The recent structural studies have narrowed down the number of possible models for sister chromatid cohesion and provide a clearer view of how the components of the cohesin complex interact. Nonetheless, several key questions regarding the mechanism of sister chromatid cohesion still need to be addressed: can cohesins form higher-order structures, how is the cohesin complex recruited to specific regions of the chromosome, how does the passage of the replication fork lead to the linkage of sister chromatids and what role does ATP hydrolysis play in sister chromatid cohesion. Undoubtedly, the establishment and maintenance of sister chromatid cohesin is a dynamic process and many of its moving parts still need to be elucidated.

Acknowledgements

We thank Ann Corsi, April Robbins, Ritu Agarwal, Paul Megee, Doug Koshland and Tatsuya Hirano for stimulating and helpful discussions.

References

1 Anderson, D.E. *et al.* (2002) Condensin and cohesin display different arm conformations with characteristic hinge angles. *J. Cell Biol.* 156, 419–424

- 2 Haering, C.H. *et al.* (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* 9, 773–788
- 3 Losada, A. *et al.* (2000) Identification and characterization of SA/Scc3p subunits in the *Xenopus* and human cohesin complexes. *J. Cell Biol.* 150, 405–416
- 4 Sumara, I. *et al.* (2000) Characterization of vertebrate cohesin complexes and their regulation in prophase. *J. Cell Biol.* 151, 749–762
- 5 Pasierbek, P. et al. (2001) A Caenorhabditis elegans cohesion protein with functions in meiotic chromosome pairing and disjunction. Genes Dev. 15, 1349–1360
- 6 Hirano, T. (2002) The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev.* 16, 399–414
- 7 Hirano, M. *et al.* (2001) Bimodal activation of SMC ATPase by intra- and inter-molecular interactions. *EMBO J.* 20, 3238–3250
- 8 Cohen-Fix, O. (2001) The making and breaking of sister chromatid cohesion. *Cell* 106, 137–140
- 9 Toth, A. *et al.* (1999) Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes Dev.* 13, 320–333
- 10 Tanaka, T. *et al.* (1999) Identification of cohesin association sites at centromeres and along chromosome arms. *Cell* **98**, 847–858
- 11 Megee, P.C. *et al.* (1999) The centromeric sister chromatid cohesion site directs Mcd1p binding to adjacent sequences. *Mol. Cell* 4, 445–450
- 12 Blat, Y. *et al.* (1999) Cohesins bind to preferential sites along yeast chromosome III, with

differential regulation along arms versus the centric region. *Cell* **98**, 249–259

- 13 Laloraya, S. *et al.* (2000) Chromosomal addresses of the cohesin component Mcd1p. *J. Cell Biol.* 151, 1047–1056
- 14 Ciosk, R. et al. (2000) Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* 5, 243–254
- 15 Hartman, T. *et al.* (2000) Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in *Saccharomyces cerevisiae. J. Cell Biol.* 151, 613–626
- 16 Panizza, S. *et al.* (2000) Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. *Curr. Biol.* 10, 1557–1564
- 17 Skibbens, R.V. et al. (1999) Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev.* 13, 307–319
- 18 Ross, K.E. *et al.* (2002) Separase: a conserved protease separating more than just sisters. *Trends Cell Biol.* 12, 1–3
- 19 Dej, K.J. *et al.* (2000) Separation anxiety at the centromere. *Trends Cell Biol.* 10, 392–399
- 20 Cohen-Fix, O. (2000) Sister chromatid separation: falling apart at the seams. *Curr. Biol.* 10, R816–819

Joseph L. Campbell

Orna Cohen-Fix*

The Laboratory of Molecular and Cellular Biology, NIDDK, NIH 8 Center Drive, Bethesda, MD 20892, USA. *e-mail: ornacf@helix.nih.gov

Protein Sequence Motif

YTH: a new domain in nuclear proteins

Peter Stoilov, Ilona Rafalska and Stefan Stamm

A novel 100–150-residue domain has been identified in the human splicing factor YT521-B and its *Drosophila* and yeast homologues. Homology searches show that the domain is typical for the eukaryotes and is particularly abundant in plants. It is predicted to adopt a mixed α -helix- β -sheet fold and to bind to RNA. We propose the name YTH (for YT521-B homology) for the domain.

