J. Biochem. 2010;148(6):651–658 doi:10.1093/jb/mvq100

JB THE JOURNAL OF BIOCHEMISTRY

Characterization and heterologous expression of a new matrix attachment region binding protein from the unicellular green alga *Dunaliella salina*

Received May 19, 2010; accepted August 24, 2010; published online October 6, 2010

Tianyun Wang[1,2], Guiqin Hou[3], Yafeng Wang[1] and Lexun Xue[1,*]

[1]Laboratory for Cell Biology, The First Affiliated Hospital, Zhengzhou University, 40 Daxue Road, Henan 450052; [2]Department of Biochemistry and Molecular Biology, Xinxiang Medical University, Henan 453003; and [3]School of Pharmaceutical Sciences, Zhengzhou University, Henan 450001 China

*Lexun Xue. Tel: +86 371 66658332, Fax: +86 371 66997182, email: xuelx@371.net

Although interactions between the nuclear matrix and special regions of chromosomal DNA called matrix attachment regions (MARs) are implicated in various nuclear functions, the understanding of the regulatory mechanism of MARs is still poor. A few MAR-binding proteins (MARBP) have been isolated from some plants and animals, but not from the unicellular algae. Here, we identify a novel MAR-binding protein, namely DMBP-1, from the halotolerant alga *Dunaliella salina*. The cDNA of DMBP-1 is 2322-bp long and contains a 1626 bp of an open reading frame encoding a polypeptide of 542 amino acids (59 kDa). The DMBP-1 expressed in *Escherichia coli* specifically binds A/T-rich MAR DNA. The DMBP-1 fused to green fluorescent protein appears only inside the nuclei of Chinese hamster ovarian cells transfected with the pEGFP–MBP, indicating that the protein is located in the nuclei. The findings mentioned above may contribute to better understanding of the nuclear matrix–MAR interactions.

Keywords: *Dunaliella salina*/matrix attachment region/MAR binding protein/nuclear matrix.

Abbreviations: ARBP, attachment region binding protein; DMBP, *Dunaliella salina* MAR-binding protein; DSM, *Dunaliella salina* MAR; HMG, high mobility group; IPTG, Isopropyl β-D-1-thiogalacto-pyranoside; MBP, MAR-binding protein; NuMA, nuclear mitotic apparatus; OD, optical density; PCR, polymerase chain reaction; pEGFP, enhanced green fluorescent protein; RT, reverse transcription; SAR, scaffold attachment region; SATB1, special AT-rich sequence binding protein 1.

Nuclear matrix is the dynamic fibroranular structure forming the skeletal framework that surrounds and penetrates the interphase nucleus. The nuclear matrix is defined biochemically as the insoluble material that remains after extraction of the nuclei with high-salt solutions and treatment with DNases after removal of chromatin and soluble nuclear proteins (*1*). It has been demonstrated that the nuclear matrix is implicated in most nuclear functions including replication, repair, transcription, RNA processing and RNA transport, etc. (*2*).

Anchorage of these loops to the nuclear matrix involves specific stretches of DNA termed matrix attachment regions (MARs) (*1*). MARs consist of AT-rich sequences extending over at least a few hundred base pairs and various AT-rich motifs as well as structural motifs such as base-unpairing regions and intrinsically curved portions, and are thereby thought to be organized into topologically constrained loops, each of which represents a sort of functional or structural domain (or both) (*3*, *4*). MARs from a range of eukaryotic organisms have been isolated and sequenced, ranging from 0.2 to several kilobases in length and occurring in non-coding DNA and, generally, have a >70% A/T-rich bases. Although MARs have no consensus base sequences, a number of motifs commonly occur, such as the A-box (AATAA AYAAA), T-box (TTWTWT TWTT), the unwinding sequence (AATATATTT) and the *Drosophila* topoisomerase II consensus cleavage site (GTNWAYATTNATNNR). MARs, in addition to their structural role in defining the anchorage points of DNA loops, have been proposed to be involved in a number of genome functions, *e.g.* preventing the linear propagation of chromatin structural changes and shielding genes from *cis*-elements lying in other loops (*5*, *6*), enhancing the transgene expressions (*7*–*11*) and acting as origins of replication and involvement in chromosome condensation (*12*).

A variety of proteins that interact with MARs have been identified in animals including maintenance of the genome stability-related protein: DNA topoisomerase II (*13*), filament proteins: lamins and NuMA (*14*), an RNA-binding protein: hnRNP-U/SAF1, which is involved in RNA processing (*15*, *16*), a methylated CpG-binding protein: ARBP (*17*), a tissue-specific transcription factor: SATB1 (*18*) and architectural chromatin proteins: histone H1 and HMG-I/Y (*19*).

One kind of filament-like proteins tomato MFP1 and the MFP1-interacting protein MAF1 from tomato and tobacco have been identified at the peripheral regions of the plant nuclei. MFP1 protein locates in both the nuclei and chloroplasts, being consistent with the findings in onion and the dicot species (*20*–*22*). A novel protein AHM1 was identified in the inner nuclear matrix of wheat, which has the special combination of a J domain-homologous region and a Zinc finger-like motif and an AT hook (*23*). Two

MAR-binding proteins, MARBP-1 and MARBP-2, which have a significant homology to yeast nucleolar proteins, were found in the nuclear matrix of pea (*24*).

However, the understanding of molecular basis of the nuclear matrix–MAR interactions is still fragmentary, and the investigations on MARs have been carried out mostly in higher plants and in animals. As the only eukaryotic Phylum lacking histone and nucleosome, unicellular Dinoflagellates still have a residual nuclear matrix similar to that of vertebrates and higher plants (*25*,*26*). *Dunaliella salina* is one of the most extremely halotolerant eukaryotes and is able to survive in a variety of salt concentrations, ranging from 0.05 to 5 M sodium chloride (*27*). The simple and cheap culture of *D. salina* means that it has a great potential in bioengineering for producing valuable polypeptides and proteins (*10*). In addition, the bias of codon usage of algae is more similar to humans than *Escherichia coli* and yeast (*28*). The nuclear matrices and three MARs from *D. salina* have been isolated in our previous experiments (*29*), yet we do not understand the mechanism of the nuclear matrix–MAR interactions in *D. salina*.

In the present study, through screening nuclear matrix-localized MAR-binding proteins, we isolated and characterized a novel type of MAR-binding protein from the unicellular green alga *D. salina,* which may contribute to better understanding of the nuclear matrix–MAR interactions.

