Sex Plant Reprod (2002) 14:355–357
DOI 10.1007/s00497-002-0129-5

SHORT COMMUNICATION

Ming Yang · Sheila McCormick

The *Arabidopsis MEI1* gene likely encodes a protein with BRCT domains

Received: 2 August 2001 / Accepted: 6 February 2002 / Published online: 15 March 2002

Abstract It was previously proposed that the *Arabidopsis MEI1* gene encodes a small protein with some similarity to an acrosin-trypsin inhibitor of humans [He and Mascarenhas (1998) Sex Plant Reprod 11:199–207]. Here we show that the complement strand of the *MEI1* gene region encodes a predicted protein that fits better with the chromosome fragmentation phenotype of the *mei1* mutant. By aligning their partial cDNA sequence with the two sources of genomic sequence available for the region, we deduced a predicted protein that contains three BRCT domains. Proteins with BRCT domains are involved in DNA repair, recombination, and cell cycle checkpoint control. This newly proposed annotation for MEI1 suggests that it might regulate the repair of DNA double-strand breaks during male meiotic recombination, which is consistent with the chromosome fragmentation defect found in early diplotene in the *mei1* mutant.

Keywords Meiosis · DNA repair · Chromosome fragmentation · BRCT domain · *Arabidopsis*

Introduction

For recombination to occur during meiotic prophase I, it is essential that the chromosomes undergo double-strand breaks and repair. If chromosome breakage is not successfully repaired, chromosome fragmentation can occur, resulting in polyad meiotic products and sterility. An *Arabidopsis* T-DNA line, designated as line CS2343 (*Arabidopsis* Biological Resource Center, Columbus, Ohio), has been described as a male sterile mutant by both He et al. (1996) and Ross et al. (1997). The T-DNA insertion mutation in this line, named *mei1* by He et al. or *mcd1* by Ross et al., exhibits chromosome fragmentation during male meiotic prophase I (Ross et al. 1997), and subsequently produces polyad meiotic products (He and Mascarenhas 1998). Thus, a reasonable prediction would be that the wild-type MEI1 protein is involved in maintaining chromosome stability during meiosis I. However, He and Mascarenhas (1998) reported that the T-DNA disrupted an open reading frame predicted to encode a protein of 89 amino acids. This 89 amino acid peptide showed some similarity to a human acrosin-trypsin inhibitor. As these authors noted, it is difficult to explain how a loss of such a protein could cause the *mei1* phenotype. Moreover, the genomic DNA fragment (8.0 kb) used by He and Mascarenhas to complement the *mei1* phenotype was much larger than the proposed *MEI1* gene, leaving open the possibility that another gene resides on this fragment. Lastly, He and Mascarenhas isolated a cDNA from the region disrupted by the T-DNA and deduced the positions of two introns by sequence alignment with the WS ecotype genomic sequence, but noted that these introns did not comply with the :GU...AG: rule (Brown et al. 1996).

In the course of studying another meiotic mutant, *tardy asynchronous meiosis* (*tam*), we found that *tam* maps in the vicinity of *MEI1*. The *tam* mutant is not an allele of *mei1* (Magnard et al. 2001). We noticed that the *Arabidopsis* Genome Initiative (AGI; The *Arabidopsis* initiative 2000) had predicted a hypothetical protein (F2P24.3), which is encoded by a region encompassing the proposed MEI1 protein, but in the opposite coding orientation. The F2P24.3 predicted protein has BRCT-domains. BRCT-domain-containing proteins are required for double-strand break repair during meiotic recombination in prophase I in *S. cerevisiae* and in mouse (Herrmann et al. 1998; Garcia-Diaz et al. 2000). The molecular nature of F2P24.3 and its chromosomal location raised the possibility that F2P24.3 is actually the protein

M. Yang · S. McCormick (✉)
Plant Gene Expression Center,
United States Department of Agriculture,
Agricultural Research Service and Department of Plant and Microbial Biology, University of California at Berkeley,
800 Buchanan Street, Albany, CA 94710, USA
e-mail: sheilamc@nature.berkeley.edu
Fax: +1-510-5595678

Present address:
M. Yang, Department of Botany, Oklahoma State University,
104 Life Sciences East, Stillwater, OK 74078–3013, USA

encoded by the *MEI1* gene. To explore this possibility, we analyzed the related sequences further.