Published online: 5 September 2002

One of the most prominent features of eukaryotic genes is their discontinuity. Analysis of the working draft of the human genome has shown that, on average, introns account for 95% of the pre-mRNA [1]. The precise removal of intron sequences by the spliceosome is crucial for gene expression. However, the sequences of the two splice sites and the branch point that are recognized by spliceosome components are clearly insufficient to identify the exons in the pre-mRNA sequence. Correct recognition of exons is achieved by the cooperative action of multiple splicing factors auxiliary to the spliceosome [2]. These factors bind in a sequence-specific manner to the pre-mRNA [3,4] and recruit the spliceosome components to the splice sites. The nuclear protein YT521 has been identified in two-hybrid screens with splicing factors [5,6]. It interacts with several splicing factors, both in two-hybrid and co-immunoprecipitation assays [5,6].

In addition, the alternative splicing patterns of the splicing factor SRp20 and hTra2- β pre-mRNAs are altered by YT521-B [6]. YT521 does not belong to any of the known splicing factor families. The only sequence feature that it shares with some of the splicing factors is an RE/D repeat. Here we report the identification of a new domain in splicing factor YT521-B and a number of proteins of unknown function that could be involved in RNA binding.

Domain characterization

During BLAST searches to identify YT521-B homologues, a conserved part of the protein was identified between residues 356 and 499 of the rat YT521-B