Materials and Methods

Algae strain and culture

Dunaliella salina strain UTEX 1644 TEOD was purchased from the University of Texas, USA, which was grown in batch cultures in liquid Ben-Amotz medium at 26°C under continuous illumination of 4500 Lux, with a 12-h light/day as described by Wang *et al.* (*10*). *D. salina* cells at the logarithmic phase were transferred to a solid medium containing 0.8% agar and then cultured for 3 weeks until a few individual colonies appeared. Subsequently, one single colony was picked out and transferred to the liquid medium mentioned above for further culture.

RNA extraction and cloning of the DMBP-1 cDNA fragment

Cells of *D. salina* from the log phase of growth were harvested by centrifugation at 4,000*g*, and then RNA extraction was carried out using TRIzol Reagent (GIBCO BRL) according to the manufacturer's instruction. The quality and concentration of RNA were examined by electrophoresis on EB-stained agarose gels and spectrophotometer analysis.

First of all, a group of MAR-binding proteins and mRNA sequences from *pea, tobacco, Arabidopsis, yeast, Homo sapiens* and others were multi-aligned on Vector NTI Suite 6 to find their conserved amino acid sequences EWYGWHF and KYGLIY, and then to design two degenerate homologous PCR primers P1 (5′-GARTG GTAYGGNTGGCAYTT-3′) and P2 (5′-RTADATNARNCCR TAYTT-3′). A total RNA of ~2.5 μg was reverse-transcribed into cDNA and then PCR was carried out using the two degenerate primers. After agarose gel electrophoresis, the target DNA band was purified for sequencing.

3′- and 5′-RACE of the D. salina *DMBP-1 gene*

The 3′-RACE System for Rapid Amplification of cDNA Ends (TaKaRa, Dalian, China) was used for 3′-cDNA end amplification of the DMBP-1 gene. According to the sequencing result of DMBP-1 cDNA fragment cloned above, three primers were designed and synthesized for 3′-RACE of DMBP-1. RNA was reverse-transcribed into cDNA. The first PCR was carried out using primers A1 (5′-CA GCTGTTCGACTACCTGA-3′) and AP (5′-GACTCGA GTCGAC ATCG-3′). The second PCR was performed using A2 primer (5′-GG TCC TGGTTGGAGAGC TTGTA-3′) and the aforementioned AP primer. The system and conditions of PCR of the target DNA fragments were accomplished according to the previous methods described by Wang *et al.* (*10*).

The 5′-RACE System for Rapid Amplification of cDNA Ends (TakaRa, Dalian, China) was used for 5′-cDNA cloning. According to sequencing result of the 3′-cDNA end, four primers were designed and synthesized for 5′-RACE of DMBP-1. A quantity of 5 μg of total RNA extraction from *D. salina* were reversely transcribed using primer RT [5′-(P) GTCAGGTTTGGT-3′] at 30°C for 10 min, 50°C for 30 min and 80°C for 2 min, and then hybrid RNA was denatured and circulated according to the Kit protocol. Primers A1 and S1 (A1: 5′-TGGACACGATGGC TGCATCC-3′; S1: 5′-GG ACTTGATGAGGATCCGTGA-3′) were used for the primary amplification using 5 μl of circulated cDNA as a template in a total volume of 50-μl reaction system under the standard PCR conditions. After being confirmed by agarose gel electrophoresis, the diluted 50-fold PCR products were used as template for the nested PCR amplification using primers (A2: 5′-CACAATCTTGGTCAT CTCCG-3′ and S2: 5′-CAGCTGTTCGACTACCTGA-3′). The nested PCR products were subjected to electrophoresis, gel extraction, ligation and transformation into *E. coli* and sequencing as mentioned above.

Cloning and analysis of the full-length cDNA of the DMBP-1

By comparing and aligning sequences of the 3′- and 5′-RACE products, the full-length cDNA sequence of the DMBP-1 was deduced and obtained through RT-PCR using primer P1(5′-ATGGTTCT C GTCTTGTTCGAGAC-3′) and P2 (5′-CTCAGCGGC CTTCTTCT TCTT-3′). PCR was carried out as the methods described above.

After agarose gel electrophoresis detection and gel extraction, DNA of the target bands was ligated to pGEM-T Easy vector (Promega, USA), and sequenced using T7/SP6 primers (Sangon, Shanghai, China). Sequence alignments, open reading frame (ORF) translation and molecular calculation of predicted protein were carried out using Vector NTI Suite 6. BLAST was conducted at the NCBI server (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi), while structural analysis of the predicted DMBP-1 protein was carried out on the website (http://cn.expasy.org). Phylogenesis was analysed according to Clustal V method of MegAling program in DNA STAR package.

Analysis of DMBP-1 expression using semi-quantitative RT-PCR

A total of RNA isolated from *D. salina* cells was prepared on Days 1, 2, 4, 8, 15, 20, 24 and 30, respectively, after being cultured under the 12-h light and dark/day. Semi-quantitative RT-PCR was performed using SYBR RT-PCR Kit (Takara, Dalian, China) according to the manufacturer's instructions. The primers for the DMBP-1 gene were designed as follows: 5′-TCGCGATACAAGCTCAAGTT CAG-3′ (sense) and 5′-GTGGATGCTGGCTGCTTGGCCTGG-3′ (antisense). The primers for 18S rRNA were 5′-TTGGGTAGTCGG GCTGGTC-3′ (sense) and 5′-AGGTCCACTTGCCCAACACTCC G-3′ (antisense). The PCR conditions were as follows: after heating at 95°C for 10 min, followed by 30 cycles of 50 s at 95°C, 30 s at 60°C and 40 s at 72°C, and a final extension at 72°C for 5 min. As targeted DNA was amplified, the fluorescence of the end of each cycle of RT-PCR was monitored by iCycler iQ (Bio-Rad Laboratories Inc., Hercules, CA, USA). The targeted DNA amplified specifically was confirmed by electrophoresis and sequencing. The relative abundance of 18S rRNA was used as an internal standard.

Expression of the recombinant DMBP-1 in E. coli

The DMBP-1 cDNA of *D. salina* was amplified by PCR using the primers 5′-CGGCATATGATGGTTCTCGTCTTGTTC GA GAC-3′ and 5′-CGGGAATTCCTCAGCGGCCTTCTTCT TCTT-3′. The PCR products cut with enzymes *Nde*I and *Eco*RI were cloned into the prokaryotic expression vector pET28a (+).

