Plant Molecular Biology **23**: 981–994, 1993.

The GAPDH gene system of the red alga *Chondrus crispus*: promotor structures, intron/exon organization, genomic complexity and differential expression of genes

Marie-Françoise Liaud[1], Christiane Valentin[1], Ulrike Brandt[1], François-Yves Bouget[2], Bernard Kloareg[2] and Rüdiger Cerff[1,*]

[1] *Institut für Genetik, Universität Braunschweig, Spielmannstr. 7, D-38106 Braunschweig, Germany (* author for correspondence);* [2] *Centre d'Etudes Océanologiques et de Biologie Marine, CNRS-UPR 4601, BP 74, F-29682 Roscoff Cedex, France*

Received 4 May 1993; accepted in revised form 17 August 1993

Key words: cis-acting elements, intron conservation, intron secondary structure, pre-mRNA splicing, CpG suppression, protoplasts, transcript levels

Abstract

Our previous phylogenetic analysis based on cDNA sequences of chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases (GAPDH; genes *GapA* and *GapC*, respectively) of the red alga *Chondrus crispus* suggested that rhodophytes and green plants are sister groups with respect to plastids and mitochondria and diverged at about the same time or somewhat later than animals and fungi. Here we characterize the genomic sequences of genes *GapC* and *GapA* of *C. crispus* with respect to promotor structures, intron/exon organization, genomic complexity, G + C content, CpG suppression and their transcript levels in gametophytes and protoplasts, respectively. To our knowledge this is the first report on nuclear protein genes of red algae. The *GapC* gene is G + C-rich, contains no introns and displays a number of classic sequence motifs within its promotor region, such as TATA, CAAT, GC boxes and several elements resembling the plant-specific G-box palindrome. The *GapA* gene has a moderate G + C content, a single CAAT box motif in its promotor region and a single intron of 115 bp near its 5′ end. This intron occupies a conserved position corresponding to that of intron 1 in the transit peptide region of chloroplast GAPDH genes (*GapA* and *GapB*) of higher plants. It has consensus sequences similar to those of yeast introns and folds into a conspicuous secondary structure of −61.3 kJ. CpG profiles of genes *GapC* and *GapA* and their flanking sequences show no significant CpG depletion suggesting that these genomic sequences are not methylated. Genomic Southern blots hybridized with generic and gene specific probes indicate that both genes are encoded by single loci composed of multiple polymorphic alleles. Northern hybridizations demonstrate that both genes are expressed in gametophytes but not in protoplasts where appreciable amounts of transcripts can only be detected for *GapC*.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X73036 (*GapC*; 2746 bp) and X73033 (*GapA*; 3091 bp).

Introduction

The nuclear genes encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase of glycolysis (GAPDH, gene *GapC*) has been widely used as phylogenetic marker to trace the evolutionary history of plants, animals and fungi [12, 37, 38, 40, 56]. Higher plants have a second, photosynthetic GAPDH [10] which is encoded in the nucleus by two related genes *GapA* and *GapB* [11] which were previously suggested to be of endosymbiotic origin [8, 9, 32, 36, 47, 53]. A phylogenetic analysis of GAPDH gene sequences from eubacteria [35] firmly established the close relationship of nuclear genes *GapA* and *GapB* to their homologue *gap2* in cyanobacteria, the free-living relatives of chloroplasts. Surprisingly, however, additional types of GAPDH genes were found in eubacteria, one of which (gene *gap1* in *Escherichia coli* and cyanobacteria) is very similar to nuclear *GapC* genes in plants, animals and fungi. In striking analogy to the cyanobacterial origin of photosynthetic *GapA*/*GapB* this strongly suggests that eukaryotic *GapC* is also of endosymbiotic origin and probably derives from purple bacteria, the presumptive ancestors of present-day mitochondria (for details see [35]).

In a recent study [33, 59] we characterized full-length cDNAs encoding chloroplast and cytosolic GAPDH of the red alga *Chondrus crispus* (genes *GapA* and *GapC*, respectively). A phylogenetic analysis of these genes in the context of their eukaryotic and eubacterial homologues suggests that rhodophytes and chlorophytes together form a monophyletic group with respect to plastids and mitochondria and that the two lineages separated at about the same time or somewhat later than animals and fungi. Here we characterize the genomic sequences of genes *GapA* and *GapC* from *C. crispus* with respect to primary structure, intron/exon organization, genomic complexity and their transcript levels in gametophytes and protoplasts, respectively. This is the first report on nuclear protein genes from red algae showing interesting novelties suggesting that rhodophyte genes may be different in some respects.

