Cloning and sequencing of an apparently recombinant promoter for napin gene from *Brassica campestris* genomic library and its evolutionary significance

Somnath Bhattacharya and Radha Kanta Mandal

Department of Biochemistry and Centre for Plant Molecular Biology, Bose Institute, Calcutta 700 054, West Bengal, India

Received 25 January 1999; revised 29 April 1999

From a genomic library of *Brassica campestris* (brown saron cv. B54), we have cloned and sequenced about 2 kb of upstream regulatory region from one of the 2S albumin-coding gene family. The sequence has several seed-specific promoter motifs. A sequence alignment of the 5' flanking regions of the available *Brassica* 2S storage protein genes showed that our sequence is a double crossover recombinant product of the two members of the napin gene family. A possible explanation of this fact is that *Brassica* species evolved through gene duplications and recombination from a common ancestor with fewer number of chromosomes and genes.

Among the various seed storage proteins, 2S storage protein (napin) is the major in the *Brassica* species. Napins are encoded by a multigene family with 4-7 numbers in *B. rapa* and 10-16 in *B. napus*. Nucleotide sequences of 2S coding regions showed about 90% homology within different *Brassica* species. With a view to construct a seed specific expression system for transgenic plant, the upstream region of 2S protein gene of *Brassica juncea* and *B. oleracea* have been amplified from genomic DNA using oligonucleotide primers corresponding the napin gene (published) and consequently sequenced. We are aware of sequence data for six different napin promoters. Among them three upstream sequences of napin gene (*B. juncea*, *B. oleracea*, *B. carinata*) showed close homology (over 90%) with that of a sequence of *B. napus* by Joseffson et al. The upstream sequences from *B. rapa* isolated by Kridl et al. showed close homology with that of another napin promoter sequenced by Scofield and Crouch. But in different Indian morphotypes (saron, toria) of *B. campestris* (syn *B. rapa*) this upstream regulatory region is not amplified by the above said oligonucleotide primers. Here we describe the isolation sequencing and evolutionary significance of a genomic napin clone from oleiferous *B. campestris*, one of the diploid ancestral parents of both *B. napus* and *B. juncea*.

### Materials and Methods

#### Plant material

Seeds of *Brassica campestris* var. B54 was obtained from Pulses and Oil Seeds Research Station, Berhampore, West Bengal, India.

#### Chemicals

Enzymes were obtained from Bangalore Genei, India or Bethesda Research Laboratories (BRL) USA. Agarose, ethidium bromide, dNTPs, X-gal etc. were obtained from Sigma, USA. [α-32P] dATP and [α-35S] dATP used for Southern blotting and sequencing were supplied by Board of Radiation and Isotope Technology, BARC, Mumbai, India. Hybond N+ and Hybond N were obtained from Amersham, England. X-ray films used for autoradiography were from OROW or Agfa, India.

#### DNA isolation, genomic library construction and screening

The high molecular nuclear DNA of *B. campestris* cv. B54 was isolated following a rapid and simple method. The same nuclear DNA was partially digested with EcoRI and ligated into the EcoRI site of EMBL3 phage vector. Recombinant bacteriophage were transfected into P2392 E. coli cells as described. Plaques were lifted onto Hybond N membrane in duplicate and screened with a nick translated PstI fragment of the napin cDNA (pN2) clone labeled with [α-32P] dATP. Some of the positive plaques were picked for recombinant phage screening. The phage DNA was isolated and used for construction of genomic library in EMBL3 vector. About 2 kb of sequence of the phage DNA was sequenced for each positive clone. Among the positive plaques generated, one phage sequence was identified to be a recombinant product of the two members of the napin gene family. The upstream regulatory region of the phage sequence contained all the regulatory motifs present in the two napin genes. The phage sequence was used for the construction of genomic library. Several positive clones were obtained from this library. The positive clones were sequenced and one of the nucleotide sequences was found to be a double crossover recombinant product of the two members of the napin gene family.
beled with $[^\alpha\text{-}32\text{P}]d\text{ATP}$. Two positive recombinant phage clones were obtained and purified according to a standard protocol\(^1\). They were subsequently found to contain overlapping genomic inserts and designated as BeGL5 and BeGL6. Using proper restriction enzymes, the BeGL6 derived overlapping genomic fragments were subcloned into the pUC18 vector.

