly understood.

Background

Cohesin is a macromolecular complex that holds sister chromatids together at the metaphase plate during mitosis. The links between the sister chromatids are formed during DNA replication and destroyed during the metaphase to anaphase

transition, when sister chromatids separate to opposite poles of the cell. In budding yeast, the 14S cohesin complex comprises at least two SMC (structural maintenance of chromosomes) proteins - Smc1 [1] and Smc3 [2] - and two SCC (sister-chromatid cohesion) proteins - Scc1 [3] and Scc3 [2].

A recent development is the identification of a separate complex, comprising two further sister-chromatid cohesion proteins, Scc2 and Scc4, that function in the loading of cohesin macromolecules onto chromosomes [4].

The Smc1 and Smc3 proteins belong to the conserved and well characterized SMC family, which also includes Smc2 and Smc4, components of the condensin macromolecular complex. The SMCs have a highly conserved structure comprising five domains arranged in a head-rod-tail architecture, including a Walker A motif in the amino-terminal domain and a DA-box (Walker B motif) in the carboxy-terminal domain (Figure 1a) [5-7]. Dimeric models of Smc1-Smc3 protein complexes have been proposed, in which the coiled-coil domains of each protomer interact in an antiparallel arrangement, bringing the Walker A and B motifs together at the termini of the structure, forming two complete ATP-binding sites (Figure 1b) [7-11]. In accordance with this model, an SMC homodimer has been observed by electron microscopy in *Bacillus subtilis* [8]. A similar model is proposed for Smc1-Smc3 heterodimers in eukaryotes [7].

A number of additional proteins are known to play key roles in the cohesion mechanism. Eco1 is involved in the establishment of cohesion during S phase of the cell cycle, but not for its maintenance during G2 or M phases [10,12]. Esp1, a separin protein, is a protease that cleaves Scc1 at the metaphase-to-anaphase transition to trigger sister chromatid separation [13]. This protein is complexed with the

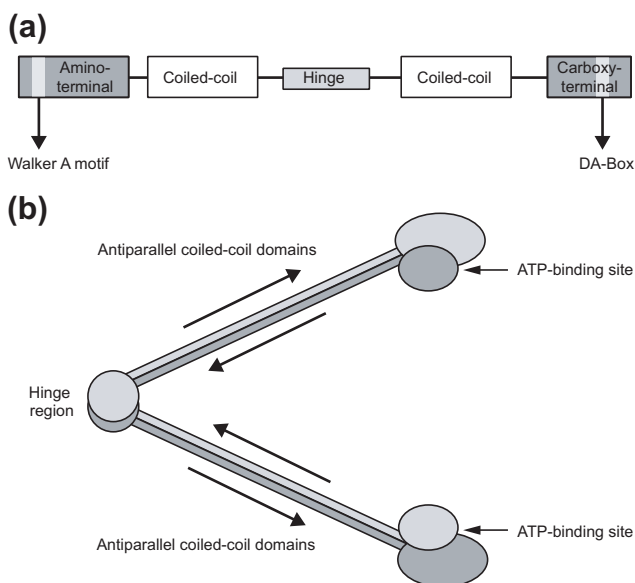


Figure 1
Structure of an SMC protein. **(a)** The five domains of SMC proteins. The amino-terminal domain includes a Walker A motif and the carboxy-terminal domain a DA-box (also known as a Walker B motif). **(b)** Proposed dimeric interaction of SMC molecules (see, for example, [8]).

securin protein Pds1 for some of the cell cycle [14], which prevents the onset of anaphase when there has been DNA or spindle damage during DNA replication. When Esp1 is separated from Pds1, it undertakes the proteolytic cleavage of Scc1 (for review see [15]).

Here, we have combined available genomic and proteomic data into a comprehensive network of information to reach a better understanding of the function of the cohesion complex. We have searched for homologs of the SMC proteins, created a new evolutionary tree for these proteins and identified an interesting homology between SMC3 and Mmip1 in mouse. We have also created a cohesion interaction network of 17 proteins using two proteomic databases. A number of pairs of proteins within this network share sequence motifs that could represent common binding sites. In addition, the genes encoding a subset of six proteins in the network share a common upstream regulatory element.

Results

Homologs of cohesin proteins

SMC proteins

A PSI-BLAST search for sequence homologs of SMC1 and SMC3 from *Saccharomyces cerevisiae* revealed homologs from many species of eukaryotes, archaea and eubacteria as previously reported [8,16-18] (Table 1). These homology searches provided the basis for a phylogenetic tree and for the analysis of new sequence homologs.

The SMC phylogenetic tree created from the alignment of SMC3 homologs (Figure 2) reveals five families: Smc1-Smc4 from eukaryotes and a fifth 'ancestral' family that includes the SMCs from eubacteria and archaea. This ancestral family also includes a number of eukaryotic proteins from *S. cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster* and humans. Each of these eukaryotes has SMC proteins from all five families. The eukaryotic proteins within the ancestral family include the Rad18 from *S. pombe* and Rhc18, the Rad18 homolog in *S. cerevisiae*. Rad18 in *S. pombe* is involved in the repair of DNA damaged by UV radiation [19]. The sequences from *C. elegans*, *Drosophila* and human that cluster with Rad18 within the ancestral family are likely to be Rad18 homologs. Also clustered within this group is Spr18, an SMC protein proposed to be the homodimeric partner of rad18 in *S. pombe* [20]. In addition, MukB from *Escherichia coli* also lies within this ancestral family. MukB is known to be essential for chromosome partitioning in this species [21-23]. The clustering of the Rad18 homologs with the ancestral SMC proteins is not observed in the phylogenetic tree constructed by Cobbe and Heck [18].

One unusual sequence homolog of SMC3 in mouse (SMCD) [6] has already been reported in the form of bamacan, a chondroitin sulphate proteoglycan [24,25]. This protein is

Table 1

Proteins involved in sister chromatid cohesion in *S. cerevisiae*, and their sequence homologs

Organism	SMC1 proteins/homologs	SMC3 proteins/homologs	Non-SMC proteins/homologs		Loading factors/homologs	
<i>S. cerevisiae</i>	Smc1p	Smc3	Scclp/Mcd1p	Scclp/IRRI	Scclp	Scclp
<i>S. pombe</i>	Psm1/SPBC29A10.04	Psm3/SPAC10F6.09C	Rad21	SPAC17H9.20 Rec11	Mis4	?
<i>C. elegans</i>	F28B3.7	Y47D3A.26	F10G7.4	F18e2.3	?	?
<i>Drosophila</i>	CG6057	DCAP	dRAD21/DMsccl	Stromalin (SA)	Nipped-B	?
<i>Xenopus</i>	XSMC1	XSMC3	XRAD21	?	?	?
Mouse	mSMCB	MSMCD	mHR21SPA	SA1 SA2 SA3	?	?
Bovine	bSMC1	BSMC3	?	?	?	?
Human	hSMC1/Sb1.8	HSMC3/HCAP	hR21PSA hHR21spB	SAG1 SAG2 SAG3	IDN3-B	?
<i>Arabidopsis</i>	?	CAP/AT2G27170	Syn1 /Dif1	SA F17A22.37	?	?