Materials and methods

Plant material

Gametophytes of *C. crispus* (Stackh.) were provided by Sanofi-Bio-Industrie (Carentan, France). The algae were kept in running sea water under controlled culture conditions (light intensity 60 μE m^{-2} s^{-1}; 12 h light/dark photoperiod; temperature ranging from 10 °C in winter to 17 °C in summer).

Isolation of DNA and RNA

Genomic DNA for cloning and Southern hybridizations was extracted from protoplasts of *C. crispus* apical tips. Protoplasts, prepared as previously described [31], were precipitated by centrifugation and the pellet was gently resuspended in a washing buffer containing 0.55 M NaCl, 5 mM KCl, 5 mM $CaCl_2$, 0.35 M sorbitol, 40 mM EDTA in 50 mM Tris-HCl pH 7.0. The cells were lysed with the lysis buffer composed of 50 mM EDTA and 4% Sarcosyl in 50 mM Tris-Cl pH 8.0. Genomic DNA was purified by a phenol extraction, a precipitation with isopropanol and a centrifugation in a CsCl gradient. Total RNA was extracted from intact cells and from protoplasts according to the method of Apt and Grossman [3]. The poly(A)$^+$-containing mRNA fraction was purified from total RNA by adsorption to oligo(dT) cellulose.

Construction of the genomic library and isolation of clones.

Chondrus crispus DNA (150 μg) was partially digested with *Mbo* I and size-fractionated on a 10 to 40% sucrose gradient. The 15–25 kb fractions were purified by dialysis against TE buffer [34]. *Mbo* I partials (1 μg) were ligated for 16 h at 12 °C to 0.7 μg of λEMBL3 vector digested with *Sal* I (Stratagene). The ligation mixture was heated at 55 °C for 5 min and packaged *in vitro* according to Hohn [25]. *Escherichia coli* strain K803 was employed as a host.

The genomic library was screened with homologous cDNA clones identified as described [33] and labelled by random priming [15]. The hybridization was performed at 60 °C overnight in 6× SSPE, 0.2% PVP, 0.2% Ficoll, 0.1% SDS, 0.5 μg/ml denatured salmon sperm DNA and the labelled probe. Filters were washed twice for 20 min at 60 °C, once in 2× SSPE, 0.1% SDS and once in 0.2× SSPE, 0.1% SDS. Positive plaques were rescreened and purified to single plaques by standard procedures. Genomic fragments digested with *Sal* I or *Hin*d III were subcloned into BlueScript.

DNA sequence analysis

An ordered set of deletion clones was prepared for each subcloned *Sal* I and *Hin*d III fragment by using exonuclease III following the Stratagene protocol. The deletion clones were sequenced by the dideoxy chain termination method as double-stranded DNA by using the T7 polymerase (Pharmacia protocol). When required, oligonucleotides (17-mers) were synthesized according to sequence information and used directly as primers for further sequencing.

PCR amplification of the intron sequence of Chondrus crispus GapA

PCR amplification was performed in the DNA Thermal Cycler (Perkin-Elmer Cetus). The reaction mixture contained 10 pg of template DNA (the *GapA* gene inserted into the BlueScript vector), 300 ng of each primer (GTACGTTTCCGCCTT and CTGCGAGATTAGATC), 200 mM of each dNTP, 0.5 units of *Taq* DNA polymerase (Boehringer) and *Taq* DNA polymerase reaction buffer (67 mM Tris pH 8.8, 2 mM $MgCl_2$). Denaturation of DNA was performed at 93 °C for 1 min. Primer annealing and primer extension reactions were carried out at 56 °C for 1 min and at 72 °C for 1 min, respectively. After 30 cycles of amplification, the PCR products were blunt-ended with mung bean nuclease and cloned into the BlueScript vector.

Southern hybridizations

C. crispus DNA (10 μg) was digested to completion with 15 U of the respective restriction enzyme, electrophoresed on a 0.7% agarose gel, capillary-transferred and UV-coupled to a Hybond N (Amersham) nylon membrane. Filters were hybridized for 24 h at 65 °C (probe *GapA*) or 75 °C (probe *GapC*) in 30 ml of 6× SSPE, 0.1% SDS, 0.2% PVP and 0.2% Ficoll containing 100 ng of random-prime-labelled DNA probes. Filters were washed twice for 10 min at 65 °C (probe *GapA*) or 75 °C (probe *GapC*), once in 2× SSPE, 0.1% SDS and once in 0.2× SSPE, 0.1% SDS. Autoradiograms were exposed for 20 h at −80 °C.