	18 20 38 40 50 60 70 85 90 108 110 130 135 140 148
decondary structure	TEXT NUMBER OF THE TOTAL TRADUCT INTERACT IN SOME TAXABLE IN SOMETHING.
gi 18938641* A. gambiae	APPENDENCES ALSO AND A A A A A A A A A A A A A A A A A A
g1[3809107] A. thaliana	A FEW TRESS DOWNED DY SAN DO THORN DAA DO ATTYCLED OR DE PLOTE BAAR OF ALL AND TY DEVELOPMENT OF AVECTING AVECTING AND THORN DE VISION OF AN AVECTING AND THE RECEIPTING AND THE
gi 18190169 A. thallana	A EP PUTET IN DEV VENTION OF COMPANY AND THAT AND A COMPANY
gi[18997519 A. thaliana	TY WAR EXCEPTED AND DESIGN AND DE
gi 18399947 A. thaliana	ALL FLIPS TREENVERTICHTWEITER SCAARSEN
gi 18000030 A. thaliana	APPETERS SEDVES IN WASTINGER CAASIGED
gi 18401473 A. thaliana	AR FYDES STODY HIS DOWNS DOWN SOFTED ACRIMINATION SECTOR IS BOOK FOR ALL OF A
gi 18402045 A. thaliana	APPYLED REDVELISYNYE ITAGEKSOR EACEINERACEINE FOR HER CONTRACTOR OF THE SAME STATE OF A STA
gi 18405306 A. thaliana	ASPECTASES DO VESTION VASTORIS CONVESTION VASTORIS CARAGONO CONSIGNATION PROVIDED AND CONTRACTOR VASTORIS CONTRACTOR V
g1 18412315 A. thaliana	A F PV I & THE DED I RELEVAN WETTAU BROCK A RECEVANT OF THE MARCE OF THE WARD OF THE MARCE - THE THE PURCH I REPORT OF THE RECOVER AND THE RECOVER
gi 18613720 A. thaliana	TEVELLED TO DE DE MARCONENTENTO VAN DE LA DA PARCO
gi 18424075 A. thaliana	ART FLY KEPS ID BY THE DOT WARD TO HER KOLOTH BOARD
gi 18424423 A. thaliana	APPWIE THE THE TRUTCH AND THE REAL AND THE ACCOUNTS FOR THE ATTRUCT OF A AND PROPERTY MODEL. THE REAL AND THE ATTRACT ATTRACT AND THE ATTRACT ATTRACT ATTRACT
gi 20453238 A. thaliona	APPVILES SUDVERCINGES STOPS ST
g1 15979210* 9. tearum	INTERNALISELE SOCIOLINETTONIC ADAMSSION - MYTANIE SOCIAL S
gi 17862688 D. melanogaster	APPVILT RECEIPTION OF THE RECORDER OF THE RECORDER OF THE RECORD FOR T
gi 5452477 D. meLenogaster	TREELES STOLD ST
gi 16885895* D. rerio	OFFICE TREETER DESCRIPTION OF TREETER DATA AND A TREETER DESCRIPTION OF TREETER DESCRIPTION
Q1 33479378* G. max	The Antiparties of the state of
g1 16347632* G. max	APPVIRTUEDDING DODARD TWENTY END ADDRESS AND ADDRESS TO THE PRODUCT OF ADDRESS TO THE POST OF THE POST
gi 18849234* C. max	ALP PUTE THE DEPARTMENT OF A DATA STRUCTURE OF
g1[33246678* G. arboreum	APPVILES STOPVES IN THE ASTOPVES IN THE RECORD FOR THE PLANT AND THE RECORD FOR T
gi 10438157 H. sapiens	VEY FIRE FLACE FOR THE RECORD FOR FOR THE RECORD FOR THE RECORD FOR THE RECORD FO
gi 15128560 H. sapiene	CEVEL IN A RECEIPTION OF CONTRACT ON A PROVE OF A VALUE AND A VALUE AN
g1 16551030 H. septers	APPETER NEEDYS LAGAGANY D'LAVAET DALAERAAC
gi 19851932 H. sapiens	GEVELLE LEEDING DATE CONTRACTOR DATE OF A MARKEN AND STATUTED AND A CONTRACTOR DATE AND A CONTRACTOR DATE OF
gi. 18209266* H. valgare	ARE FILLED STORAGED AND THE RECOMPTING AND THE RECOMPTING AND THE RECOMPTING AND THE RECOVER OF THE
gi 18222703* H. valgare	ARPYTER DEDUCED RELEVAN ANTHOR RECORDING TO AND ADDRESS OF AND ADDRESS AND ADDRESS ADDR
gi 13778702* L. esculentum	THY PUTELISHED CLARENESS TO AN A DESCRIPTION OF A DESCRIP
gi 13370421" M. trancetula	AUPIVIES STOTES DOTES DOTES DOTES DOTES AND AND ADDRESS OF A LAST OF A LAST OF A LAST OF A DRESS OF ADDRESS OF ADDRE
gi 13783728* M. truncatula	YTA AND A REPARED AND AND AND AND AND AND AND AND AND AN
gill4985474" M. trancatala	ARPETER SECONSECTON AND THE RECOVER OF A LAST OF A
gi 14996178* M. truncatula	APPVIDE SEGNARS DOWNED DOWNED AND AND CAR EXCENTION OF THE BARACE OF AN EVER PARTNER OF THE BARACE OF A DESCENT OF THE BARACE OF THE BARA
gi 20907368" M. trancatala	APPNIE PRODUCT NO. THE CANAD GRAVEL AT THE PROVIDE PROVIDE A CONTRACT OF
gi 15920653 M. masculus	GRVFTIRESEDDTRESTONTE CTERCECOLORS DOWN - GRVVTL PERIODOR - GWERKAWOYN CAVAR (G
gi 15887003 C. sativa	
gil4882965* G. sativa	APPE DEMANDED AT A DEVELOPMENT AND ADDRESS
gi 19173797 R. DOTVegicus gi 6320182 R. Derevisias	APPY TEAL OF THE PARTY OF THE REPORT OF THE PARTY OF THE
g1 8320182 S. cerevisiae g1 33674434* S. tropicalis	APPETERSNELWARANSVELVENESSAAPEA
gillife/eele" 5. toberosum	
gi 9251721* S. tabercoum	OF PERSONNEL CONNECTION AND ADDRESS CARACTERIZED TO ADDRESS FOR THE PERSONNEL CARACTERIZED AND ADDRESS FOR THE PERSONNEL CARACTERIZED
gill#965595* 9. bicolor	APPY IS SED REVENUE AND ADDRESS AD
gi 12819080" T. sestivum	APPETER AND ADDRESS OF ADDRE
gi 20442447* T. asstivum	APPOTATION NOT THE RECEIPTION OF THE RECEIPTION
gi 9410695* T. sestivus	APPELIES DECOMPLETED IN MATCHINE AND ADDRESS OF ADDRESS
gi 9658741* T. memoroorum	APPETER NUMBER OF A DESCRIPTION OF A DES
gi 11406675* X, laevis	APPETRISHENGLAMMON CONTRACTORS INTO A DATE OF THE DESCRIPTION OF THE D
gi 9731123* Z. mays	APPLY TRUE TRUE TO A THE CALL AND
gi 9731916* E. maps	APPETERS SERVICE AND CONTRACTOR AND CONT
consensus/set	sparling.sps)phurthalwach.spt-arcaug.p. hhbwwyths.p. Gawes.ashtts.w.s. http://wwhite-iphall.s.w-covp.usmass.p.o.phu.h
and a second sec	
	T/BS