The question marks indicate where no sequence homologs could be found.

known to have 100% sequence identity to SMCD [18]. Here we identify another new homolog, Mmip1, that also shares an extremely high sequence identity with mouse SMCD. Mmip1 (Mad interacting protein 1) was identified from a yeast two-hybrid screen for proteins that bind Mxi, a basic helix-loop-helix (bHLH) transcription factor [26]. Mmip1 is a basic helix-loop-helix zipper (bHLH-ZIP) protein that strongly dimerizes with Mad1, Mxi, Mad3 and Mad4, but not with Max or c-Myc [26]. A Clustal X alignment of Mmip1 with SMCD reveals that Mmip1 lacks the first globular domain and the first coiled-coil domain common to SMC proteins. In the alignment there is 40% sequence identity between Mmip1 and SMCD over the entire length of the SMCD (1,217 amino acids). Over the length of the Mmip1 protein (485 amino acids), however, the protein shares 99% sequence identity with SMCD. These high percentage sequence identities are also reflected in the DNA sequences that encode these proteins. The cDNA encoding the Mmip1 protein is 100% identical to the cDNA encoding SMCD over the 2,612 base pairs of the Mmip1 sequence.

It has previously been suggested that eubacteria contain a single ancestral SMC protein [27]. The PSI-BLAST search for SMC homologs in the current work identified two SMC-related proteins in two species of eubacteria, *B. subtilis* and *Aquifex aeolicus*. In both species one sequence has previously been identified as an SMC homolog, whereas the function of the second is unknown. The two sequences from *B. subtilis* share 95% sequence identity, whereas the two sequences from *A. aeolicus* share 20% sequence identity. All four homologs contain a Walker A and B motif, and the two homologs from *B. subtilis* contain the five domains

characteristic of the SMC proteins (Figure 1a). The *A. aeolicus* protein known to be an SMC homolog (TrEMBL accession number O60878) also contains the five domains, including the two coiled-coil domains separated by a hinge region of 180-200 residues. However, the second homolog in *A. aeolicus* (TrEMBL accession number O67124) has the two coiled-coil domains (predicted using Coils [28]) but the hinge region separating them consists of only approximately 10-20 residues. In the current model of SMC dimers the hinge region allows the folding of the structure into an approximately symmetrical complex (Figure 1b). For this *A. aeolicus* homolog, however, the very short hinge region would restrict the range of folding. In this species, two homodimeric SMC structures could be formed, one from the five-domain SMC and one from the four-domain SMC homolog lacking the hinge domain. The presence of two potential SMC homologs in *B. subtilis*, however, could mean that the heterodimeric model of SMC interactions proposed for eukaryotes (for example, [10]) could also be extended to some prokaryotes. The presence of two SMC homologs in some eubacteria is not shown in the SMC phylogenetic tree constructed by Cobbe and Heck [18].

SCC proteins

The SCC proteins are only present in eukaryotes and are not as well characterized as the SMC proteins. Sccl (also identified as MCD1) is physically associated with the SMC1 protomer in the complex [3]. Homologs in *S. pombe* [29], *Xenopus laevis* [30], humans [31] and *Drosophila* are identified as Rad21 proteins (Table 1), involved in the repair of DNA double-stranded breaks induced by ionizing radiation. Sccl (previously identified as IRR1 [10]) contains a nuclear

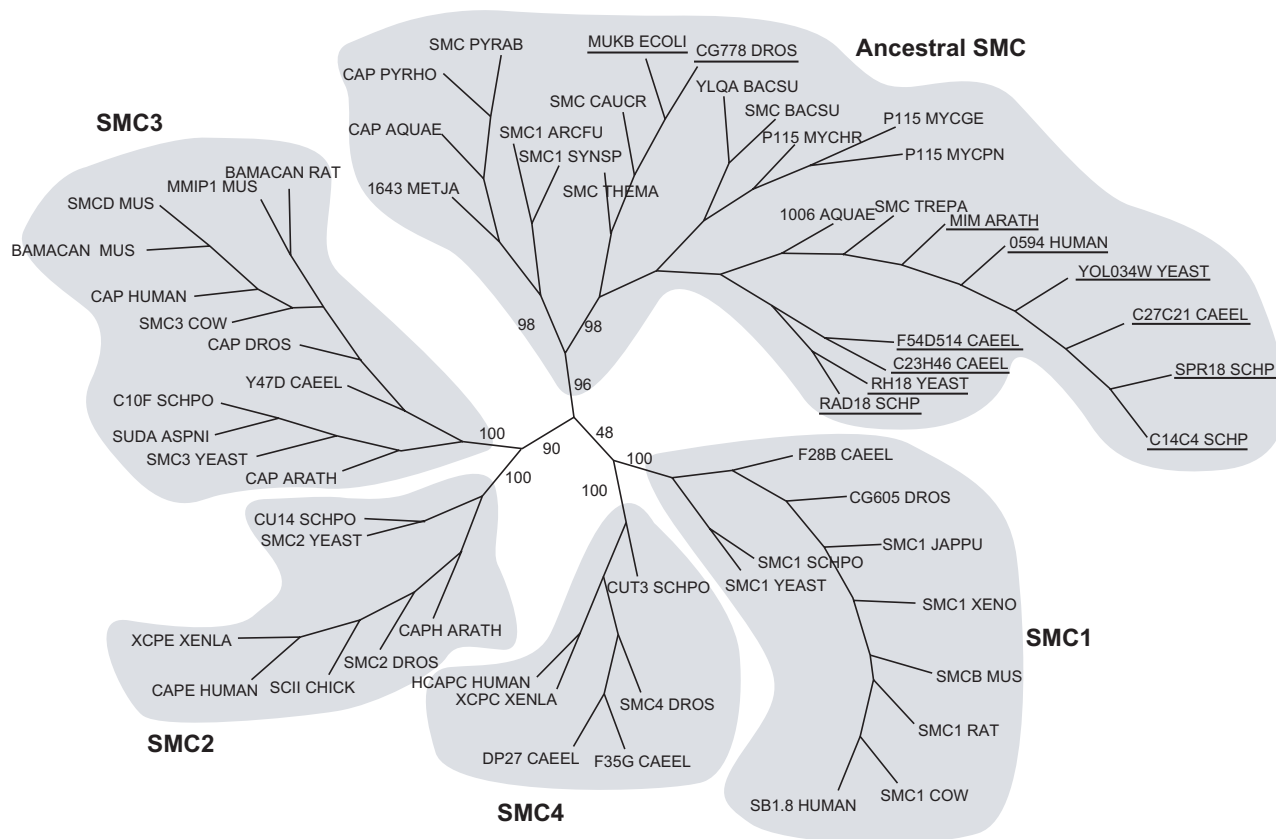


Figure 2

Evolutionary tree for SMC proteins, created using PHYLIP [69,70]. Each of the five SMC families is highlighted and labeled. The names of the eukaryotic proteins present in the ancestral family are underlined. Bootstrap values from 100 bootstrap trials are shown on the primary branches of the tree. AQUAE, *Aquifex aeolicus*; ARATH, *Arabidopsis thaliana*; ARCFU, *Archaeoglobus fulgidus*; ASPN, *Aspergillus niger*; BACSU, *Bacillus subtilis*; CAEEL, *Caenorhabditis elegans*; CAUCR, *Caulobacter crescentus*; DROS, *Drosophila*; ECOLI, *Escherichia coli*; JAPPU, Japanese pufferfish; METJA, *Methanococcus jannaschii*; MUS, mouse; MYCGE, *Mycoplasma genitalium*; MYCHR, *Mycoplasma hyorhinis*; MYCPN, *Mycoplasma pneumoniae*; PYRAB, *Pyrococcus abyssi*; PYRHO, *Pyrococcus horikoshii*; SCHP, *Schizosaccharomyces pombe*; SYNSP, *Synechocystis* sp.; THEMA, *Thermotoga maritima*; TREPA, *Treponema pallidum*; XENLA, XENO, *Xenopus laevis*; YEAST, *Saccharomyces cerevisiae*.

localization sequence (see later) and a number of homologs have been identified (Table 1). Scc3 homologs in *Drosophila*, mouse, human and *Arabidopsis* are a family of stromalin proteins [32,33] which share between 20-25% sequence identity (Table 1). In *Drosophila*, mouse and human there are two stromalin proteins (dSA, dSA2; SA1, SA2; and STAG1, STAG2, respectively), which are located in the nucleus, but their function is unknown. In addition, STAG3 has been identified in humans [33] and is proposed to be involved in chromosome pairing during meiosis.