Northern hybridizations

For RNA gel blot analysis, 3 μg of poly(A)$^+$ from *C. crispus* gametophytes and *C. crispus* protoplasts was denatured in formamide/formaldehyde, fractionated on formaldehyde-agarose gel [51], transferred and UV-coupled to a Hybond N nylon membrane. The hybridization with cDNA probes encoding *GapC* and *GapA* was performed for 20 h at 62 °C in the same buffer as specified above for Southern hybridizations. Filters were washed twice for 15 min at 62 °C in 2× SSPE, 0.1% SDS. Autoradiograms were exposed for 12 h at −80 °C.

Results

Isolation and characterization of genomic clones encoding GapC *and* GapA *from* Chondrus crispus

The genomic library was screened with the homologous cDNAs encoding GAPC and GAPA from *C. crispus* identified as described previously [33]. From the positive genomic clones one for each gene was purified and submitted to sequence analysis. The structural features of the two genes *GapC* and *GapA* from *C. crispus* (PGCDB codes *GapC*Cc.1* and *GapA*Cc.1*) are summarized in

Fig. 1. The *GapC* gene contains no intron and differs in seven nucleotide positions from the corresponding cDNA characterized: two differences were found within the 5′ leader (one A→C transversion and one insertion of a single C), four in the coding region (three synonymous substitutions and one non-synonymous mutation leading to a conservative exchange in position 259, Ala→Thr; see alignment in Fig. 3B) and one change in the 5′ trailor region (insertion of a single G).

The *GapA* gene is interrupted by a single intron of 115 bp at the 5′ end of its transit peptide region (see Fig. 3). As shown in Fig. 2A, the 5′ and 3′ splice junctions and the branch point of the *GapA* intron are compatible with the corresponding consensus sequences of higher plants and vertebrates but resemble also the more conserved sequences of *Schizosaccharomyces pombe* and yeast. The intron contains a number of inverted repeats which can stabilize a secondary structure of −61.3 kJ (Fig. 2B). If compared to its cDNA [33, 59] the *GapA* gene differs in two nucleotide positions within the transit peptide region: one synonymous substitution and one replacement mutation leading to a conservative exchange at position −52, Thr→Ser (see Fig. 3A).

As depicted in Fig. 1, the promotor region of the *GapC* gene contains a number of classic sequence motifs: two copies each of potential TATA and CAAT boxes and multiple GC box motifs [14]. In addition, upstream the CAAT box