Fig. 1. Multiple alignment of sequences containing YT homology (YTH) domains. The sequences are denoted with the GeneBank identifier of the corresponding nucleotide sequence followed by the species name. The GeneBank identifier for the *Saccharomyces cerevisiae* refers to protein entry in the database. Translations of partial EST sequences are indicated with an asterisk. The sequences were identified using searches of the translated EST and the nonredundant nucleotide databases at NCBI using rat YT521-B (shown in italic). Wherever available, the conceptual translation provided in the database was used for the alignment. If no conceptual translation was provided (i.e. for the EST sequences), the reading frame identified by the BLAST search was used if it did not contain stop codons or unknown residues. Redundant sequences originating from the same species were omitted. The similar residues that are present in all sequences are shaded (green) using the PAM250 matrix. The identical residues are shaded in yellow. The secondary structure, as predicted by the PHD program, is shown on top. E denotes extended (β-strands) structure and H denotes the predicted α -helices. The multiple sequence alignment (accession number ALIGN_000432) has been deposited at the European Bioinformatics Institute (ftp://ftp.ebi.ac.uk/pub/databases/embl/align/).

protein. We therefore performed a PSI-BLAST [7] search with this portion of the YT521-B protein, using an exclusion threshold of 0.005. After four iterations, multiple proteins from *Arabidopsis thaliana, Oryza sativa, Homo sapiens, Mus musculus, Drosophila melanogaster, Plasmodium falciparum* and

Saccharomyces cerevisiae where found to have similar regions, with E values in the range 10^{-11} – 10^{-60} . Many of these protein sequences were derived from automated gene prediction and were not confirmed by mRNA or EST sequences. To identify sequences of existing proteins, we performed BLAST [7] searches with the conserved region against the translated nonredundant nucleotide database at NCBI (filtered using 'biomol_mrna[PROP]' as a keyword) and the translated EST database. The sequences from these searches that aligned to the full length of the query and had an E value $<10^{-6}$ were aligned using CLUSTALW [8] after the redundancies had been removed (Fig. 1). Additional BLAST searches against the genome databases confirmed that the conserved region is present exclusively in eukaryotic genomes.

The conserved region appears to define a new domain in these proteins, which we termed the YT homology (YTH) domain. The YTH domain is usually located in the middle of the protein sequence. The domain shows remarkable conservation across a wide species range, with 14 invariant and 19 highly conserved residues. The proteins present in the alignment do not share significant similarity outside the YTH domain, with the exception of the closely related vertebrate homologues of YT521.

We used SMART [9,10] to search the SMART and PFAM databases for additional known domains in the proteins that contain YTH. The search failed to identify any known domains except for three C–C–C–H-type zinc finger domains in GI:18397519, an *A. thaliana* predicted protein of unknown function. Therefore, we conclude that the YTH domain defines a new protein family.

The putative secondary structure was determined using the PHD program [11]. The domain is predicted to have a mixed $\alpha\beta$ -fold, with four α -helices and six β -strands. The conservation pattern follows the predicted secondary structure, with three blocks of conserved sequence separated by loops of variable size. Notable features of the domain are the highly conserved aromatic residues located in the β -sheet.

Most of the proteins identified in the BLAST searches are of plant origin, with 13 distinct sequences coming from a single species (*A. thaliana*). It is unclear whether this protein family is more widespread in plants or whether the observed species distribution is because of bias in the databases.

Predicted function

Biochemically and functionally, YT521-B has been extensively characterized as a pre-mRNA splicing factor [6,12]. In contrast to the other known splicing factors, it lacks a recognizable domain that can confer RNA binding. The conservation of aromatic residues in the β -sheets of the YTH domain is similar to the RNA recognition motif (RRM) domain. In the RRM domain conserved aromatic residues located in the β -sheet are crucial for RNA binding [13]. In addition, experimental evidence shows that the YTH domain is not involved in protein-protein interactions [6]. Based on this, we predict that the biological function of the YTH domain is to bind to RNA.