Scc2 and Scc4 are the recently identified cohesin loading factors [4]. Homologs to Scc2 have been identified in *S. pombe* (Mis4 [34]) and *Drosophila* (Nipped-B [35]), *Coprinus cinereus* (Rad9 [36]) and human (IDN3-B; TrEMBL accession number Q9Y6Y3) (Table 1). Mis4 in *S. pombe* is required for equal chromatid separation in anaphase and has

a function distinct from cohesin [34]. The Rad9 gene product in *C. cinereus* is essential for the normal completion of meiosis. The Nipped-B gene product is proposed to function architecturally between transcription enhancers and promoters to facilitate enhancer-promoter interactions [35]. The function of the *IDN3-B* gene in humans is unknown, other than it is preferentially expressed in hepatocellular carcinomas (HCC) [37]. It has been proposed that these SCC molecules represent a family of 'adherins' that share a large central core domain of sequence homology [34].

Scc4 was identified as a product of open reading frame (ORF) YER147C [4], and comprises a sequence of 624 amino acids that includes an AMP-binding motif. However, other than interacting with Scc2 and being involved in the establishment of sister-chromatid cohesion, little is known about this protein. Scc4 has no identifiable sequence homologs in

Table 2

Three cohesin-interacting proteins from *S. cerevisiae* and their sequence homologs

Organism	Cohesin-interacting proteins			
<i>S. cerevisiae</i>	Trf4	Prp11	Tid3p/NCD80	Esp1p
<i>S. pombe</i>	SPAC12g12.13c	SPBC21C3.05	SPBC11C11.03	Cut1
<i>C. elegans</i>	ZK858.1	F11a10.2	W01B6.9	Y47G6A.12 ZK430
<i>Drosophila</i>	EG:8D8.8	CG10754	?	CG10583
Mouse	ESTs	SAP62	ESTs	ESTs
Human	hTRF4-1 hTRF4-2	SAP62	Hec1	KIAA0165
<i>Arabidopsis</i>	F17k2.15	T26B15.16	T14E10.200	F7H19.150

The question marks indicate where no sequence homologs could be found.

either the full-sequence or EST databases, and therefore could be the product of an orphan gene.

Cohesion interaction network

A cohesion interaction network was created by collating information from two proteome databases and the literature (Figure 3). In Figure 3, lines are drawn between proteins to indicate known or potential interactions. The data from which the interactions are derived are indicated in a detailed key that differentiates between the two proteomic databases (and between the different sources of data within each database) and the literature. Four proteins (Esp1, Trf4, Prp11 and Tid3) interact directly with SMC or SCC proteins in *S. cerevisiae*. The interaction of Esp1 and Scc1 is currently known at a functional level [13], and its importance has already been discussed. This interaction is time-dependent and has not been identified in the yeast two-hybrid screen, and this information is not currently recorded in the YPD.

Trf4 is a protein involved in both mitotic chromosome condensation [38] and sister-chromatid cohesion [39]. In *X. laevis* Trf4 interacts with Smc1 and Smc2 [38], and in *S. cerevisiae* Trp4 interacts with Smc1 and Trf5 [40], another member of the TRF family. Trf4 homologs have been identified in *S. pombe*, *C. elegans*, *Drosophila*, human and *Arabidopsis* (Table 2). Trf4 has very recently been identified as a DNA polymerase with β -polymerase-like properties and is now designated DNA polymerase κ (the fourth class of nuclear DNA polymerases) [41]. Remote homologs of *S. cerevisiae* Trf4 include the caffeine-induced cell death protein I (Cid1) in *S. pombe* [42] (13.4% sequence identity) and the polynucleotide adenylyltransferase enzyme from a number of organisms including *S. pombe* and humans (10.2% and 9.7% sequence identity respectively). Cid1 is of particular interest as it thought to play a part in the S-M

checkpoint pathway in *S. pombe* [42]. As a homolog of Trf4, Cid1 could be the link between sister-chromatid cohesion and this checkpoint pathway.

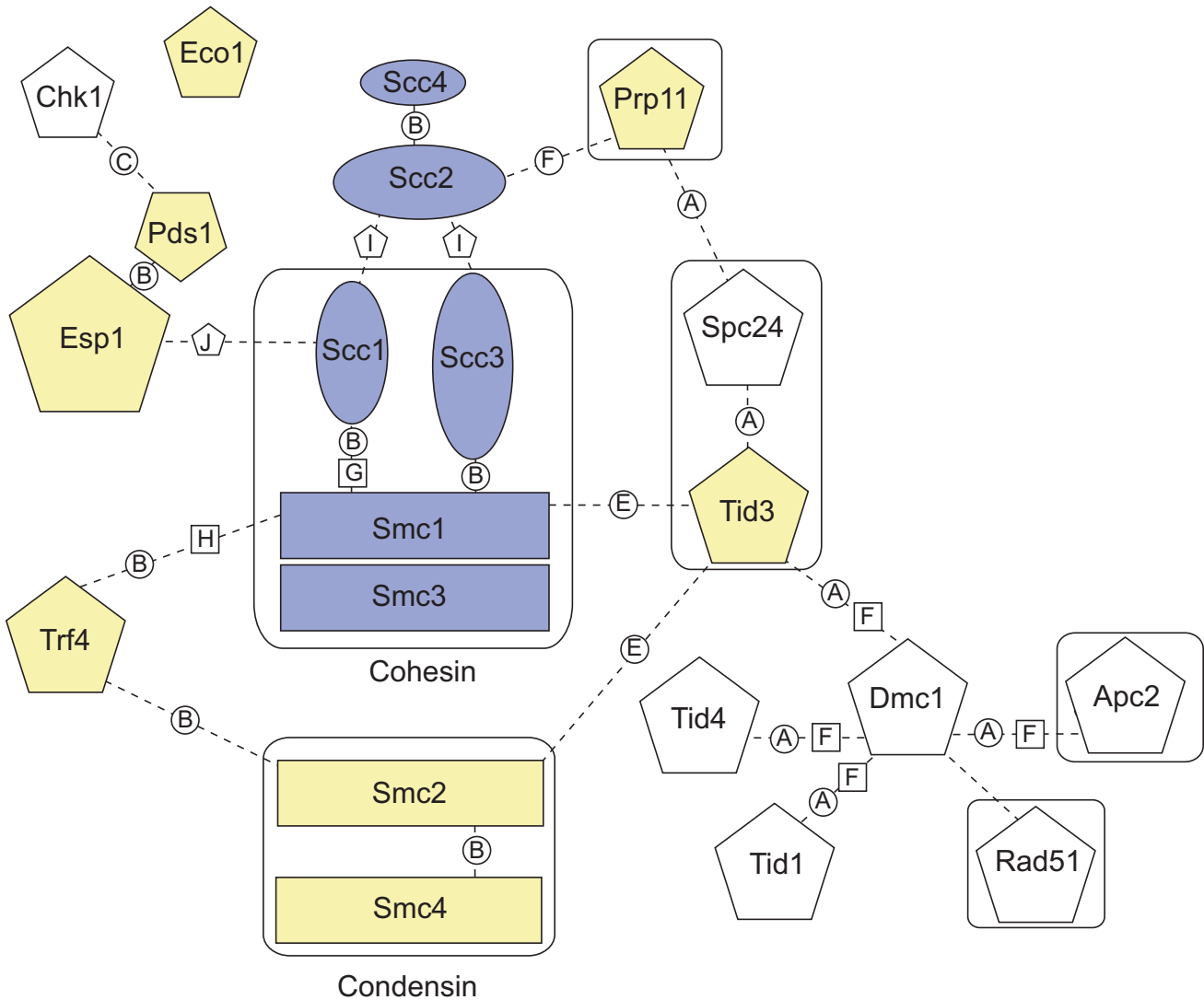
Prp11 is a yeast splicing factor involved in the early stages of the spliceosomal assembly pathway [43]. Prp11 is a 266 amino-acid protein that includes a zinc-finger domain common to RNA-binding proteins [44]. This splicing factor forms a complex with two others, Prp9 and Prp21, which together with Prp5 are required for the binding of U2 snRNP to pre-mRNA [45,46]. There are homologs of this splicing factor in *S. pombe*, *C. elegans*, *Drosophila*, *Arabidopsis*, mouse and human (Table 2) and all include the RNA-binding motif. In mouse and humans, the homolog is SAP62 (spliceosome-associated protein), a spliceosomal protein that binds to pre-mRNA in the prespliceosomal complex [47].