Database search and sequence analysis

The sequence of BAC F2P24 (AC078898, Columbia ecotype) was retrieved from GenBank. The second gene (*F2P24.3*) on the BAC overlaps with the proposed *MEI1* gene (WS ecotype). We used the Pairwise Blast program (http://www.ncbi.nlm.nih.gov/BLAST) to align the genomic sequences of *F2P24.3* and the corresponding region sequenced from the WS ecotype (accession AF074849). We predicted the gene structure of *F2P24.3* using the GeneScan program (http://genes.mit.edu/GENSCAN.html). We used the predicted protein sequences for Blast searches (http://www.ncbi.nlm.nih.gov/BLAST) and searched for conserved motifs with the ProfileScan program (http://www.isrec.isb-sib.ch/software/PFSCAN_form.html).

Results and discussion

The sequence of *F2P24.3* was compared to the reported *MEI1* cDNA. The positions of the two introns predicted by He and Mascarenhas (1998) were the same as the positions predicted for two of the introns in *F2P24.3*, and in *F2P24.3* these introns comply with the :GU...AG: rule. Thus, it is possible that the cDNA reported by He and Mascarenhas corresponds to a partial cDNA of *F2P24.3* (Fig. 1). Note that the first intron predicted by the AGI annotation is not supported by the partial *MEI1* cDNA from the WS ecotype, and is therefore likely to be mis-annotated.

The alignment of the WS ecotype and Columbia ecotype genomic sequences revealed differences at 16 sites. These differences make the predicted gene structures in the two ecotypes very different if the exon/intron junctions used by the AGI annotation are followed. Therefore, it is likely that there are sequence errors in one or both of the genomic sequences. To better predict the gene structure of *F2P24.3*, we used the cDNA reported by He and Mascarenhas but in the opposite orientation, and we made two assumptions. First, we assumed that both the beginning and end of the open reading frame predicted by AGI were correct. Second, we assumed that the amino acid sequences of F2P24.3 in the Columbia and WS ecotypes should be very similar, if not identical.

It was impossible to find a single open reading frame with the same exon/intron positions that would code for two identical or similar proteins in both ecotypes. The AGI sequence might have an error or errors of nucleotide duplications at 11 of the16 locations, in addition to mismatches with the WS sequence at the other 4 locations. In contrast, there are only two possible errors in the WS sequence: missing nucleotides at positions 2,827 and 8,336. We therefore primarily used the WS genomic sequence to predict the structure of *F2P24.3*, with reference to the AGI annotation. The composite prediction identified an open reading frame that encodes a peptide of 812 amino acids. This version of *F2P24.3* has 12 introns. The predicted *F2P24.3* in GenBank encodes a peptide of 783 amino acids and has 13 introns.