A single colony of *E. coli* BL21 containing the DMBP-1 was grown in 5 ml of L-broth medium containing 50 μg/ml kanamycin for 12 h at 37°C at 200*g*. The 50-μl cultures were inoculated into 50 ml L-broth media containing kanamycin and then grown until the OD 600 value achieved was 0.4–0.6. Subsequently, the cultures were harvested at 10,000*g* for 10 min after being induced by 0.2 mM IPTG at 37°C for another 4 h. Each pellet was resuspended in 500 μl loading buffer prior to 10% SDS–PAGE.

Southern–Western Blots

Proteins resolved by 10% SDS–PAGE were transferred onto nitrocellulose (0.45 μm) by electro blotting in 50 mM Tris, 380 mM glycine, 20% methanol and 0.1% SDS. Binding was performed using a modified method according to the report by Morisawa *et al.* (*23*) and von Kries *et al.* (*16*). In brief, binding was carried out in a DNA-binding buffer containing 20 mM Tris–HCl (pH 7.5), 150 mM NaCl, 4 μg/ml sheared *E. coli* DNA and 2 ng/ml end-labelled MAR fragment in the presence or absence of competitors, unless specified. After incubation at room temperature (RT) for 15 h, the filters were washed three times with the DNA-binding buffer and then autoradiographed. The probes used included a 0.88-kb fragment of β-globin MAR (*30*), a 0.44-kb fragment containing the mouse lgκ locus MAR (*31*) and a 0.50-kb fragment corresponding to DSM 2 MAR from *D. salina* (*10*). The blots were blocked for 30 min (5% skimmed milk powder, 10 mM Tris–HCl pH 7.5, 50 mM NaCl, 2 mM EDTA pH 8.0). The blots were then incubated for 1 h in binding buffer containing Dig end-labelled probe and the appropriate concentration of competitor DNA. Finally, the protein blots were rinsed three times for 10 min in binding buffer before being dried and exposed to X-ray film.

Expression and subcellular localization of the DMBP-1 in CHO cells

The *D. salina* DMBP-1 cDNA amplified by PCR was cloned into the expression plasmid pEGFP (Clontech) to generate a plasmid pEGFP-MBP, and then the pEGFP-MBP would be transfected into CHO cells by Lipofectamine 2000. Western blots were performed to determine the expression of the fusion proteins and their cellular localization. Briefly, a total of recombinant proteins expressed as DMBP-1 in CHO cells was run on 12% SDS–polyacrylamide gel and then transferred to nitro-cellulose membrane. Blots were hybridized with mouse anti-hexahistidine monoclonal antibody at 4°C overnight. After being washed, the blots were incubated with horseradish peroxidase-conjugated secondary antibody at RT for 5–10 min and then visualized using the Amersham enhanced chemiluminescence detection system.

Results

Amplification of 3′- and 5′-ends of DMBP-1 cDNA

Agarose gel analysis revealed that amplification with degenerate primers resulted in an ∼500 bp of DNA band. BLAST-n analysis indicated that the DNA fragment of PCR had wide similarities to known MAR-binding protein genes from *pea, tobacco, Arabidopsis, yeast, Homo sapiens* and others.

Electrophoresis analysis of primary PCR product of 5′-RACE showed only smear, not specific band. However, when the nested PCR was carried out using the PCR product as a template, a specific band of ∼800 bp appeared. There was a similarity between this sequence and the 3′-cDNA end by BLAST analysis. Based on the 272-bp overlap between the 5′ and 3′-fragments, the full-length DMBP-1 cDNA was 2232 bp in length [poly (dA) not included].