**Restriction mapping and DNA sequence analysis**

Approximately a 6 kb recombinant phage clone was subcloned into the EcoRI site of pUC18 plasmid vector. This 6 kb fragment was restriction digested by \(\text{HindIII}, \text{HaeIII}\) and \(\text{XhoI}\) separately and blotted on Hybond N+ membrane according to standard procedure. Blots were hybridized with the labelled 537 bp \(\text{XhoI/XhoI}\) DNA fragment containing coding region and 1137 bp EcoRI/HindIII fragment containing promoter region of napin gene from \(B.\) juncea\(^7\). Blots were prehybridized and hybridized in solutions containing 5X SSPE, 0.1% sodium pyrophosphate, 0.25% SDS, 0.02% polyvinyl pyrolidone, 0.02% BSA (fraction V), 10 \(\mu\text{g/mL}\) yeast tRNA, 50% deionised formamide at 40°C. Washes were in 0.1% SSC and 0.1% SDS at 40°C.

A 2kb HindIII fragment was chosen as a promoter region for sequencing. This fragment was subcloned into various smaller pieces through \(\text{ExoIII-mung bean nuclease treatment according to suppliers manual (Stratagene)}\) into the pBsk(+) vector. Sequence of this upstream regulatory region was determined by the chain terminating dideoxy method using sequanase version 2.0 (modified T\(_7\) polymerase) DNA sequencing kit (USB, Cleaveland, USA) and computer aligned by UW-GCG program.

**Results and Discussion**

**Isolation and characterization of a promoter from \(B.\) campestris genomic library**

Using napin cDNA (pN2) as a probe, 6.2 kb genomic clones from EMBL3 phage library of \(B.\) campestris was isolated. To identify the 5' upstream region from the genomic clone this 6.2 kb fragment was restriction digested by \(\text{HindIII}\) and \(\text{HaeIII}\) and Southern hybridized separately by napin promoter\(^5\) and coding region\(^7\) (from \(B.\) juncea) as probes. From the autoradiogram (Fig. 1a, b) it is found that only a 2 kb HindIII fragment and 1.4 kb HaeIII fragment are present when hybridized with promoter region. But two HindIII fragments (2.2 kb and 1.1 kb) and same size of a HaeIII fragment (1.4 kb) are present when hy-

---

**Fig. 1**—Southern blot analysis of a 6 kb genomic fragment isolated from clone BeGL6. (Undigested fragment (lanes 1); digested with \(\text{HindIII}\) (lanes 2); and \(\text{HaeIII}\) (lanes 3). The blot was probed with (a), EcoRI/HindIII fragment (1137 bp) encompassing napin promoter of \(B.\) juncea\(^7\); (b), \(\text{PstI}\) fragment of the napin coding region (537 bp) of \(B.\) juncea\(^7\). The numbers indicate \(\lambda\text{HindIII DNA size markers in kb. (c), The restriction map of clone as derived from above Southern blot analysis results is depicted (a and b). The filled box indicates the coding region. H. HindIII; Ha, HaeIII. The size of 1 kb is indicated.}\)
bridized with coding region (Fig. 1b). So from the banding pattern it is ascertained that both the enzymes have at least one restriction site each in 5’ upstream region and coding region. As evidenced from the autoradiogram, one HindIII fragment from 5’ upstream region is missing (lane 3 of Fig 1a). The reason behind this is discussed later on after complete sequencing of the 2 kb 5’ upstream regulatory region. The restriction map of the clone BeGL6 derived from the Southern blotting results is depicted in Fig. 1c.

The 2 kb HindIII fragment and a 1.4 kb Haelll fragment were subcloned into pBSk(+) vector for getting 5’ upstream regulatory region of B. campestris. A 2 kb HindIII fragment was further fragmented into five smaller pieces by exonuclease treatment and cloned into the same pBSk(+) vector for sequencing. The complete sequence of 2084 bp upstream region was submitted to the EMBL data bank under the accession number Y13108 and is given in Fig. 2 (upper line).

The sequence has several universal and seed specific transcription regulatory motifs. A possible TATA box lies from nt. 2020-2026 and a putative CAAT box from nt. 2001-2005. The hexameric, G-box motif (CACGTG) and seed specific alternate purine pyrimidine motif (CATGCA) were approx. 65 and 77 nt. upstream respectively from the transcription start point deduced by analogy with that of napin gene^5. A four base repeat CATT is found ten times as shown by double underlining.