Tid3 (NCD80) is a spindle pole body protein that has homologs in a number of eukaryotes (Table 2). Tid3 is predicted to interact with Smc1 and Smc2, and has been shown experimentally to interact with Spc24, another component of the spindle pole body. Interactions between the human homolog of Tid3, Hec1, and human Smc1 and Smc2 homologs have also been observed [48]. The interactions of Tid3 with subunits from both the cohesin and condensin

Table 3

PROSITE sequence motifs located using ProfileScan in proteins involved in the sister chromatid cohesion mechanism

Protein class	<i>S. cerevisiae</i> proteins	PROSITE motifs
Cohesin subunits	Smc1	NLS_BP: Bipartite nuclear localization signal ATP_GTP: (P loop) (Walker A motif) DA_BOX: Walker B motif Leucine zipper
	Smc3	NLS_BP: Bipartite nuclear localization signal ATP_GTP: (P loop) (Walker A motif) DA_BOX: Walker B motif Leucine zipper
	Scc1	None
	Scc3	NLS_BP: Bipartite nuclear localization signal
Loading factors	Scc2	Leucine zipper
	Scc4	AMP binding domain
SMC1-interacting protein	Trf4	NLS_BP: Bipartite nuclear localization signal PAP: poly(A) polymerase PAP_associated: poly(A) polymerase associated PAP_core: poly(A) polymerase core
SMC1-interacting	Tid3/NCD80	None
Scc1-interacting	Esp1	PROKAR_LIPOPROTEIN prokaryotic membrane lipoprotein attachment site NLS_BP: Bipartite nuclear localization signal
SCC2-interacting	Prp11	ZF_MATRIN domain: RNA-binding zinc finger



Key to interactions		
YPD	CuraGen	Literature
(A) Yeast two-hybrid	(F) Yeast two-hybrid	(I) [4]
(B) Co-immunoprecipitation	(G) Co-immunoprecipitation	(J) [13]
(C) Affinity column	(H) Genetic	
(D) Co-sedimentation		
(E) Prediction		

Figure 3

The cohesion interaction network. Lines connecting proteins indicate known or potential interactions as derived from two proteomic databases and the literature. Cohesin and the loading factors are in yellow; additional proteins involved in cohesion or interacting with cohesin or the loading factors are in blue; all other proteins in the network are in white. Proteins outlined with boxes are part of macromolecular complexes. Prp11 is part of a complex in the spliceosomal pathway, and Apc2 is part of the anaphase-promoting complex (APC). Tid3p and Spc24 are both part of the spindle-pole body. Solid black lines indicate proteins that form dimeric interactions. The cohesion network of 17 proteins includes all those labeled, excluding Apc2, Tid4, Tid1 and Rad51.

macromolecules, places it alongside Trf4 and Scc1, as a protein integrally involved in both mechanisms. It is also proposed that Hec1 may be involved in chromatin assembly in the centromere and regulation of the kinetochore [48]. Spc24, one interaction partner of Tid3, also interacts with Prp11, the yeast splicing factor which is linked to the cohesin loading factors through its interaction with Scc2 (Figure 3).

A common upstream DNA element

The upstream regions of the genes encoding 17 proteins in the cohesion network (Figure 3) were searched for shared motifs using AlignACE. Three consensus motifs were identified that were common to subsets of the 17 genes. Only one motif was found to be relatively specific, however, matching upstream sequences of only 29 genes in the SGD [49,50] (see the Materials and methods). This motif has the consensus sequence A6[X]ACGCGTH2[X]RXAAX and includes the MluI cell-cycle box (MCB) element (consensus sequence ACGCGT) [51,52]. The extended consensus motif found in the current work was present in upstream regions of the genes encoding Scc1, Scc3, Smc3, Pds1, Eco1 and Spc24. This motif was located between 123-299 base pairs (bp) upstream of the genes encoding these six proteins. A search of the SGD revealed 23 additional genes containing this upstream motif. Eight of these additional genes encoded hypothetical proteins of unknown function. However, these additional genes also included those encoding chaperones (JEM1 and PDI1n), transcription factor components (TFA1, RFA2, RNA polymerase II, SPT20 and PRT1), and a YC component of the proteasome. When the search was extended to 2,000 bp upstream of the 5' untranslated regions of the yeast genome, the gene encoding Trf4 was also found to contain this consensus motif (1,560 bp upstream).

Shared motifs within the cohesion interaction network

Teiresias, a pattern discovery algorithm [53,54], was used to search for common motifs between two or more sequences in the 17 proteins of the cohesion network. The highest number of proteins sharing a common motif was three, and these were the three SMC proteins, which have a high sequence identity and share known Prosite motifs (Table 3). More interesting were 24 pattern matches found between pairs of proteins in the network. A number of proteins share more than one sequence motif with the same protein. All shared motifs were either specific to the two proteins in the cohesion network, or in the case of three motifs, shared by one other protein sequence.

One motif shared by two sequences in the network and one additional sequence, is the DXXPENIXLXKN motif shared by the sequences of Scc2, Chk1 and a third *S. cerevisiae* protein PKH1 (yeast ORF YDR490C) (Figure 4). Both Chk1 and PKH1 are serine/threonine (ST) protein kinases, and the motif they share with Scc2 includes part of the PROSITE ST kinase signature motif ([LIVMFYC]X[HY]XD[LIVMFY]KXXN[LIVMFYCT](3), where X indicates any residue, (3) indicates that

```

SCC2_YEAST  459  YSFCNQMKDIAPENIDLLKNEYKQEEFL
CHK1_YEAST  134  VECGVAHRDIKPENILLDKNGNLKLADFG
PKH1_YEAST  241  HNIGLIHRDIKPENILLDKNMKVKLTDFG

ST_KINASE  [LIVMFYC]X[HY]XD[LIVMFY]KXXN[LIVMFYCT](3)
    
```

Figure 4

Sequence alignment of the conserved motif in Scc2, Chk1 and Pkh1, which includes the PROSITE serine/threonine (S/T) protein kinase motif. In the alignment the conserved residues of the motif identified using Teiresias are in red and additional conserved positions are in green. The residues that coincide with the S/T kinase motif are outlined with a box. The number before each motif indicates the position of the first residue within the complete sequence. The PROSITE S/T kinase motif is shown beneath the alignment. The alternative residues are shown in squared brackets; X denotes any residue; the active-site aspartic acid is in blue.

the previous residue is repeated three times, and D is the active site residue). The sequence of Scc2 does not match the ST kinase signature motif exactly. Of the 13 residues in the ST kinase motif, Scc2 has four mismatches but, importantly, the active-site aspartic acid is conserved.