Sequence and characterization of DMBP-1

As anticipated, the specific 2232-bp DNA band was amplified by PCR using full-length primers P1 and P2. Sequencing result of this sequence completely coincided with the deduced full-length cDNA sequence (Fig. 1). The cDNA of DMBP-1 possessed a 1623 bp of an ORF from 105–1729 bp of the sequence, including a 105 bp of 5′-UTR and a 473 bp of 3′-UTR. The 3′-UTR possessed a typical polyadenylation signal TGTAA at position 15 upstream of the poly (A)-tails.

```
1     ACCTGACATTACTTTTGGTTTGGTTAATAGGTTGATGCTTTAGTGTCCTTTGCTAGCAAC
61    GACTGCCCTTTGCCAGCAAAGTAGCTTTGCAGTCTGCTGGCAAGCATGGTTCTCGTCTTG
                                                  M  V  L  V  L
121   TTCGAGACTGCTGGAGGCTTGGCGTTGTTCAAGGTGCTCCAAGAAGGCAAACTGAAGCAG
       F  E  T  A  G  G  L  A  L  F  K  V  L  Q  E  G  K  L  K  Q
181   ACAGAGGATGTGAGCGCAGACTTTGGGACGCCCGAGCAGGCGCAAAAGATGGTCAAGCTC
       T  E  D  V  S  A  D  F  G  T  P  E  Q  A  Q  K  M  V  K  L
241   AAGGCGTTCAGCAAGTTCCAGGACACGACGGAGGCGATGGAGGCCGCCACGGCCCTGGTG
       K  A  F  S  K  F  Q  D  T  T  E  A  M  E  A  A  T  A  L  V
301   GACAGCAAGCTCAGCAAGCCCATGAAGAAGTTCCTGAGGAAGAACGTGGATGGAGAGGAG
       D  S  K  L  S  K  P  M  K  K  F  L  R  K  N  V  D  G  E  E
361   CTGGCCATCCTCGATAAGAAGCTGGGCGGCATCGTGCAGGAGAAGCTCGGCATTCCTTGT
       L  A  I  L  D  K  K  L  G  G  I  V  Q  E  K  L  G  I  P  C
421   GTGTACAGCAACGCGGTGCTGGAGCTCACGCGTGGCATCAGGAACCAGCTGCAGGGCTTG
       V  Y  S  N  A  V  L  E  L  T  R  G  I  R  N  Q  L  Q  G  L
481   ATTAGTGGCCTGTCTGGTGTGGACCTGAAGCCCACGTCCCTGGGTCTATCTCACAGCTTG
       I  S  G  L  S  G  V  D  L  K  P  T  S  L  G  L  S  H  S  L
541   TCGCGATACAAGCTCAAGTTCAGCCCAGACAAGGTCGACACCATGATTGTGCGAGCCATT
       S  R  Y  K  L  K  F  S  P  D  K  V  D  T  M  I  V  R  A  I
601   GGGCTGCTGGACGGCCTGGACAAGGAGCTGAACACGTATGCCATGCGTGTGCGTGAGTGG
       G  L  L  D  G  L  D  K  E  L  N  T  Y  A  M  R  V  R  E  W
661   TATGGCTGGCACTTCCCGGAGATGACCAAGATTGTGAACGACAACATCGCATACGCCAAG
       Y  G  W  H  F  P  E  M  T  K  I  V  N  D  N  I  A  Y  A  K
721   GTCGTCAAGCTGATGGGCACCCGTGACCAGGCCGCCTCCCACGACTTTTCGGGCATCATC
       V  V  K  L  M  G  T  R  D  Q  A  A  S  H  D  F  S  G  I  I
781   GAGGAGGACACAGAGCAGCACCTAAAGGATGCAGCCATCGTGTCCATGGGCACAGAGATC
       E  E  D  T  E  Q  H  L  K  D  A  A  I  V  S  M  G  T  E  I
841   AGCCAAGAGGACTTGATGAGGATCCGTGAGCTAGCAGACCAGGTGATTGACCTGTACGCG
       S  Q  E  D  L  M  R  I  R  E  L  A  D  Q  V  I  D  L  Y  A
901   TACAGAGGCCAGCTGTTCGACTACCTGAAGTCGCGAATGAACGCCATCGCACCAAACCTG
       Y  R  G  Q  L  F  D  Y  L  K  S  R  M  N  A  I  A  P  N  L
961   ACGGTCCTGGTTGGAGAGCTTGTAGGTGCACGCCTCATCTCCCGTGCCGGCTCCCTCATC
       T  V  L  V  G  E  L  V  G  A  R  L  I  S  R  A  G  S  L  I
1021  AACCTGGCCAAGCAGCCAGCATCCACCGTCCAAATCCTAGGAGCAGAGAAGGCCTTGTTC
       N  L  A  K  Q  P  A  S  T  V  Q  I  L  G  A  E  K  A  L  F
1081  AGGGCCTTGAAAACCAAGCACGAGACCCCCAAGTACGGCCTCATCTACCACGCATCGCTC
       R  A  L  K  T  K  H  E  T  P  K  Y  G  L  I  Y  H  A  S  L
1141  ATCGGCCAGAGCTCATCCAAATACAAGGGCAAGGTGTCACGTGTGCTGGCCGCCAAGTGT
       I  G  Q  S  S  S  K  Y  K  G  K  V  S  R  V  L  A  A  K  C
1201  GCGCTGGCAACGCGCGTGGATGCCCTTGGGGAGGGTGAAGACGCACAGATCGGTATTGAC
       A  L  A  T  R  V  D  A  L  G  E  G  E  D  A  Q  I  G  I  D
1261  GCACGTTCCAAGGTGGAGGCGCGCCTGCGCCGGCTGGAAGGCAAGACGCTCATCACTGAT
       A  R  S  K  V  E  A  R  L  R  R  L  E  G  K  T  L  I  T  D
1321  GGTGGCAAGGCAAAGGGCAAGGAGCAGCCCGCGCCCTACGACAAGACCAAGCAGCAGAGC
       G  G  K  A  K  G  K  E  Q  P  A  P  Y  D  K  T  K  Q  Q  S
1381  CCGGCTGCAGGGCTATCCACAGTGCCCAAGGCTTACAACGCGGACGCGGATGTAGCGGTG
       P  A  A  G  L  S  T  V  P  K  A  Y  N  A  D  A  D  V  A  V
1441  CCAGAGAAGAAAAAGAAGAAGAAGGACAAGAGCGGCGAGCAGGCCGGCCCCAGCGAAGAG
       P  E  K  K  K  K  K  K  D  K  S  G  E  Q  A  G  P  S  E  E
1501  CAGCCACAGCAGCCCCAGCAGGCGGACGGCCAGCAAGAAGGCGAGCAGCCAAAGAAGAAG
       Q  P  Q  Q  P  Q  Q  A  D  G  Q  Q  E  G  E  Q  P  K  K  K
1561  AAGAAGGAGAAGAAGGCGGCGGAGGCAGCAGCAGGAGACGCGCAGAATGGTGGTGGTGGC
       K  K  E  K  K  A  A  E  A  A  A  G  D  A  Q  N  G  G  G  G
1621  GCGGAAGAGGAGGCGCCCAAGAAGAAGAAGAAGAAGGCCGAGGCTGGCGCTGCAGATGGC
       A  E  E  E  A  P  K  K  K  K  K  K  A  E  A  G  A  A  D  G
1681  GGTGCTGCTGAGGAGGCGCCCAAGAAGAAGAAGAAGGCCGCTGAGTAAGCTGGTGTC
       G  A  A  E  E  A  P  K  K  K  K  K  K  A  A  E
1741  CTTCGGCCTCCTCCTCCTGCCTCTCCCTTTCCCTCCATCACCACGGAGTCTGCGAAAAAC
1801  CCTTCACCTCGAAGATCCCTCCTTCCTCTCTACCTCCTGTGTTGGCCTGTCCCATGGTCA
1861  AGGCGCCATGATGGTGGCTGAGAGCACTTGCCAACGGACTTTTTTCCCCCAGTTCTGCCC
1921  TCAATTATTGCTACCCATTTCTTGCGCAACTGTGTCTGCTTGCTCGGACCCTTCCAGCTC
1981  TGCTTTGTTGCGTTCACTGCGGCTTTTGTGCCACTTTTTCGCACCTGCTTTTTTGGCTGG
2041  AGTGGCTGTGCTGTGCTAGCGATCCTGGCGGTTTGTGATATGCTTATTGTTTGAGAAGGG
2101  CTGCTTTCCATTCATTGAGAGAGGCTGCCCTCGGACCGCAGGTATTTGATTTACTGGACT
2161  GTACCAGACTCAATCTTATTTTGTAACTATTAACATGCGCCAAAAAAAAAAAAAAAAAAA
2221  AAAAAAAAAAAA
```

Fig. 1 Nucleotide sequences of the gene encoding the DMBP-1 from *D. salina* and deduced amino acid sequences (GenBank no: DQ124215). The nucleotide sequences used as degenerate oligonucleotide primers are boxed. The putative initiation codon (ATG), the putative polyadenylation signal and the stop codon (TAA) in italic are underlined.