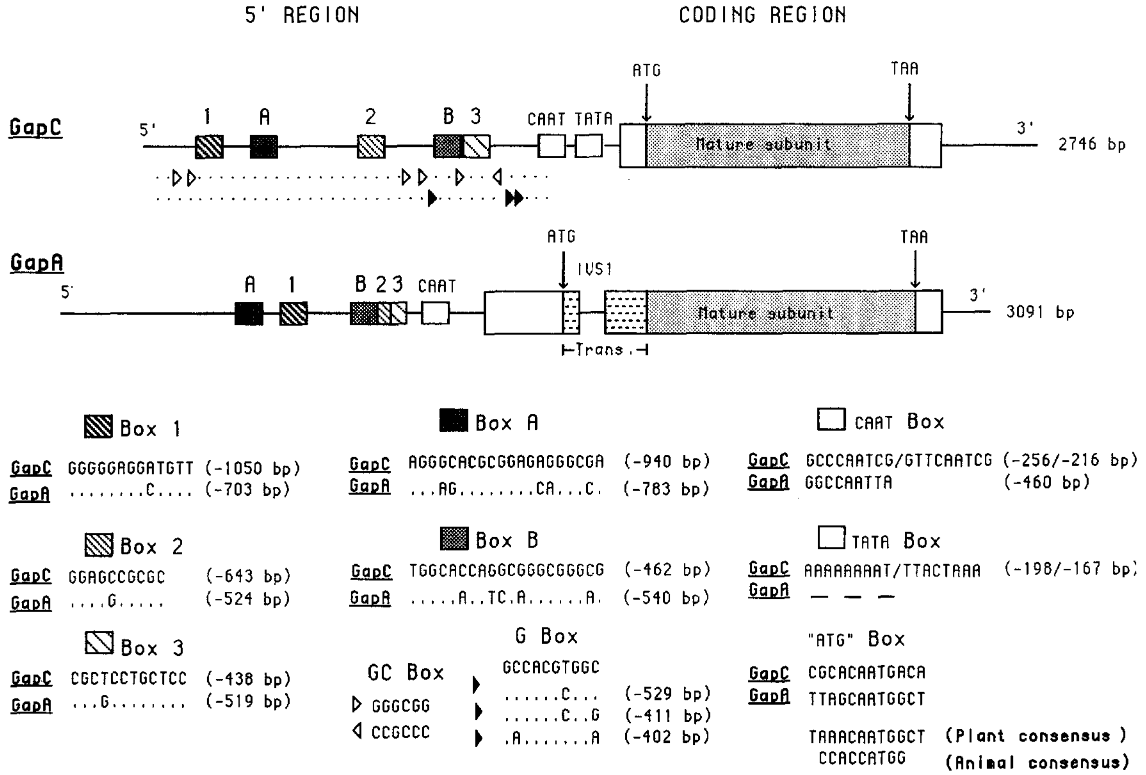
5' REGION
CODING REGION
GapC
5'
1
A
2
B 3
CAAT TATA
ATG
TAA
Mature subunit
3'
2746 bp
GapA
5
A
1
B 2 3
CAAT
ATG
IVS1
TAA
Mature subunit
3'
3091 bp
Trans.
Box 1
GapC GGGGGAGGATGTT (-1050 bp)
GapA ....C.... (-703 bp)
Box 2
GapC GGAGCCGCGC (-643 bp)
GapA ....G..... (-524 bp)
Box 3
GapC CGCTCCTGCTCC (-438 bp)
GapA ...G........ (-519 bp)
Box A
GapC AGGGCACGCGGAGAGGGCGA (-940 bp)
GapA ...AG........CA...C. (-783 bp)
Box B
GapC TGGCACCAGGCGGGCGGGCG (-462 bp)
GapA .....A..TC.A......A. (-540 bp)
GC Box
▷ GGGCGG
◁ CCGCCC
G Box
GCCACGTGGC
▶ ......C... (-529 bp)
▶ ......C..G (-411 bp)
▶ .A.......A (-402 bp)
CAAT Box
GapC GCCCAATCG/GTTCAATCG (-256/-216 bp)
GapA GGCCAATTA (-460 bp)
TATA Box
GapC AAAAAAAAT/TTACTAAA (-198/-167 bp)
GapA — — —
"ATG" Box
GapC CGCACAATGACA
GapA TTAGCAATGGCT
TAAACAATGGCT (Plant consensus)
CCACCATGG (Animal consensus)

Fig. 1. Structural features of genes *GapC* and *GapA* from *Chondrus crispus*. Large boxes symbolize exons; small boxes and open and closed triangles designate potential *cis*-acting elements in the promotor and 5′-upstream regions. Shaded, hatched and open large boxes symbolize exon regions encoding the mature subunit, the transit peptide and the 5′/3′ non-coding regions (present in the cDNAs) respectively. The position of the intron in the transit peptide region of the *GapA* gene is indicated (IVS1). Elements called Box 1, Box 2, Box 3, Box A and Box B are '*Chondrus*-specific' motifs conserved between genes *GapC* and *GapA*. All other elements are motifs related to known eukaryotic *cis*-acting elements (TATA, CAAT, GC Box and G Box). Nucleotide sequences of elements are specified below the genes and positions of elements are given as number of bases upstream of the AUG initiation codon.

A) Intron consensus sequences

	5' Donor	Branch point	3' Acceptor
GapA C. crispus	CAG:GTACGT ··· 64 ···	TGCTAACG ·· 33 ···	CGCAG:G
Yeast Consensus	G:GTATGT ········· * *	TACT**A**ACA ·········· *	YAG:
S. pombe Consensus	:GTAAGT .... T	CTA**A**C .... T	ATAG: C
Plant Consensus	CAG:GTAAGT .... A	TTCTR**A**Y .... RRT	TGCAG:G
Vertebrate Consensus	CAG:GTAAGT .... A	CTG**A**C .... Y R Y	YNCAG:G n