A second motif shared by a third protein not included in the cohesion network was SXXSXLKKKXLXT; this is found in Scc1, Scc2 and yeast ORF YHR011W, a putative seryl-tRNA synthetase (Figure 5a). However, this motif was not part of the tRNA ligase motif of YHR011W, or of any other known motif within this sequence. A third motif shared by a protein from outside the cohesion network was NDXNXDDXDN,

(a)	SCC2_YEAST	1177	V N STLS S NS I LKKK L LK T NRVE
	TRF4_YEAST	251	R N NLY S LA S HLKK K NLA T EV E
	YHH1_YEAST	104	Y N DQ N S K S S ELKK K I L ET C K S L
(b)	SCC1_YEAST	248	G I TEK N D Q N D D D N S V E Q G
	SMC1_YEAST	67	Y R GV L N D S N DD D Y D NEGA A S
	PATP_PLAYO	1342	D N DD D N D D N DD D N D N D N
(c)	SCC4_YEAST	597	T V AM R G K D V AL T NA K LEAL V
	SMC3_YEAST	144	N I VP Q G K I V AL T NA K DKER L

Figure 5

Sequence alignments for three motifs shared by proteins in the cohesion network. (a) A motif shared by Scc2 and Trf4 in the network and a putative seryl-tRNA synthetase (YHH1) from yeast. (b) A motif shared by Scc1, Smc1 and a P-type ATPase from *Plasmodium yoelii*. (c) A motif shared by the cohesin loading factor Scc4 and SMC3. In each alignment the conserved residues of the motif identified using Teiresias are in red and additional conserved positions are in green. The number before each motif indicates the position of the first residue within the complete sequence.

shared by Scc1, Smc1, and a P-type ATPase from *Plasmodium yoelii* (Figure 5b). Scc4 is one of the cohesin loading factors for which no known homolog has been found. This protein was, however, found to share a 10-residue sequence motif (GKXVALTNAK) with Smc3 (Figure 5c).

The securin Pds1 is an anaphase inhibitor that contains a destruction box motif (RXXXLXXXXN) [55], which targets this protein for destruction by the APC ubiquitin ligase. We found three destruction box motifs in Smc3, one in the hinge region (at position 682, RTRLESLKN) and two in the second coiled-coil domain (one at position 744 (RTSLNTKKN) and one at position 920 (RLLKLDN)). We also found a KEN-box motif (an additional APC recognition signal [56]) in SMC2 at position 304 (KENGLLN), in the first coiled-coil domain.

Discussion

The mechanism of cohesion is essential to the successful completion of the cell cycle. Cohesin is the macromolecular complex, comprising SMC and SCC proteins, that plays a central role in this mechanism.

SMC3 and Mmip1

In the mouse, the sequence of SMC3 (SMCD) is 100% identical to two proteins with distinctly different functions, bamacan and Mmip1. The possibility that SMCs can 'moonlight' outside the cell in the basement membrane in the guise of the proteoglycan bamacan has recently been proposed [18]. Proteoglycans are glycoproteins present in connective tissues and on the cell surface, and modified proteoglycans are found in human tumors of epithelial origin (for example breast, colon and lung). In the current work we found that Mmip1 was 100% identical to the hinge, second coiled-coil and carboxy-terminal domain of SMCD. It is possible that Mmip1 is a protein resulting from alternative gene splicing or post-transcriptional modification. Mmip1 is thought to compete with the dimerization of Max and Mad family proteins, indirectly upregulating the transcriptional function of c-Myc, which plays a central role in cell proliferation, differentiation and apoptosis [26]. Hence, a single gene appears to encode three products, each with a different function.

Ancestral SMC sequences

The evolutionary tree presented here identifies five SMC families: Smc1, Smc2, Smc3, Smc4 and an ancestral SMC family that includes the SMC sequences from the Eubacteria and the Archaea. However, some eukaryotic SMC sequences, namely Rad18 and its homologs, are also clustered within the ancestral SMC family. Humans, *Drosophila*, *C. elegans*, *S. cerevisiae* and *S. pombe* have SMCs from all five families. A number of these eukaryotes have more than one sequence clustered in the ancestral family, which leads us to suggest that these proteins might form heterodimeric structures, similar to those observed in other SMC proteins (see, for example [8]). This evolutionary tree is similar in some

respects to that recently reported by Cobbe and Heck [18], in that five families are identified. However, in our tree the Rad18 homologs are more closely related to the ancestral eubacterial and archaeal sequences. Rad18 is involved in DNA repair in response to damage caused by UV irradiation. Rad21 in *S. pombe* (homologous to Scc1 in the cohesin complex of *S. cerevisiae*) also functions in the repair of DNA damaged by ionizing radiation. From our analysis it is evident that the SCC proteins are not present in eubacteria or archaea. Hence, it is possible that the SMCs of the ancestral family cluster with the eukaryotic Rad18 homologs because the eubacterial and archaeal proteins function both as DNA repair proteins and as structural proteins holding sister chromatids together during cell division.

The phylogenetic tree we have constructed is similar to that of Cobbe and Heck [18], but three new important results are revealed in our tree. First, Rad18 homologs cluster with the ancestral SMC proteins, as opposed to clustering as a separate family; second, two SMC homologs are identified in two species of eubacteria, as opposed to only single SMC homologs identified in these species [18]; and third, Mmip1 protein is identified as a new homolog of the SMC mouse protein, a protein not present in the tree previously constructed [18].

Cohesion interaction network

The interaction network is based on a combination of data sources, and includes data from two large-scale yeast two-hybrid screens [40]. It is known that this technique can lead to false-positive results [57,58] and false negatives (due to protein instability, for example [59]). However, it is still the only high-throughput technique that can be applied effectively to an entire genome. To screen more than 6,000 *S. cerevisiae* ORFs, some errors are inevitable and these have to be accepted as part of the data. We have combined yeast two-hybrid data with that available in the YPD and in the literature.

The cohesin macromolecular complex is shown to be at the center of a complex interaction network that connects the mechanism of cohesion to proteins in the condensation and spliceosomal pathways and in the spindle pole body (Figure 3). These interactions also show the potential overlap of proteins involved in sister-chromatid cohesion during mitosis and meiosis. Trf4 and Tid3 are both proteins that interact with Smc1 from the cohesin complex and Smc2 from the condensin complex. Scc2, one of the cohesin loading factors, interacts with Prp11, a splicing factor, which in turn interacts with Spc24, a protein of the spindle pole body that interacts with Tid3. Tid3 makes an important interaction with Dmc1, a meiosis-specific protein, which is required for recombination, synaptonemal complex formation and cell-cycle progression [60]. Dmc1 interacts with large number of other proteins, including Apc2 which is a component of the anaphase-promoting complex (APC). The macromolecular

APC is a ubiquitin protein ligase that is essential for mitotic cyclin proteolysis and sister chromatid separation [61].