The ORF was obtained by ORF Finder on NCBI (http://www.ncbi.nlm.nih.gov/gorf/ gorf.html) and the predicted DMBP-1 protein precursor was 541 amino acids in length, with a 58.55 kDa of theoretical

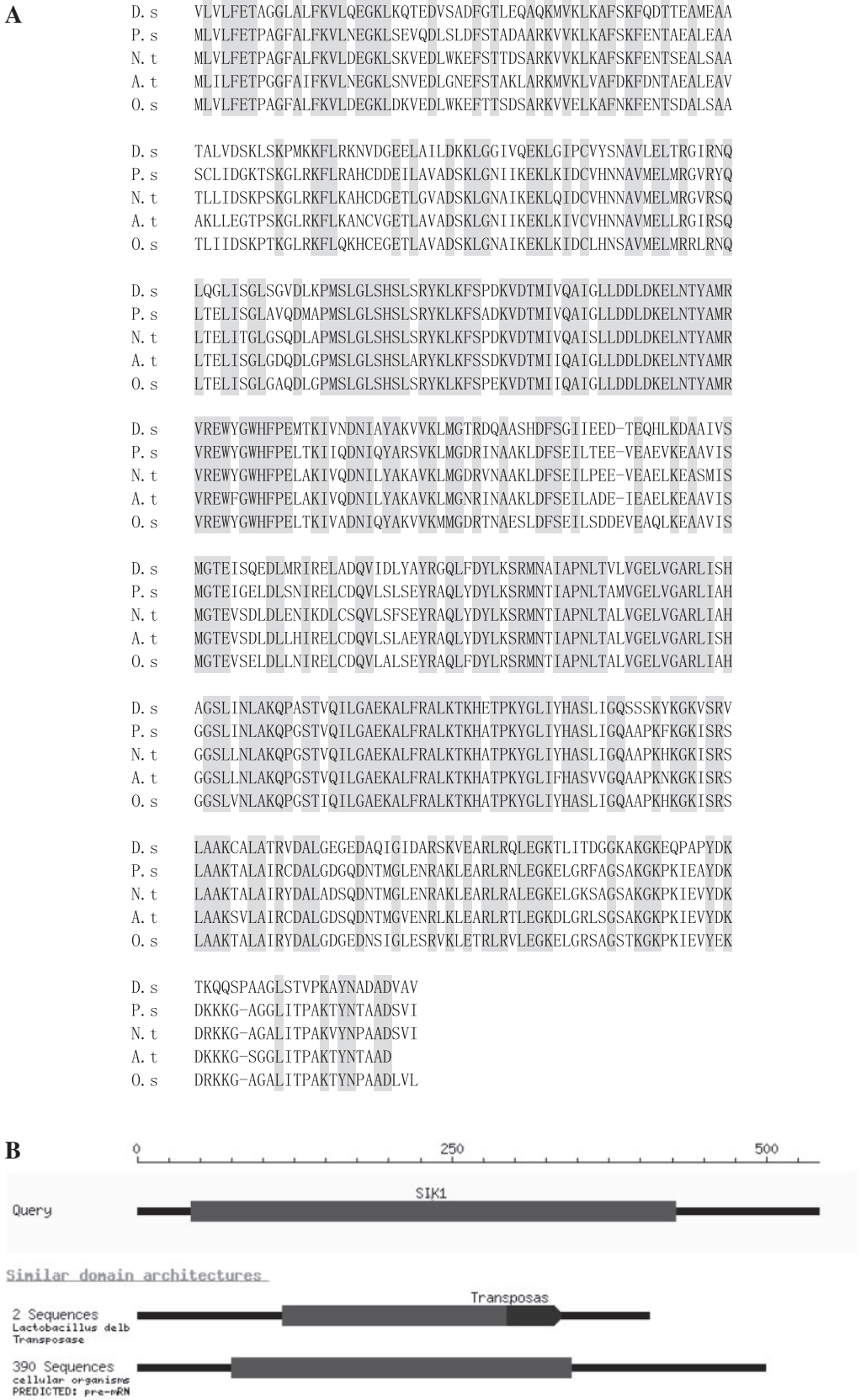

Fig. 2 Alignment and CCD search results. (A) The alignment of DMBP-1 protein using DNAMAN software. Multi-alignment was made by DNAMAN of the deduced DMBP-1 amino acid sequence with the four most homologous MAR/SAR binding proteins from *Pisum sativum* (GenBank accession no: AAC16330.1), *Nicotiana tabacum* (GenBank accession no: BAB41076.1), *Arabidopsis thaliana* (GenBank accession no: NP_187157.1), *Oryza sativa* (GenBank accession no: BAA31260.1). D.s: *Dunaliella salina*; P.s: *Pisum sativum*; N.t: *Nicotiana tabacum*; A.t: *Arabidopsis thaliana*; O.s: *Oryza sativa*. The completely identical amino acids are shaded. (B) The CCD search result of *D. salina* DMBP-1 protein via NCBI Blast. Analysis of the functional domain by CCD search showed that there was a SIK1 domain.

molecular weight and a 9.1 of pI value (GenBank accession no. DQ124215). Multi-alignment of the DMBP-1 with other organisms was conducted (Fig. 2A). According to NCBI protein–protein BLAST, the deduced DMBP-1 amino acid sequence had 75% (344/458) identities and 86% (395/458) positives in local alignments to nucleolar protein, a component of C/D snoRNPs from *Chlamydomonas reinhardtii,* and 66% (300/454) identities and 78% (355/454) positives to *Pisum sativum* SAR DNA-binding protein-1. It also shared very broad and high local identities and positives to MAR- binding proteins from species of broad classes such as *Oryza sativa* (61% identity and 76% positive), *Arabidopsis thaliana* (63% identity and 76% positive), *Nicotiana tabacum* (61% identity and 75% positive), *Lotus japonicus* (63% identity and 76% positive), *Medicago truncatula* (64% identity and 75% positive) and *Beta vulgaris* (65% identity and 78% positive).

We identified functional domains of the *D. salina* DMBP-1 protein via CCD search of NCBI, and found that there was a main putative conserved domain: SIK1 from position 40 to 400 (Fig. 2B). The SIK1 protein is implicated in ribosomal biogenesis, Nop56p homologue, which may be involved in translation, ribosomal structure and biogenesis.

The two-dimensional structure prediction of the *D. salina* DMBP-1 protein was conducted. Based on the Hierarchical Neural Network 23 result, the DMBP-1 mature peptide was composed of 56.19% α-helix, 7.95% extended strand and 35.86% random coil. α-Helices and β-sheet reside penetrating through all parts of the *D. salina* DMBP-1 protein.