B) GapA intron structure

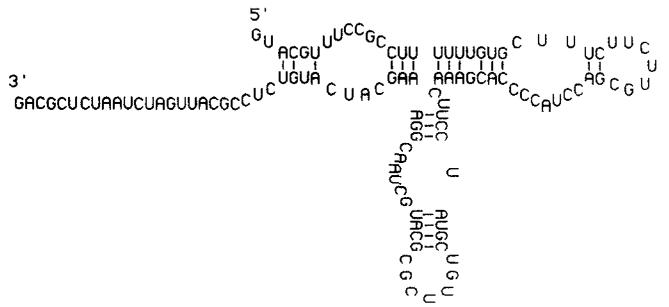

Fig. 2. The single intron of *Chondrus crispus GapA* contains splicing junctions similar to those of yeast and folds into a conspicuous secondary structure of about −61.3 kJ. Sources of consensus sequences: yeast [29, 48]; plants and vertebrates [29, 55]; *Schizosaccharomyces pombe* [45]. The essential branchpoint adenosine is indicated in bold. The yeast consensus sequences are 100% conserved except for the positions indicated by asterisks.

sequences there are three imperfect direct repeats resembling the palindromic G-box motif GCCAGCTGGC of angiosperms [27, 30, 41]. In the *GapA* promotor a single CAAT box can be found but no TATA box nor any other classic motifs of eukaryotic genes. The 'ATG boxes' of the two genes seem related somewhat more closely to the consensus of higher plants than to that of animals [26]. It should be noted that the potential TATA and CCAAT boxes of the *GapC* gene are surrounded by G + C-rich sequences (see Fig. 4B and below) which makes it seem unlikely that these motifs originated by chance and do not have regulatory functions.

While the two promotors differ considerably with respect to classic motifs, they share a number of '*Chondrus*-specific' elements: two 20-mers with 5 mismatches each (called Box A and Box B) and three elements of 10 to 13 bp with one mismatch each (called Box 1, Box 2 and Box 3 in Fig. 1). Each set of boxes is arranged in the same order (A→B; 1→2→3) but spaced differently in the two genes. Box B, Box 2 and Box 3 in the *GapC* gene overlap by 1 and 3 bp respectively (compare sequence elements in Fig. 1).

In Fig. 3A the derived transit peptide of *C. crispus GapA* (line 6) has been aligned with *GapA* and *GapB* transit peptides of higher plants and positions of introns 1 and 2 are indicated by arrow heads. The *Chondrus* transit peptide shows up to 15% amino acid sequence similarity with *GapA*/*GapB* transit peptides [33] and the corres-

A) Intron positions in GapA/GapB transit peptide regions.

```
  1        10        20        30        40        50        60        70        80        90
1 MATHAALASTRIPTNTRFPSKTS-HSFPSQCASKRLEVGEFSGLK-STS--CISYVH--SARDSSFYDVVAAQLTSKANG-STAV--KG-VTVA  GapB  Pea          -51/-13
2 M.......VS...VTQ.LQ..SAI....A..S.....A.....R--M.--S.-------GGEA..F.A....IIP..VTT..P.--R.-E..A  GapB  A thaliana   -46/-14
3 M.SATFSVAK--.AIK----------------ANGKGFS.....RN.SRHLP-------FS.K..DDFHSLVTFQTN.V.-.SGGHK.SL.VEA  GapA  Pea          -54/-15
4 M.SSM-.SA.TV.LQQ------------------GGGLS.....RS.A.-LPMRRNATSDD----.MSA.SFR-.H-.V.-TSGGPRRAP-.EA  GapA  Maize        -51/-14
5 M.SVTFSVPK------------------------GFT.....RS.SASLP------FGKKL..DEF.SIVSFQTS.M.-.SGGYR..-..EA  GapA  A thaliana   -50/-14
6 M.FV.PV.TV.AT.KSSVCQVQ-----------G.SSFAQ...M.KVNQSSRLQP----AQSG.A.GGYSD.NDAFYT--RVSGIVAATFGPTM  GapA  C crispus    -55/---
                                 IVS 1                                  IVS 2
                               -46 to -55                             -13 to -15
```