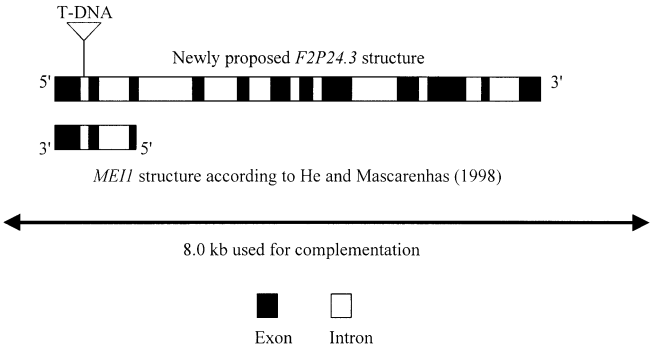

Fig. 1 Diagram aligning newly proposed *F2P24.3* structure with the previously proposed *MEI1* structure. The lengths of *exons* and *introns* are drawn to scale

```
MKTTQLFKGANVFMSRNLVPPEVFDTLLDAFKLNGAEIFLCCDPSRSGPSDFHVIASPDH 60
VRNVLSYISCNCLFRYIICSYCHFFGVLQEKFKDLKAKGCNLIGPQCALFCAKEGRPLPQ 120
RGFTCCLAMDGLKVLASGFLVDEKVKIKELVTSMGGVLLSRASSDVNFVIVKNVLAAKYK 180
WALNKKPIVTLNWLHRCWNEHRVVPQEPYKIPPFSGLTICVTRIPAGDKYKVARKWGHIQ 240
IVTRKWFQQSIDKKGDVGCFCILYILLLISDRTRWRPNGKAQEQNIDGCTARDSESEDND 300
LYLSDCRIFLLGFEASEMRKLAKLVRRGGGSRYMLLNERMTHIVVGTPSESEKREARSVA 360
ASGVIQVVIPSWLEDCDREKKEIPVHNIYTANHLILPRDSACLTKGSFARMSSMEQTKNT 420
HDQTMVYDSSSRSINVSNGPATLLGKNKEAMQEFGRKDEIHTGRKIVSPTQKETLIQLVT 480
CESKEQRSIQCEFSGQNDQERKSSVFKGETFCFSHSFPEDRTPLPGFESLCICSSQHNEK 540
NVELLRNLSVVLGADFVERLTRKVTHLICNFAKGDKYVRASKWGIISVTPDWLMNVLDRE 600
LTTQDREAGSQFHTQFVPMASRDSMSLPVSHSEDREKIQSFAGKSGCGKGEVYNRLGEIG 660
KEQTFPSKKAKLLRDGQESDVFPVRELPSNCDRSSHSGDGIVTGYDVASGREVPDVADTM 720
KDLLEQTSKHFSTSEQYNTGNHSVTGLSRHWINRVHKNDDMGSPPGDATTALTETLVRRR 780
QNHRLYITLALIILYLEDSIHNLKLRSYLQISRNLQVVGYEEDLSGRQMLIDRVRTRSSL 840
T
```

Fig. 2 Proposed F2P24.3 protein sequence. The three BRCT domains are indicated by *highlighted bold letters*. *Arrowheads* Positions of exon/intron junctions

Alignment with the partial cDNA of *MEI1* suggests that the first intron in the composite and GenBank versions of *F2P24.3* were incorrectly predicted, and there are no obvious intron donor/acceptor sites farther upstream. However, eliminating the first intron disrupts the open reading frame of *F2P24.3*. The open reading frame can be restored by insertion of a T nucleotide at position 2,827, a possible error in the WS sequence. By eliminating the first intron and adding the nucleotide T, *F2P24.3* would encode a peptide of 842 amino acids and would have 11 introns (Fig. 1). The larger version of F2P24.3 has three BRCT domains (Fig. 2). It is worth noting that even when numerous other possible versions of the F2P24.3 protein sequence were used in the analysis, three BRCT domains were always predicted. The F2P24.3 amino acid sequence predicted in GenBank has three typical BRCT domains and one shorter BRCT domain of half the typical length. Thus, it is highly likely that there is indeed a BRCT domain-containing protein encoded in this region. Unfortunately, except for the

cDNA reported by He and Mascarenhas (1998), there is no other cDNA available in any EST database that can be used to validate the proposed *F2P24.3* structure; this is reflective of the probable low abundance of the mRNA. As noted by He and Mascarenhas (1998), screening of an inflorescence cDNA library was not successful and it was only possible to isolate the cDNA they obtained via a PCR-based approach.

The *mei1* mutant phenotype was complemented by an 8.0 kb genomic fragment (He and Mascarenhas 1998). The coding region (including introns) of the newly proposed *F2P24.3* gene is 5,957 bp, and it is entirely contained within the 8.0 kb fragment used for their complementation experiment (Fig. 1).