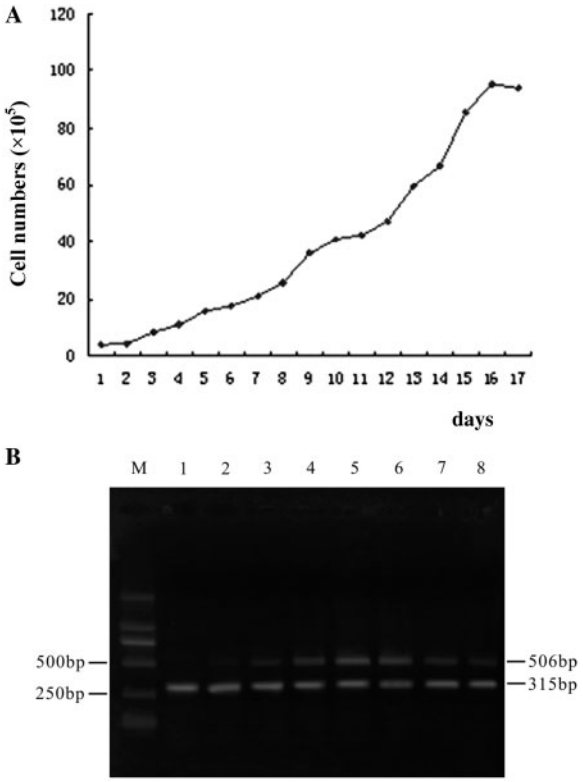

Fig. 3 Expression of the DMBP-1 gene in the different growth periods. (A) Cell density of *D. salina* cells grown in the different periods. (B) The relative proportion of the DMBP-1 mRNA was markedly increased from days 8 to 15 with the highest level of DMBP-1 mRNA on Day 15. Lanes 1–8 are representative of Days 1, 2, 4, 8, 15, 20, 24 and 30, respectively.

Expression of the gene encoding DMBP-1 protein

The semi-quantitative RT-PCR was used to analyse the expression profile of the DMBP-1 gene from *D. salina* cells grown under the 12-h light and dark/day. As shown in Fig. 3, the cell density ($\times 10^5$) of *D. salina* was elevated gradually from Day 3 after being cultured under the 12-h light and dark/day (Fig. 3A). The relative proportion of the DMBP-1 mRNA was significantly increased from Days 8 to 15 with the highest level on Day 15 (Fig. 3B), suggesting that expression of the DMBP-1 protein of *D. salina* is cell cycle dependent.

Southern–Western blots

In order to investigate the MAR DNA-binding properties of the DMBP-1 protein, the coding sequence was cloned into prokaryotic expression plasmid pET28a (+) and then expressed in *E. coli*. The expressed proteins in *E. coli* were recovered from inclusion bodies and analysed by SDS–PAGE. An ~75 kDa of recombinant protein was found, which agrees with the theoretical molecular weight of the recombinant DMBP-1-protein with the N-terminal tag.

The recombinant DMBP-1 protein from *E. coli* inclusion bodies was fractionated by 10% SDS–PAGE and blotted on nitro-cellulose. Proteins blots were probed with Dig-labelled MAR DNA and finally were exposed to X-ray film. The recombinant DMBP-1 protein bound to MAR fragments from human (GM), mouse (lgκ) and *D. salina* in the presence of up to 2,000-fold in excess of non-specific competitor DNA. As shown in Fig. 4, the *E. coli*-expressed DMBP-1 protein specifically interacted with DNA of MARs (Fig. 4A). The interaction of the recombinant DMBP-1 protein with MAR DNA probe was sensitive to 25 μM distamycin (Fig. 4B) and to poly[dA-dT]. poly[dA-dT] competitor DNA (Fig. 4C), but not to poly[dG-dC].poly[dG-dC] (Fig. 4D). These experiments demonstrate that the *E. coli*-expressed DMBP-1 protein specifically interacts with MAR DNA *in vitro*.

Expression and subcellular localization of the DMBP-1

In order to determine the subcellular localization of the DMBP-1, an expression vector with the DMBP-1 fused to green fluorescent protein was constructed, and then transfected into CHO cells via Lipofectamine 2000. Western blots showed that the DMBP-1 protein was expressed in CHO cells (Fig. 5A). Also, the green fluorescence was full of the cytoplasma of the cells transfected with the control plasmid pEGFP (Fig. 5 B) under fluorescence microscope, while the green fluorescence appeared only in the nuclei of CHO cells transfected with the pEGFP-MBP (Fig. 5C), indicating that the DMBP-1 is located in the nuclei.

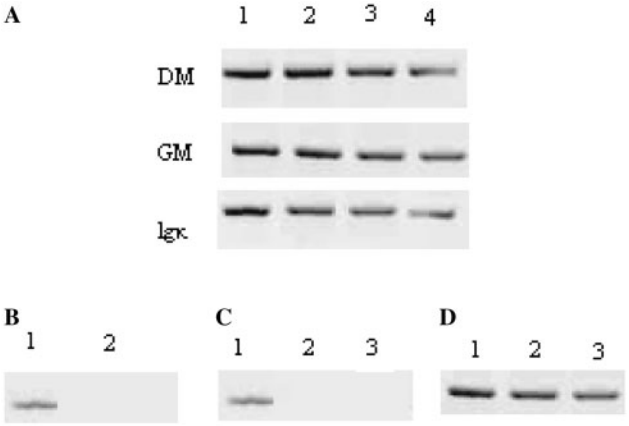

Fig. 4 DNA-binding properties of DMBP-1. The recombinant protein DMBP-1(0.1 μg) from *E. coli* was fractionated by 10% SDS–PAGE and blotted onto nitrocellulose. Protein blots were probed with ^{32}P-labelled MAR DNA and then autoradiographed. The probing conditions were varied as follows: (A) 0 (lane 1), 500-fold (lane 2), 1000-fold (lane 3) and 2000-fold (lane 4) in excess of non-specific competitor DNA (sheared *E. coli* genomic DNA) over probe DNA. The length of MARs used and their AT contents (in parentheses) were as follows: DM, 0.55 kb (62.5%); GM, 0.88 kb (66.0%); lgκ, 0.44 kb (72%). (B) A 1000-fold in excess of the unlabelled competitor (lane 1) and in the presence of 25 μM distamycin (lane 2). (C and D) 0 (lane 1), 500-fold (lane 2) and 1000-fold (lane 3) in excess of the unlabelled, poly[dA-dT].poly[dA-dT] and poly[dG-dC].poly[dG-dC].