References

- 1 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- 2 Smith, C.W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388

- 3 Lui, H.-X. *et al.* (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* 12, 1998–2012
- 4 Lui, H.-X. *et al.* (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.* 20, 1063–1071
- 5 Imai, Y. et al. (1998) Cloning of a gene, YT521, for a novel RNA splicing-related protein induced by hypoxia/reoxygenation. Brain Res. Mol. Brain Res. 53, 33–40
- 6 Hartmann, A.M. *et al.* (1999) The interaction and colocalization of Sam68 with the splicingassociated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Mol. Biol. Cell* 10, 3909–3926
- 7 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
- 8 Thompson, J.D. *et al.* (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
- 9 Schultz, J. *et al.* (1998) SMART, a simple modular architecture research tool: identification of

signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5857–5864

- 10 Letunic, I. *et al.* (2002) Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* 30, 242–244
- 11 Rost, B. (1996) PHD: predicting onedimensional protein structure by profile-based neural networks. *Methods Enzymol.* 266, 525–539
- 12 Nayler, O. *et al.* (2000) The ER-repeat protein YT521-B localizes to a novel subnuclear compartment. *J. Cell Biol.* 150, 949–961
- 13 Hoffman, D.W. et al. (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. Proc. Natl. Acad. Sci. U. S. A. 88, 2495–2499

Peter Stoilov Ilona Rafalska

Stefan Stamm

Institute of Biochemistry, Friedrich-Alexander-University Erlangen-Nuremberg, Fahrstrasse 17, 91054 Erlangen, Germany.

BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms

Mark Gomelsky and Gabriele Klug

A novel FAD-binding domain, BLUF, exemplified by the N-terminus of the AppA protein from *Rhodobacter sphaeroides*, is present in various proteins, primarily from Bacteria. The BLUF domain is involved in sensing blue-light (and possibly redox) using FAD and is similar to the flavin-binding PAS domains and cryptochromes. The predicted secondary structure reveals that the BLUF domain is a novel FAD-binding fold.

The prototype and identification of the BLUF domain

AppA is a multidomain protein from the phototrophic proteobacterium *Rhodobacter sphaeroides* (accession number L42555). The N-terminus of the AppA protein is involved in the regulation of photosynthesis gene expression, although its mechanism of action is not well understood [1–4]. The N-terminus of the AppA protein was identified as being involved in repression of photosynthesis genes by blue-light [1], and Gomelsky and Kaplan showed that the N-terminal ~120 residues bind flavin adenine dinucleotide (FAD) noncovalently with an apparent 1:1 stoichiometry [4]. At the time of publication of Ref. [4], two additional bacterial proteins showing sequence similarity to the N-terminus of AppA had been identified [4]. One of these, YcgF from *Escherichia coli* (also known as F403 and b1163; accession number P75990), has been purified and shown to bind FAD [4].

Using the region most conserved between AppA and YcgF (residues 16-108 of AppA), we performed a BLAST search of the nonredundant protein database and a TBLASTN search of the microbial genome database at NCBI, as well as a TBLASTN search of the individual unfinished microbial genomes at the sequencing centers listed in our Acknowledgements. Our searches revealed a variety of uncharacterized proteins containing domains with significant similarity to the N-terminus of AppA. We designated these domains BLUF, for 'sensors of blue-light using FAD'. Most of these proteins are from two branches of Bacteria, Proteobacteria and

Cyanobacteria (Fig. 1). Bacterial genomes contain up to three BLUF domains per genome. No BLUF domains are encoded by the currently available genomes of Archaea. Four BLUF domains are found in Eukarya, all from the unicellular flagellate *Euglena gracilis* [5] (Fig. 1).

Involvement of the BLUF domain in sensory transduction

To our knowledge, the functions of only two proteins containing BLUF domains have been tested experimentally. Similar to the BLUF domain in R. sphaeroides AppA, the BLUF domains from the recently described photoactivated adenylyl cyclase (PAC) from E. gracilis are also involved in the blue-light-dependent control of enzyme activity. Two BLUF domains belong to the α -subunit of the enzyme, PAC α , and two to the PAC β subunit [5] (Fig. 1). To gain an insight into the putative role of the BLUF domains in other proteins we analyzed their domain architecture using the SMART [6] and Pfam [7] databases. Based on the deduced domain structures. all