Blast searches revealed that the 842 amino acid version of F2P24.3 is most similar to a human protein (accession number D87448) that is similar to the rad4+/cut5+ product in *Schizosaccharomyces pombe* (Nagase et al. 1996a, b), and to a human DNA topoisomerase II binding protein (accession number XP003014). The similarity (38%) is largely due to the conserved BCRT domains in these proteins. BRCT-domains were first described in the protein encoded by the human BRCA1 gene, mutations in which result in susceptibility to breast and ovarian cancers (Feunteun and Lenoir 1996). BRCT domains are flexible modules that can interact with other BCRT domains, with non-BRCT proteins, and with DNA strand breaks (Bork et al. 1997; Huyton et al. 2000). Although BRCT domain-containing proteins seem to have diverse functions, their participation in DNA damage-responsive checkpoints appears to be a unifying theme (Bork et al. 1997). In *mei1*, the chromosome fragmentation seen in early diplotene is the first noticeable defect (Ross et al. 1997). It is therefore reasonable to hypothesize that *MEI1* is involved in repair of double-strand breaks during meiotic recombination. Such a function would be consistent with *MEI1* encoding a BRCT-domain-containing protein. If this is true, it would suggest that the mechanism of repair of double-strand breaks during meiotic prophase I is likely to be conserved in yeast, mammals, and higher plants.

In conclusion, it is possible that *MEI1* is *F2P24.3* and therefore encodes a protein with BRCT domains. However, confirmation of both the newly proposed *MEI1* identity and the structure of *F2P24.3* will require further experiments. In *Arabidopsis*, there are at least ten genes that are predicted to encode proteins with between one and four BRCT domains (Butcher et al. 2001). It will be interesting to determine the roles that these proteins play during plant development.

Acknowledgement This work was supported by United States Department of Agriculture CRIS 5335-21000-011-00D.

References

Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB J 11:68–76

Brown JWS, Smith P, Simpson CG (1996) *Arabidopsis* consensus intron sequences. Plant Mol Biol 32:531–535

Butcher SJ, Coleman MJ, Bowater RP (2001) Investigating BRCT domains in *Arabidopsis thaliana*. In: Abstracts of the 674th Meeting of the Biochemical Society, 11-13 July, Dublin, Eire

Feunteun J, Lenoir GM (1996) BRCA1, a gene involved in inherited predisposition of breast and ovarian cancer. Biochem Biophys Acta 1242:177–180

Garcia-Diaz M, Dominguez O, Lopez-Fernandez LA, de Lera LT, Saniger ML, Ruiz JF, Farraga M, Garcia-Ortize MJ, Kirchhoff T, del Mazo J, Bernad A, Blanco L (2000) DNA polymerase lambda (pol lambda), a novel eukaryotic DNA polymerase with a potential role in meiosis. J Mol Biol 301:851–867

He C, Mascarenhas JP (1998) *MEI1*, an *Arabidopsis* gene required for male meiosis: isolation and characterization. Sex Plant Reprod 11:199–207

He C, Tirlapur U, Cresti M, Peja M, Crone DE, Mascarenhas JP (1996) An *Arabidopsis* mutant showing aberration in male meiosis. Sex Plant Reprod 9:54–57

Herrmann G, Lindahl T, Schar P (1998) *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. EMBO J 17:4188–4198

Huyton T, Bates PA, Zhang X, Sternberg MJ, Freemont PS (2000) The BRCA1 C-terminal domain: structure and function. Mutat Res 460:319–332

Magnard JL, Yang M, Chen YCS, Leary M, McCormick S (2001) The *Arabidopsis thaliana* gene *TARDY ASYNCHRONOUS MEIOSIS* (*TAM*) is required for the normal pace and synchrony of cell division during male meiosis. Plant Physiol 127:1157–1166

Nagase T, Seki N, Ishikawa K, Ohira M, Kawarabayasi Y, Ohara O, Tanaka A, Kotani H, Miyajima N, Nomura N (1996a) Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. DNA Res 3:321–329

Nagase T, Seki N, Ishikawa K, Ohira M, Kawarabayasi Y, Ohara O, Tanaka A, Kotani H, Miyajima N, Nomura N (1996b) Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain (supplement). DNA Res 3:341–354

Ross KJ, Fransz P, Armstrong SJ, Vizir I, Mulligan B, Franklin FCH, Jones GH (1997) Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines. Chromosome Res 5:551–559

The *Arabidopsis* Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815