B) Amino acid alignement of mature subunits GAPC and GAPA.

```
    1        10        20        30        40        50        60        70        80        90       100       110       120       130
1   MTAPKVGINGFGRIGRLVLRAAIEKGT--CQVVAINDPFIDLDYMAYMLKYDSTHGRYA-GDVSIKDG-KLQVDG-NSITVFAHRDPAEIPWATAAADYIVEATGVFTLKDKAAAHFKGGAKKVVISAPS
2  MGAKI.I............A.V.LKRDD--VEL..V.....TT...T..F....V..QWKNDELTV..S-NTLLF.QKPV......N.E.....STG..I...S.....D.......L.......I.....
3     MV...V..........T...VLS.K--V..............N..V..F.......HFK-.T.KAEN.-..VIN.-HA..I.QE...SN.K..D.G.E.V..S.....TME..G..L.....R.I.....
4     MVR.A...........M.I.LSRPN--VE...L.....TN..A...F..........-.E..HD.K-HII...-KK.ATYQE....NL..GSSNV.IAIDS....KEL.T.QK.IDA.......T...
5    MTI............I.F...QKRSD--IEI.....LL-.A.................FD-.T.EV...-H.I.N.-KK.R.T.E....NLK.DEVGV.VVA....L.LTDET.RK.ITA......MTG..
6   MAKL.................G.NNPN--IEF.G...LV-PP.NL..L.........KLR-SQ.EA..D-GIVI..-HF.PCVSV.N...L..GKLG...V..S..L..DSEG.SK.LQA...R.I....T
7             ++++++++   +              ++              +++  +                +          + +     +               ++ +      +   +   +++ +     +
8    >KVRVAINGFGRIGRNFIRCWAGRSDSNMEVVCINDTS-GVKTASHLLKYDSILGTFD-ADVSAGED-TISVNG-KTIKIVSNRNPLQLPWKEMNIDIVVEATGVFVDAPGAGKHIEAGAKKVLITAPG
9   >KQLK...............L...H..K..PLD.IA....G-...Q..........T..I..-...KPVGTDG...D.-.V..V..D...AN.....LG..L.I.G......RE...R..T...........
10   >.LK...............L...H..K..PL..IVV..SG-...N..........M....K-.E.KILNNE..T.D.-.P..V..S.D..K...A.LG....I.G......G.......Q.......I....A
11   MVI...............A...L..EN..I.L.AV....-DPR.NA...N...S..VKK-V.IT.DDN-S.T...-....C..D...EN.....WE..LII.S....TSKE..L..VN...........

          140       150       160       170       180       190       200       210       220       230       240       250       260
1   KD---APMFVCGVNEAKYTPDLD-IISNASCTTNCLAPLVKVIHEKYGIEEGLMTTVHATTATQKTVDGPSNKDWRGGRGRGRNIIPSSTGAAKAVGKVMPELNGKLTGMAFRVPTPDVSVVDLTVRLTS
2   ..---.....V....NE.K.EF.-..................A...NDRF..V.........SI...........S........AASF.................L.A........S.....V...........EK
3   A.---.....M...HE..DKS.K-.V................A....DNF..V..........I...........G.L..DD..AAQ....A...............I.................N........C..EK
4   ST---.....M....V...S..K-.V................A...NDAF.............SL...........H........TASG.................L...Q..............V.........K.DK
5   ..--NT....K.A.FD..AGQD--.V................A...NDNF..I.......................H.........ASQ.................L.................N...........EK
6   ..PDRVRTLLV...HDLFD.SK.V.V...........IA...NDNF.LT.........M...P.....K........AAQ............AL.L...K...........I.........FKT.K
7               + +          + +++++++++ ++  ++       +    + +++ +   +   +   +          +  +     ++ +  +++++ +  + + + +++ +    +++   ++  +
8   KGD-GIGTFVVGVNEKDYSHDKYDIVSNASCTTNCMAPFMKVLDDEFGVVRGMMTTTHSYTGDQRLLDA-GHRDLRRARSAALNIVPTTTGAAKAVALVVPSLKGKLNGIALRVPTPNVSVCDVVMQVNK
9   ...--.P.Y.....ADA.T.AD-..I.........L...V....QK..IIK.T.................-S........A........S...........L.T...................V.L.V..S.
10  ..A-D.P.Y.I....Q..G.EVA..I.........L...A....E...I.K.T.................-S........A........S.......S..L.Q...................V.L.VN.A.
11  .N--ED....I...HH..D.NVHH.I.........L..IA...N.K..IIK.S.................-S........A..I.....S....R....I.E.........V..........MV.F.V..E.

          270       280       290       300       310       320       330       340
1   E-TTYEDIKATMKAAAD--DSMKGIMKYTEDAVVSTDFIHDDASCIFDASAGIMLNSKFCKLVAWYDNEWGYSNRVVDLIAHISKVQ*
2   A-A..DE...AI.EESE--GKL...LG....D.......G.TR.S....K...A..D..V...S.....L...T......V..A-K.L*
3   P-AK.D...RVV.....--GPL...LG....Q...C..NG.SH.ST...G...A..DH.V...S.....F.........MV.MASKE*
4   .-...DE..KVV....E--GKL..VLG........S..LG.SH.S........Q.SP..V...S.....Y...T.....VE..A.A*
5   A-A...Q...AV....E--GE...VLG....D......NGEVCTSV...K...A..DN.V...S.....T....K.L.......K*
6   A-.S.KE.C.A..Q.SE--G.LA..LG..DEE......QG.TH.S....G...E...N.F.V...................MLSMIQKEQLAAV*
7                          +          ++ +            +            +   +++++ +++   + ++
8   K-TFKEEVNGALLKASE--GAMKGIIKYSDEPLVSCDHRGTDESTIIDSSLTMVMGDDMIKVVAWYDNEWGYSQRVVDLGEVMARQWK*
9   .-..A....E.FRESAA--KELT..LSVC......V.F.C..V.STV............LV..I................ADIV.NNWK*
10  .GISA.D..A.FR..A.--.PL...LDVC.V....V.F.CS.V..T.............V.........................AHLV.NK.P<
11  R-.IT....Q..KD...--.PL...LD..ELQ...S.YQ...A.S.V.A...L...N.LV..M...............L..A.LV.EK.V*
```