The identification of this extensive network of interacting proteins provided us with a data set of genes and proteins that we could search for common motifs. We identified a large number of protein sequence motifs that were shared by pairs of proteins in the network, and which could potentially represent shared binding sites. The most interesting discovery was the motif shared by *Scc2* and *Chk1*, which comprised part of the serine/threonine protein kinase motif, including the active-site residue. Eukaryotic protein kinases are characterized by 12 subdomains and 12 conserved residues [62]. In a ClustalX alignment of *Scc2* and *Chk1* from *S. cerevisiae*, *Scc2* (which comprises 1,493 residues compared to the 527 residues of *Chk1*) does not have the 12 subdomains, but it does have 5 of the 12 conserved residues, only two of which are in the shared catalytic-site motif identified using the Teiresias algorithm. The potential for *Scc2* to function as a kinase needs to be validated experimentally. If *Scc2* functioned as a kinase it would provide further evidence for the regulation of sister-chromatid cohesion by phosphorylation. It has been already been shown that phosphorylation of *Scc1* is a requirement for its subsequent cleavage by the *Esp1* separin, which occurs at the metaphase-to-anaphase transition of mitosis [13]. If *Scc2* is a kinase, then the cohesin-loading function of the *Scc2-Scc4* complex might simply be the phosphorylation of *Scc1* or one of the other cohesin subunits to allow DNA binding.

The detection of regulatory motifs in gene promoter regions may provide evidence for the functional assignment of genes of unknown function [63]. We found genes encoding six (or seven if an extreme upstream region is considered) cohesion network proteins sharing a common upstream regulatory sequence that includes the MCB box element. The MCB element is found in upstream regions of genes that encode proteins preferentially synthesized at the G1/S boundary of the cell cycle. MCB or MCB-like elements are common upstream of genes for DNA replication, repair and recombination and of a number of cyclin genes [51]. An upstream DNA motif common to a number of genes could be indicative of DNA-binding preferences of transcription factors and hence coexpression or co-regulation of the genes. It would be possible to investigate coexpression levels experimentally using microarray technology. We found that three of the subunits of the cohesin complex - *Scc1*, *Scc3*, *Smc3* - shared this common upstream motif, along with *Eco1*, *Pds1* and *Spc24* (a protein not directly associated with the cohesin complex). We searched for the expression levels of these six proteins in the data from the microarray hybridization experiment of cell-cycle-regulated genes in *S. cerevisiae* [64,65]. The data for these six proteins was incomplete, but we found that the mRNA levels of five of the six genes (*Scc1*, *Scc3*, *Smc3*, *Pds1* and *Eco1*) peaked in G1 phase of the cell

cycle, and that *Spc24* appeared not to be cell-cycle regulated. We also found that the mRNA levels of *Smc1* peaked in G1. Spellman [64] observed from the expression data that *Smc1*, *Smc3*, *Scc1*, *Pds1* and *Pds5* showed tight temporal regulation. This set of proteins includes three of those we found to contain the MCB upstream regulatory region. Further specific microarray data is needed to investigate the expression levels in more detail. However, the coexpression of *Pds1* with subunits of the cohesin complex (*Scc1*, *Scc3* and *Smc3*) would fit with the theory that *Pds1* acts a protector of cohesin, inhibiting its breakdown which signifies the onset of anaphase.

The securin *Pds1* is an anaphase inhibitor that is complexed to *Esp1* for part of the cell cycle. The destruction of *Pds1* by the ubiquitin ligase APC is the control point for the separation of sister chromatids [66,67]. APC is activated by *Cdc20* during mitosis and the proteolytic function of *Cdc20*-APC depends on a destruction box motif [54] which has been identified in *Pds1*. We found three destruction box motifs in *Smc3*. It is possible that two of the potential destruction box motifs in *Smc3* are linked to the periodicity of residues in the coiled-coil domain. But one destruction box motif is in the hinge region, and if this targets the proteins for destruction by APC, then proteolysis of *Smc3* could also be connected with the metaphase-to-anaphase transition. APC is activated by *Cdh1* during late mitosis/G1 and *Cdh1*-APC requires a KEN box recognition signal (KENXXXN) to identify its substrates [56]. We identified a KEN box in the sequence of *Smc2*. If this targets *Smc2* as an APC substrate, then proteolysis of this protein could be an event that occurs in late mitosis/G1 of the cell cycle.

Conclusions

We have combined available genomic and proteomic data into a comprehensive network of information to reach a better understanding of the function of the cohesin complex. We have identified a new sequence homolog to SMC that highlights the capability of a protein to conduct different functions in different cellular locations, and have created a new phylogeny for SMC proteins that includes a number of eukaryotic sequences (predominantly *Rad18* homologs) within an ancestral family of SMC proteins. A complex network of interacting cohesion proteins was identified. Six of these networked proteins contain a known upstream regulatory sequence. In addition we have identified a number of protein pairs within the network that share protein motifs, which could indicate a common structural feature such as a binding site. In this way we discovered a motif shared by *Scc2* and *Chk1* that suggests *Scc2* could have kinase activity. This hypothesis needs to be verified experimentally. The potential for *Scc2* to function as a kinase could provide further evidence for the regulation of sister-chromatid cohesion by phosphorylation mechanisms, which are currently poorly understood.

Materials and methods

Homologs of cohesin proteins

A PSI-BLAST [68] search of the nonredundant protein sequence database (GenBank CDS translations, PDB, SWISS-PROT, PIR and PRF) for sequence homologs of Smc1, Smc3, Scc1, Scc2, Scc3 and Scc4 from *S. cerevisiae* was conducted. PSI-BLAST was used for five iterative rounds of searching and a cut-off threshold of 0.001. Homologs from nine eukaryotic species are listed in Table 1. A PSI-BLAST search was also conducted for four proteins (Trf4, Prp11, Tid3 and Esp1) that interact with cohesin proteins (Table 2).

A phylogeny for SMC proteins

The sequence homologs to Smc3 from *S. cerevisiae* identified from the PSI-BLAST search were globally aligned in ClustalX (using default parameters). This alignment was used as input for the PHYLIP program [69,70]. PHYLIP (PHYLogeny Inference Package) includes a number of programs for the creation of evolutionary trees. PROTPARS was used to estimate the phylogeny from the ClustalX sequence alignment using the parsimony method. CONSENSE was used to create a consensus tree using the majority-rule consensus tree method, and DRAWTREE was used to plot an unrooted phylogeny (Figure 2).

Identification of known sequence profiles, motifs and domains

Each cohesin protein sequence (Smc1, Smc3, Scc1, Scc2, Scc3 and Scc4) and the four interacting proteins (Trf4, Prp11, Tid3 and Esp1) from *S. cerevisiae* were searched for PROSITE motifs [71] and PFAM domains [72]. The PROSITE database was searched using ProfileScan [73]. PFAM was searched using the protein search facility at the PFAM website [74].

Identification of the interaction network

A large-scale yeast two-hybrid screen has been conducted to identify protein-protein interactions in *S. cerevisiae* [40]. This screen revealed 957 putative interactions involving 1,004 proteins, and the data is publicly available from the CuraGen Corporation [75]. The Yeast Proteome Database (YPD) [76] collates information from the literature, and includes data on protein-protein interactions. Some of these interaction data are derived from the yeast two-hybrid screen. Both the YPD and the yeast two-hybrid data were searched for each of the six cohesin proteins (Smc1, Smc3, Scc1, Scc2, Scc3 and Scc4) to find their interaction partners. This enabled the creation of an interaction network that surrounds the cohesin molecule (Figure 3).

Identification of shared protein sequence motifs

Teiresias, a two-phase general purpose pattern discovery algorithm developed at IBM [53,54], was used to search for common motifs between two or more protein sequences in a network of 17 proteins (Figure 3). The sequence pattern discovery search engine was used for 'exact discovery' with

parameter $L = 4$ (minimum number of literals, that is, non-wild characters in pattern) and $W = 6$ (maximum extent spanned by any L consecutive literals in the reported pattern). All other parameters were used with default settings. Significance values (calculated by assuming each pattern is used as a predicate to search a database with the size and composition of GenPept) were calculated, and only those with a probability value of ≤ 30.0 were considered significant.