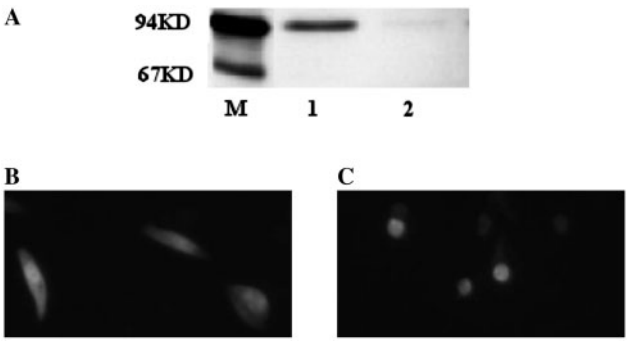

Fig. 5 Expression and subcellular localization of the DMBP-1 in CHO cells. An expression vector pEGFP-MBP, in which the DMBP-1 was fused to green fluorescent protein, was constructed and the vector pEGFP was used as a control. Subsequently, the vectors were transfected, respectively, into CHO cells via Lipofectamine 2000. (A) Blots were hybridized with mouse anti-hexahistidine monoclonal antibody. Lanes 1 and 2 represent, respectively, proteins from CHO cells transfected with the vectors pEGFP and pEGFP-MBP. (B) In CHO cells transfected with the control pEGFP, the green fluorescence was full of the entire cells under fluorescence microscope. (C) The green fluorescence appeared only in the nuclei of CHO cells transfected with the pEGFP-MBP under fluorescence microscope.

Discussion

The MAR-binding protein is a type of specific proteins that interact with and bind to MAR sequences. Some MAR-binding proteins have been isolated, so far, from plants and animals (*20–24*, *32*), but not from algae. In the present study, we isolated and identified a novel protein DMBP-1 from the unicellular microalga *D. salina*. DMBP-1 homologues have been identified in four plant species including both dicots and monocots. Whereas the N- and C-terminal parts of the proteins have little sequence similarity, the central part is well conserved.

The recombinant DMBP-1 was expressed in *E. coli* and could bind to MAR-DNA fragment in the presence of non-specific and G/C-rich competitor, but this interaction was competed by A/T-rich DNA and distamycin. The amino acid sequences of the DMBP-1 with pI of ~9 are like the Nop protein, which belongs to the MARBP family. The DMBP-1, as other Nop proteins, is more similar to the primary sequence of Nop 5 from yeast, which contains many KKX motifs in the C-terminal region.

A number of amino acid sequences interacting with MAR DNA have been characterized, including the SPRK or SPKK motif found in sea urchin sperm histone H1 and H2B, the repeated KRPRGRPKK motif present in HMG-1/Y, the KRKRGRPKK motif of the *Drosophila* D1 protein, the RKRGB motif consensus of homeodomain proteins and the KRPR motif in the phage 434 protein (*33*). However, no amino acid sequences resembling these motifs are present in DMBP-1.

The bipartite nuclear localization motif is characterized as two basic amino acid separated by a spacer of any 10 amino acids from a second cluster of basic amino acids, in which three out of five are basic and has been shown to be sufficient for nuclear localization of some proteins (*34*). In the present study, furthermore, motifs conforming to this definition were identified in the DMBP-1. KKX motifs repeats toward the C-terminus of DMBP-1 may contribute to the potential bipartite nuclear localization sequences. Domains containing repeats of these KKX motifs also occur in mammalian microtubule-associated proteins MAP1A and MAP1B, and in the yeast protein Cbf5p that is involved in mitotic chromosome segregation and rRNA biosynthesis (*35*).The KKX motifs have been demonstrated to have a microtubule-binding function *in vivo* in MAP1B and Cbf5p, and a microtubule-binding function *in vitro* in MAP1B (*36*). Hence, the KKX motif confers an ability to bind microtubules on the DMBP-1.

To investigate the location of the DMBP-1 in the cells, the expression vector with the GFP gene was transfected into CHO cells. The localization experiments indicated that the DMBP-1 was located in the nuclei, which is similar to Nop58 (*37*). Several experiments have demonstrated that some Nop proteins constitute the snoRNP, which is involved in the splicing of RNA, and Nop 5 protein can bind the rRNA precursor (*38*). The DMBP-1 shows a homology to yeast Nop5 family, in which Nop58p and 56p bind MAR DNA and are involved in ribosome biogenesis (*38*). Based on the findings of the present study, it is concluded that the DMBP-1 from *D. salina* is a novel MAR-binding protein, which may participate in RNA processing.

Supplementary Data

Supplementary Data are available at *JB* online.

Funding
National Natural Science Foundation of China (No. 30970055); International Science and Technology Cooperation Program of The Ministry of Science and Technology of China (No. 2007DFA01240); Henan Key Laboratory for Molecular Medicine.

Conflict of interest
None declared.