			Homology (%)	
1	*Chondrus crispus*	GAPC	100	
2	Pea	GAPC	68	
3	Chicken		69	
4	Yeast		67	
5	*E.coli*	GAP1	67	
6	*Anabeana variabilis*	GAP1	60	
7	Consensus GAPC/GAPA		28	
8	*Chondrus crispus*	GAPA	43	100
9	Pea	GAPA	46	71
10	Pea	GAPB	46	71
11	*Anabaena variabilis*	GAP2	45	67

Fig. 3. A. Alignment of *GapA* and *GapB* transit peptides from higher plants and *Chondrus crispus* and positions of introns 1 and 2 (IVS 1 and IVS 2). All transit sequences are compared to that of pea *GapB*. The intron positions (−46 to −54 for IVS 1 and −13 to −15 for IVS 2) are given relative to the higher-plant cleavage site Ala/Lys, with alanine counted as residue −1. Identical amino acids are marked by dots and indels by dashes. The proteolytic cleavage site of the *C. crispus* transit peptide is probably Met/Lys rather than Ala/Lys as in higher plants [33]. Sources of sequences: pea, sequences 1 and 3 [8]; maize, sequence 4 [47]; *Arabidopsis thaliana*, sequences 2 and 5 [54]; *Chondrus crispus*, this paper. B. Amino acid sequence alignment of 11 GAPDH sequences comprising 4 eukaryotic GAPC, 3 eukaryotic GAPA/GAPB and 3 eubacterial GAPDH polypeptides from 6 different species as specified at the bottom of the figure. The sequences are presented in two blocks, a GAPC block (lines 1 to 6) and a GAPA block (lines 8 to 11). The sequences in each block are compared to *C. crispus* GAPC and GAPA, respectively, the only sequences written in full. Only amino acids not identical to these reference sequences are shown in either group. The first and the last residues of each sequence are shown irrespective of similarity. Amino-terminal (transit peptides of GAPA and GAPB) and carboxy-terminal (GAPB) extensions are not shown and, where present, are indicated by arrowheads. Line 7: + signs indicate residues conserved across all GAPC and GAPA sequences. Sources of sequences: *C. crispus*, sequences 1 and 8, this paper; pea, sequences 2, 9, 10 [8, 38]; chicken and yeast, sequences 3 and 4 [40]; *E. coli*, sequence 5 [7]; *Anabaena variabilis*, sequences 6 and 11 [35].

ponding gene region is interrupted by a single intron at a position very similar to that of intron 1 in genes *GapA* and *GapB* of higher plants. In Fig. 3B the derived amino acid sequences of the mature GAPC and GAPA polypeptides of *C. crispus* are aligned in two separate blocks with the GAPC sequences of pea, chicken, yeast, *E. coli* (gene *gap1*), *Anabaena variabilis* (gene *gap1*) and the GAPA/GAPB sequences of pea and *A. variabilis* (gene *gap2*), respectively. The table at the bottom of Fig. 3B shows that GAPC and GAPA of *C. crispus* are almost as similar to their homologues in *E. coli* (gene *gap1*, 67%) and A. variabilis (gene *gap2*, 67%), respectively, than to