A search was then made for the presence of the significant motifs in SWISSPROT, TREMBL, and PIR, using the 'find-patterns' facility of GCG (Wisconsin Package Version 10.0). This facility in GCG was also used to search for destruction box (RXXLXXXXN) and KEN box (KENXXXXN) motifs in the 17 proteins in the cohesion network.

Identification of shared DNA upstream elements

AlignACE, a Gibbs sampling algorithm for identifying motifs over-represented in DNA sequences [77,78] was used to search for common upstream regions of the genes encoding 17 proteins in the cohesion network. The 'Full Analysis' option (which includes three procedures: Extraction, AlignACE and CompareACE) was selected to search the upstream regions of the selected *S. cerevisiae* genes. Extraction selects the upstream regions of each gene, AlignACE is the motif finding algorithm, and CompareACE compares the extracted motifs with a set of previously identified motifs from *S. cerevisiae*. Only motifs with a Map score of > 10.0 were considered significant.

The significant motifs were then used to search the upstream regions of the complete yeast genome using the PatMatch search facility of the SGD [49,50]. PatMatch was used to search 1,000 and 2,000 bp upstream of the 5' untranslated regions, searching both strands, with no mismatches, deletions or insertions.

Acknowledgements

We thank Frank Uhlmann for helpful discussions and comments on the manuscript.

References

1. Strunnikov AV, Larionov VL, Koshland D: **SMC1: an essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family.** *J Cell Biol* 1993, **123**:1635-1648.
2. Michaelis C, Ciock R, Nasmyth K: **Cohesins: chromosomal proteins that prevent premature separation of sister chromatids.** *Cell* 1997, **91**:35-45.
3. Guacci V, Koshland D, Strunnikov: **A direct link between sister chromatid cohesion and chromosome condensation revealed through analysis of MCD1 in *S. cerevisiae*.** *Cell* 1997, **91**:47-57.
4. Ciosk R, Shirayama M, Scevcenko A, Tanaka T, Toth A, Shevcenko A, Nasmyth K: **Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins.** *Mol Cell* 2000, **5**:243-254.
5. Strunnikov AV: **SMC proteins and chromosome structure.** *Trends Cell Biol* 1998, **8**:454-459.

6. Darwiche N, Freeman L, Strunnikov A: **Characterisation of the components of the putative mammalian sister chromatid cohesin complex.** *Gene* 1999, **233**:39-47.
7. Hirano T: **SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates.** *Genes Dev* 1999, **13**:11-19.
8. Melby TE, Ciampaglio CN, Briscoe G, Erickson HP: **The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge.** *J Cell Biol* 1998, **142**:1595-1604.
9. Nasmyth K: **Separating sister chromatids.** *Trends Biochem Sci* 1999, **24**:98-104.
10. Toth A, Ciosk R, Uhlmann F, Galova M, Schleiffer A, Nasmyth K: **Yeast cohesin complex requires conserved protein, Eco1p (Ctf7), to establish cohesion between sister chromatids during replication.** *Genes Dev* 1999, **13**:320-333.
11. Uhlmann F, Lottspeich F, Nasmyth K: **Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesion subunit Scc1.** *Nature* 1999, **400**:37-42.
12. Skibbens RV, Corson LB, Koshland D, Hieter P: **Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery.** *Genes Dev* 1999, **13**:307-319.
13. Uhlmann F, Wernic D, Poupard M-A, Koonin EV, Nasmyth K: **Cleavage of cohesion by the CD clan separin triggers anaphase in yeast.** *Cell* 2000, **103**:375-386.
14. Ciosk R, Zachariae W, Michaelis C, Shevchenko A, Mann M, Nasmyth K: **An ESPI/PDS1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast.** *Cell* 1998, **93**:1067-1076.
15. Nasmyth K, Peters J-M, Uhlmann F: **Splitting the chromosome: cutting the ties that bind sister chromatids.** *Science* 2000, **288**:1379-1384.
16. Saitoh N, Goldberg I, Earnshaw WC: **The SMC proteins and the coming of age of the chromosome scaffold hypothesis.** *Bio-Essays* 1995, **17**:759-766.
17. Strunnikov A, Jessberger R: **Structural maintenance of chromosomes (SMC) proteins. Conserved properties for multiple biological functions.** *Eur J Biochem* 1999, **263**:6-13.
18. Cobbe N, Heck MMS: **Review: SMCs in the world of chromosome biology from prokaryotes to higher eukaryotes.** *J Struct Biol* 2000, **129**:123-143.
19. Lehmann AR, Walicka M, Griffiths DJF, Murray JM, Watts FZ, McCready S, Carr AM: **The rad18 gene of *Schizosaccharomyces pombe* defines a new subgroup of the SMC superfamily involved in DNA repair.** *Mol Cell Biol* 1995, **15**:7067-7080.
20. Foustier MI, Lehmann AR: **A novel SMC protein in *Schizosaccharomyces pombe* contains the rad18 DNA repair protein.** *EMBO J* 2000, **19**:1691-1702.
21. Niki H, Imamura R, Kitaoka M, Yamanaka K, Ogura T, Hiraga S: ***E. coli* MukB protein involved in chromosome partition forms a homodimer with a rod-and-hinge structure having DNA binding and ATP/GTP binding activities.** *EMBO J* 1992, **11**:5101-5109.
22. Yamazoe M, Onogi T, Sunako Y, Niki H, Yamanaka K, Ichimura T, Hiraga S: **Complex formation of MukB, MukE and MukF proteins involved in chromosome partitioning in *Escherichia coli*.** *EMBO J* 1999, **18**:5873-5884.
23. van de Ent F, Lockhart A, Kendrick-Jones J, Lowe J: **Crystal structure of the N-terminal domain of MukB: a protein involved in chromosome partitioning.** *Structure Fold Des* 1999, **7**:1181-1187.
24. Wu R, Couchman J: **cDNA cloning of the basement membrane chondroitin sulfate proteoglycan core protein, Bamacan: a five domain structure including coiled-coil motifs.** *J Cell Biol* 1997, **136**:433-444.
25. Ghiselli G, Siracusa LD, Iozzo RV: **Complete cDNA cloning, genomic organization, chromosomal assignment, functional characterisation of the promoter and expression of the murine bamacan gene.** *J Biol Chem* 1999, **274**:17384-17393.
26. Gupta K, Anand G, Yin X, Grove L, Prochownik EV: **Mmp1: a novel leucine zipper protein that reverses the suppressive effects of Mad family members on c-myc.** *Oncogene* 1998, **16**:1149-1159.
27. Hirano M, Hirano, T: **ATP-dependent aggregation of single-stranded DNA by a bacterial SMC homodimer.** *EMBO J* 1998, **17**:7139-7148.
28. Lupas A: **Coiled coils: new structures and new functions.** *Trends Biochem Sci* 1996, **21**:375-382.
29. Birkenbihl RP, Subramani S: **The rad21 gene product of *Schizosaccharomyces pombe* is a nuclear, cell cycle-regulated phosphoprotein.** *J Biol Chem* 1995, **270**:7703-7711.
30. Losada A, Hirano M, Hirano T: **Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion.** *Genes Dev* 1998, **12**:1986-1997.
31. Mckay M, Troelstra C., van der Spek P, Kanaar R, Smit B, Hagemerijer A, Bootsma D, Hoeijmakers, JHJ: **Sequence conservation of the Rad21 *Schizosaccharomyces pombe* DNA double-strand break repair gene in human and mouse.** *Genomics* 1996, **36**:305-315.
32. Carramolino L, Lee BC, Zaballos A, Peled A, Barthelemy I, Shav-Tal Y, Proeto I, Carmi P, Gothelf Y, de Buitrago GG, et al.: **SA-I, a nuclear protein encoded by one member of a novel gene family: molecular cloning and detection in hemopoietic organs.** *Gene* 1997, **195**:151-159.
33. Pezzi N, Prieto I, Kremer L, Jurado LAP, Valero C, Del Mazo J, Martinez, Barbero JL: **STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion.** *FASEB J* 2000, **14**:581-592.
34. Furuya K, Takahashi K, Yanagida M: **Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G₁ phase.** *Genes Dev* 1998, **12**:3408-3418.
35. Rollins RA, Morcillo P, Dorsett D: **Nipped-B, a *Drosophila* homolog of chromosomal adherins, participates in activation of remote enhancers in the cut and Ultrabithorax genes.** *Genetics* 1999, **152**:577-593.
36. Seitz LC, Tang, KL, Cummings WJ, Zolan ME: **The rad9 gene of *Coprinus cinereus* encodes a proline-rich protein required for meiotic chromosome condensation and synapsis.** *Genetics* 1996, **142**:1105-1117.
37. Kondoh N, Wakatsuki T, Ryo A, Hada A, Aihara T, Horiuchi S, Goseki N, Matsubara O, Takenaka K, Shichita M, et al.: **Identification and characterization of genes associated with human hepatocellular carcinogenesis.** *Cancer Res* 1999, **59**:4990-4996.
38. Castano IB, Broska PM, Sadoff BU, Chen H, Christman MF: **Mitotic chromosome condensation in the rDNA requires TRF4 and DNA topoisomerase I in *Saccharomyces cerevisiae*.** *Genes Dev* 1996, **10**:2564-2576.
39. Walowsky C, Fitzhugh DJ, Castano IB, Ju JY, Levin NA, Christman MF: **The topoisomerase-related function gene TRF4 affects cellular sensitivity to the antitumour agent camptothecin.** *J Biol Chem* 1999, **274**:7302-7308.
40. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al.: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.
41. Wang Z, Castano IB, De Las Penas A, Adams C, Christman MF: **Pol k: a DNA polymerase required for sister chromatid cohesion.** *Science* 2000, **289**:774-779.
42. Wang SW, Norbury C, Harris AL, Toda T: **Caffeine can override the S-M checkpoint in fission yeast.** *J Cell Sci* 1999, **112**:927-937.
43. Legrain P, Chapon C: **Interaction between Prp11 and Spp91 yeast splicing factors and characterization of a Prp9-Prp11-Spp91 complex.** *Science* 1993, **262**:108-110.
44. Matsushima Y, Matsumura K, Kitagawa Y: **Zinc finger-like motif conserved in a family of RNA binding proteins.** *Biosci Biotechnol Biochem* 1997, **61**:905-906.
45. Hodges PE, Beggs JD: **RNA splicing. U2 fulfils a commitment.** *Curr Biol* 1994, **4**:264-267.
46. Wiest DK, O'Day CL, Abelson J: **In vitro studies of the Prp9-Prp11-Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation.** *J Biol Chem* 1996, **271**:33268-33276.
47. Bennett M, Michaud S, Kingston J, Reed R: **Protein components specifically associated with prespliceosome and spliceosome complexes.** *Genes Dev* 1992, **6**:1986-2000.
48. Zheng L, Chen Y, Lee, WH: **Hec1p, an evolutionarily conserved coiled-coil protein, modulates chromosome segregation through interaction with SMC proteins.** *Mol Cell Biol* 1999, **19**:5417-5428.
49. Chervitz SA, Hester ET, Ball CA, Dolinski K, Dwight SS, Harris MA, Juvik GG, Malekian A, Roberts S, Roe T, et al.: **Using the Saccharomyces Genome Database (SGD) for analysis of protein similarities and structure.** *Nucleic Acids Res* 1999, **27**:74-78.