References

1. Mirkowitch, J., Mirauld, M.E., and Laemmli, U.K. (1984) Organization of the higher order chromatin loop:Specific DNA attachment sites on nuclear scaffold. *Cell* **39**, 223–232
2. Nickerson, J.A., Blencowe, B.J., and Penman, S. (1995) The architectural organization of nuclear metabolism. *Int. Rev. Cytol.* **162A**, 67–123
3. Laemmli, U.K., Kas, E., Poljak, L., and Adachi, Y. (1992) Scaffold-associated regions: *cis*-acting determinants of chromatin structural loops and functional domains. *Curr. Opin. Genet. Dev.* **2**, 275–285
4. Bode, J., Stengert-Iber, M., Kay, V., Schlake, T., and Dietz-Pfeilstetter, A. (1996) Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. *Crit. Rev. Eukaryot. Gene Expr.* **6**, 115–138
5. Bonifer, C., Vidal, M., Grosveld, F., and Sippel, A.E. (1990) Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J.* **9**, 2843–2848
6. Geyer, P.K. (1997) The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**, 242–248
7. Girod, P.A., Zahn-Zabal, M., and Mermod, N. (2005) Use of the chicken lysozyme 5′ matrix attachment region to generate high producer CHO cell lines. *Biotechnol. Bioeng.* **91**, 1–11
8. Goetze, S., Baer, A., Winkelmann, S., Nehlsen, K., Seibler, J., Maass, K., and Bode, J. (2005) Performance of genomic bordering elements at predefined genomic loci. *Mol. Cell Biol.* **25**, 2260–2272
9. Hughes, T.S., Langer, S.J., Johnson, K.W., Chavez, R.A., Watkins, L.R., Milligan, E.D., and Leinwand, L.A. (2009) Intrathecal injection of naked plasmid DNA provides long-term expression of secreted proteins. *Mol. Ther.* **17**, 88–94
10. Wang, T., Xue, L., Hou, W., Yang, B., Chai, Y., Ji, X., and Wang, Y. (2007) Increased expression of transgene in stably transformed cells of *Dunaliella salina* by matrix attachment regions. *Appl. Microbiol. Biotechnol.* **76**, 651–657
11. Wang, T.Y., Yang, R., Qin, C., Wang, L., and Yang, X.J. (2008) Enhanced expression of transgene in CHO cells using matrix attachment region. *Cell Biol. Int.* **32**, 1279–1283
12. Kim, M.S., Lim, H.S., Ahn, S.J., Jeong, Y.K., Kim, C.G., and Lee, H.H. (2007) Enhanced expression of EGFP gene in CHSE-214 cells by an ARS element from mud loach (Misgurnus mizolepis). *Plasmid* **58**, 228–239
13. Adachi, Y., Kas, E., and Laemmli, U.K. (1989) Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. *EMBO J.* **8**, 3997–4006
14. Ludérus, M.E., den Blaauwen, J.L., de Smit, O.J., Compton, D.A., and van Driel, R. (1994) Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. *Mol. Cell Biol.* **14**, 6297–6305
15. Fackelmayer, F.O., Dahm, K., Renz, A., Ramsperger, U., and Richter, A. (1994) Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA *in vivo* and *in vitro*. *Eur. J. Biochem.* **221**, 749–757
16. von Kries, J.P., Buhrmester, H., and Stratling, W.H. (1991) A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. *Cell* **64**, 123–135
17. Weitzel, J.M., Buhrmester, H., and Stratling, W.H. (1997) Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. *Mol. Cell Biol.* **17**, 5656–5666
18. de Belle, I., Cai, S., and Kohwi-Shigematsu, T. (1998) The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) *in vivo* in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J. Cell Biol* **141**, 335–348
19. Zhao, K., Käs, E., Gonzalez, E., and Laemmli, U.K. (1993) SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. *EMBO J.* **12**, 3237–3247
20. Meier, I., Phelan, T., Gruissem, W., Spiker, S., and Schneider, D. (1996) MFP1, a novel plant filament-like protein with affinity for matrix attachment region DNA. *Plant Cell* **8**, 2105–2115
21. Gindullis, F., Peffer, N.J., and Meier, I. (1999) MAF1, a novel plant protein interacting with matrix attachment region binding protein MFP1, is located at the nuclear envelope. *Plant Cell* **11**, 1755–1768
22. Samaniego, R., Jeong, S.Y., Meier, I., and de la Espina, S.M. (2006) Dual location of MAR-binding, filament-like protein 1 in Arabidopsis, tobacco, and tomato. *Planta* **223**, 1201–1206
23. Morisawa, G., Han-Yama, A., Moda, I., Tamai, A., Iwabuchi, M., and Meshi, T. (2000) AHM1, a novel type of nuclear matrix-localized, MAR binding protein with a single AT hook and a J domain-homologous region. *Plant Cell* **12**, 1903–1916
24. Hatton, D. and Gray, J.C. (1999) Two MAR DNA-binding proteins of the pea nuclear matrix identify a new class of DNA-binding proteins. *Plant J.* **18**, 417–429
25. Mínguez, A., Franca, S., and Moreno Díaz de la Espina, S. (1994) Dinoflagellates have a eukaryotic nuclear matrix with lamin-like proteins and topoisomerase II. *J. Cell Sci.* **107**, 2861–2873
26. Moreno Díaz de la Espina, S., Alverca, E., Cuadrado, A., and Franca, S. (2005) Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: the amazing dinoflagellates. *Eur. J. Cell Biol* **84**, 137–149
27. Ben-Amotz, A. and Avron, M. (1990) The biotechnology of cultivating the halotolerant alga *Dunaliella*. *Trends Biotech.* **8**, 121–126
28. Campbell, W.H. and Gowri, G. (1990) Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol.* **92**, 1–11
29. Wang, T.Y., Hou, W.H., Chai, Y.R., Ji, X., Wang, J.M., and Xue, L.X. (2005) Nuclear matrices and matrix attachment regions from Green alga: *Dunaliella salina*. *Acta Genet. Sin.* **32**, 1312–1318
30. Yu, J., Bock, J.H., Slightom, J.L., and Villeponteau, B. (1994) A 5′ beta-globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. *Gene* **139**, 139–145
31. Cockerill, P.N. and Garrard, W.T. (1986) Chromosomal loop anchorage of the kappa immunoglobulin gene

occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* **31**, 273–282
32. Barboro, P., D'Arrigo, C., Repaci, E., Bagnasco, L., Orecchia, P., Carnemolla, B., Patrone, E., and Balbi, C. (2009) Proteomic analysis of the nuclear matrix in the early stages of rat liver carcinogenesis: identification of differentially expressed and MAR-binding proteins. *Exp. Cell Res.* **1315**, 226–239
33. Reardon, B.J., Winters, R.S., Gordon, D., and Winter, E. (1993) A peptide motif that recognizes A.T tracts in DNA. *Proc. Natl Acad. Sci. USA* **90**, 11327–11331
34. Dehesh, K., Smith, L.G., Tepperman, J.M., and Quail, P.H. (1995) Twin autonomous bipartite nuclear localization signals direct nuclear import of GT-2. *Plant J.* **8**, 25–36
35. Cadwell, C., Yoon, H.J., Zebarjadian, Y., and Carbon, J. (1997) The yeast nucleolar protein Cbf5p is involved in rRNA biosynthesis and interacts genetically with the RNA polymerase I transcription factor RRN3. *Mol. Cell Biol.* **17**, 6175–6183
36. Jiang, W., Middleton, K., Yoon, H.J., Fouquet, C., and Carbon, J. (1993) An essential yeast protein, CBF5p, binds *in vitro* to centromeres and microtubules. *Mol. Cell Biol.* **13**, 4884–4893
37. Fujiwara, S., Matsuda, N., Sato, T., Sonobe, S., and Maeshima, M. (2002) Molecular properties of a matrix attachment region-binding protein located in the nucleoli of tobacco cells. *Plant Cell Physiol.* **43**, 1558–1567
38. Gautier, T., Bergès, T., Tollervey, D., and Hurt, E. (1997) Nucleolar KKE/D repeat proteins Nop56pand Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol. Cell Biol.* **17**, 7088–7098