Comparison of B. campestris with other Brassica species and their evolutionary significance

One hundred per cent DNA homology up to 68 bp upstream from the transcription start site was observed among the three sequences (Fig. 2). Moderate amount of homology (~70%) was found up to 200 bp upstream after the above mentioned region i.e. upto

```plaintext
Be2bP
AAGCTTAATT CT...TTTCTG AAGCTTTCTG GCTGCTTTCTG 100
Br2bP
AAGCTTAATT CTCTTTTCTG GCTGCTTTCTG 100

GTGTAATTTT GTGCTTTCTG GCTGCTTTCTG 100
TCTCAATTCT CATGACTCAGC 295
GTTTTTACG TACACGAGCT

ACGTTCG CAGTTTCA CAGTTTCA CAGTTTCA CAGTTTCA CAGTTTCA CAGTTTCA 295
GTTTTTACG TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT 295
GTTTTTACG TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT 295
GTTTTTACG TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT 295

ACGTTCG CAGTTTCA CAGTTTCA CAGTTTCA CAGTTTCA CAGTTTCA CAGTTTCA 295
GTTTTTACG TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT 295
GTTTTTACG TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT 295
GTTTTTACG TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT TACATTTTTT 295
```

(Contd.)
Fig. 2.—Alignment of 5' upstream sequences of 25 protein genes from R. campinestris in the present study denoted by Bc2SP (upper lines), B. rapa (middle lines) and B. juncea (lower lines). [Capital letters indicate sequence identity at least in three successive nucleotides. Base differences are indicated by small letters where at least three consecutive bases are dissimilar. Dots represent gaps for missing nucleotides from one sequence to allow maximum homology alignment. The TATA box (solid line), CAAT like box (dotted line), G-box (dashed line), seed specific CATGCCCA asterisk under it), four base repeats like CATT (double lines) and presumed transcription start point (arrow head) are indicated. The sequence of Bc2SP was submitted to the EMBL data bank under the accession number Y13108].
-270 bp among those sequences. After this region, the promoter under study showed high degree of homology (>95%) with the B. rapa or B. napus promoter upto -420 bp (Fig. 3) and in this region, no homology is found with the B. juncea promoter. It is interesting to note that our cloned napin gene promoter has approx. 90% homology with the B. juncea or B. napus promoter pN36 (ref. 5) in the region 1312 to 1655 nt. From Fig. 2 it is clear that -420 to -741 nt. region of the promoter shows close homology (~95%) with that of B. juncea from -331 to -648 nt. region. Remaining upstream region 1 to 1311 i.e. the region beyond upstream -741 bp has more than 90% homology with the B. rapa promoter, BcNA1 (ref. 4) or B. napus promoter gNa (ref. 6) but absolutely no homology with B. juncea promoter Fig. 3). It has been found from restriction mapping and sequence analysis that there are three HaeIII sites on the napin gene; one at 3' end of the coding region, one at -784 nt. and other at -1665 nt. So when we used B. juncea promoter or coding region as probes only the 1.4 kb HaeIII fragment (-784 to 3' end of coding region) was lighted up in both the autoradiograms (Fig.1a and b). This is due to the fact that B. juncea promoter has no homology above the -741 nt. region (Fig. 3) of B. campestris promoter. Therefore, it is concluded that 5' upstream region of B. campestris napin gene, we have sequenced might be a double cross over recombinant product of the two members of the napin gene family. The recombination may have occurred during evolution of B. campestris through interspecific or intraspecific crosses among the wild species of Brassica as evidenced from the cytological or RFLP analysis. Three diploid species of Brassica evolved through two evolutionary pathways. One pathway gave rise to B. nigra and another common pathway to the diploid cultivated species, B. oleracea and B. campestris. Thus, according to the previous study, the diploid cultivated species, B. oleracea (CC) and B. campestris (AA) are secondary polyploids and share common ancestral genome of chromosome number six. Conserved regions of the Brassica lineage (i.e. shared linkage between species) evolved through duplicated loci, and hence served as orthologous linkage blocks. These orthologous linkage blocks allow sexual recombinations between genomes. So being an essential gene, this recombinant member of the napin gene isolated from B. campestris evolved during change of ascending chromosome numbers from a common ancestor of basic chromosome number below nine. Thus the napin gene evolved much earlier than the time of divergence of cultivated diploid and amphidiploid Brassica species. It could be interesting to compare the napin promoter sequences isolated from B. nigra to shed further light on the origin of Brassica amphidiploids. Further studies along these lines are in progress.

Acknowledgement
The research was supported by a UNDP/FAO project (IND/87/017) and CPMB, DBT, Govt. of India. Thanks are due to Drs N Das and S Ghosh for providing the genomic library and to Distributive In-
formation Centre, Bose Institute for help in sequence analysis.

References
1. Lounerdal B & Janson J C (1972) Biochim Biophys Acta, 278, 175-183