50. **Saccharomyces Genome Database** [<http://genome-www.stanford.edu/Saccharomyces>]
51. McIntosh EM: **MCB elements and the regulation of DNA replication genes in yeast.** *Curr Genet* 1993, **24**:185-192.
52. Machado AK, Morgan BA, Merrill GF: **Thioredoxin reductase-dependent inhibition of MCB cell cycle box activity in *Saccharomyces cerevisiae*.** *J Biol Chem* 1997, **272**:17045-17054.
53. Rigoutsos I, Floratos A: **Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm.** *Bioinformatics* 1998, **14**:55-67.
54. **Teiresias** [<http://www.research.ibm.com/bioinformatics/teiresias.html>]
55. Glotzer M, Murray AW, Kirschner MW: **Cyclin is degraded by the ubiquitin pathway.** *Nature* 1991, **349**:132-138.
56. Pfleger CM, Kirschner MW: **The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1.** *Genes Dev* 2000, **14**:655-665.
57. Fields S, Sternglanz R: **The two-hybrid system: an assay for protein-protein interactions.** *Trends Genet* 1994, **10**:286-292.
58. Hengen PN: **Methods and reagents: false positives from the yeast two-hybrid system.** *Trends Biochem Sci* 1997, **22**:33-34.
59. Warbrick E: **Two's company, three's a crowd: the yeast two hybrid system for mapping molecular interactions.** *Structure* 1997, **5**:13-17.
60. Bishop DK, Park D, XU LZ, Kleckner N: **DMC1 - a meiosis-specific yeast homolog of *Escherichia coli* RECA required for recombination, synaptonemal complex-formation, and cell-cycle progression.** *Cell* 1992, **69**:439-456.
61. Zachariae W, Nasmyth K: **Whose end is destruction: cell division and the anaphase-promoting complex.** *Genes Dev* 1999, **13**:2039-2058.
62. Hanks SK, Hunter T: **The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.** *FASEB J* 1995, **9**:576-596.
63. Goffeau A: **Genomic-scale analysis goes upstream?** *Nat Biotechnol* 1998, **16**:907-908.
64. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell-cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.** *Mol Biol Cell* 1998, **9**:3273-3297.
65. **Yeast Cell Cycle Analysis Project** [<http://genome-www.stanford.edu/cgi-bin/cell-cycle/search>]
66. Tinker-Kulberg RL, Morgan DO: **Pds1 and Esp1 control both anaphase and mitotic exit in normal cells and after DNA damage.** *Genes Dev* 1999, **13**:1936-1949.
67. Cohen-Fix O, Koshland D. **Pds1p of budding yeast has dual roles: inhibition of anaphase initiation and regulation of mitotic exit.** *Genes Dev* 1999, **13**:1950-1959.
68. Altschul SF, Madden TL, Schaffer AA: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
69. Felsenstein J: **PHYLIP: Phylogeny interface package (Version 3.2).** *Cladistics* 1993, **5**:164-166.
70. **PHYLIP (PHYLogeny Inference Package)** [<http://evolution.genetics.washington.edu/phylip/phylip.html>]
71. Hofmann K, Bucher P, Falquet L, Bairoch A. **The PROSITE database, its status in 1999.** *Nucleic Acids Res* 1999, **27**:215-219.
72. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer ELL: **The Pfam protein families database.** *Nucleic Acids Res* 2000, **28**:263-266.
73. **ProfileScan** [<http://www.isrec.isb-sib.ch/software/PFSCA-form.html>]
74. **PFAM website** [<http://www.sanger.ac.uk/software/pfam>]
75. **Curagen Corporation** [<http://curatools.curagen.com>]
76. Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, Hodges PE, Kondu P, Lengjeza C, Lew-Smith JE, Lingner C, et al.: **The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism information.** *Nucleic Acids Res* 2000, **28**:73-76.
77. Hughes JD, Estep PV, Tavazoie S, Church GM: **Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*.** *J Mol Biol* 2000, **296**:1205-1214.
78. **AlignACE** [<http://atlas.med.harvard.edu>]