A) G+C Content

		Total sequence	5' Region	Leader	Coding region	3' Region	Intron
GapC	bp	2746 bp	-1244 to -101	-100 to -1	1 to 1008	1009 to 1503	———
	% G+C	**62 %**	**66 %**	**71 % (60 % C)**	**63 % [50/89]**	**50 %**	
	obs./exp. CG	1.12	1.06		1.21		———
	GC	0.96	0.97		0.96		
	CNG	0.84	0.90		0.79		
	GNC	1.02	0.75		1.32		
GapA	bp	3091 bp	-1443 to -308	-307 to -1	1→66 & 182→1360	1361 to 1649	67 to 181
	% G+C	**53 %**	**49 %**	**57 %**	**57 % [49/73]**	**48 %**	**49 %**
	obs./exp. CG	1.00	0.91		1.04		———
	GC	1.02	1.16		0.88		
	CNG	0.90	0.92		0.88		
	GNC	1.07	0.99		1.16		

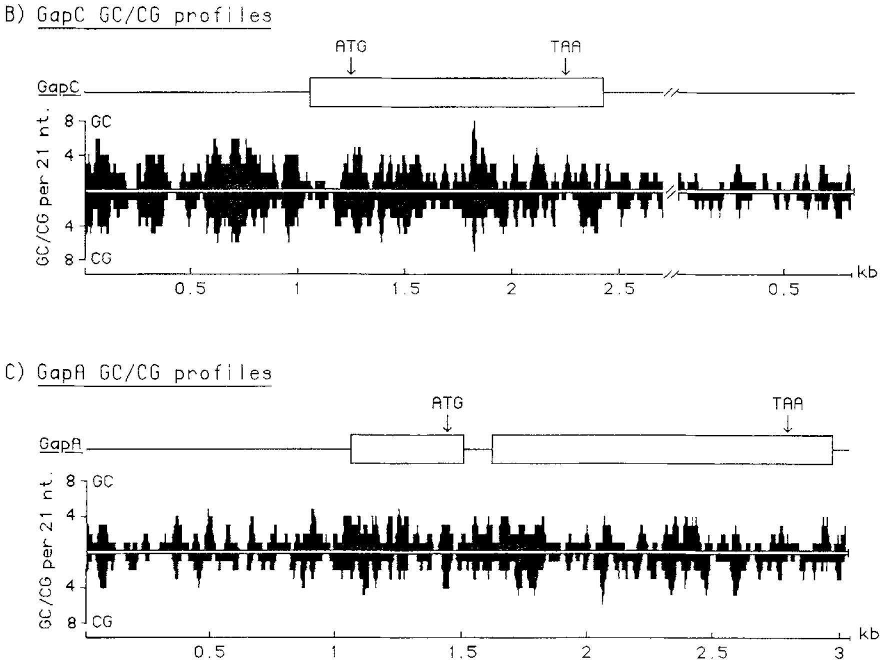

Fig. 4. Genes *GapC* and *GapA* and their genomic surroundings differ in G + C content and show no CpG depletion. A. G + C distribution (bold numbers) and calculated observed/expected values for CpG, GpC, CpNpG and GpNpC. Bold numbers in brackets (50/89 and 49/73) show the G + C content within codons as two separate values for first and second positions/third position, respectively. B and C. GpC and CpG profiles for genes *GapC* and *GapA*, respectively. Each vertical line represents the number of doublets per 21 bases.

their homologues in eukaryotes (67% to 69% for GAPC, 71% for GAPA/GAPB) [33].

Figure 4A demonstrates that the two genes differ significantly in their G + C content. The *GapC* gene is richer in G + C than the *GapA* gene (62% versus 53%), in particular in the promoter (66% versus 49%) and leader (71% versus 57%) regions. The 100 bp leader present in the *GapC* cDNA contains 60% C and three tandem repeats of the sequence ACCCCGAT/CCG starting 66 bp upstream of the ATG codon. The trailor regions and the *GapA* intron are relatively G + Cpoor (48% and 49% respectively). The CpG and GpC profiles of *GapC* and *GapA* (Fig. 4B and C) and the observed/expected values for CpG, GpC, CpNpG and GpNpC (Fig. 4A) show no significant deviations except for the coding and 3′-flanking region of gene GapC. In this region of about 1.5 kb CpNpGs are somewhat suppressed (observed/expected = 0.79) while GpNpCs (observed/expected = 1.32) and CpGs (observed/expected = 1.21) are overrepresented.

GapC *and* GapA *from* Chondrus crispus *are encoded by multiple genes (alleles) of very similar structure*

To enumerate the genes encoding GAPC and GAPA in *C. crispus* we hybridized Southern blots of genomic DNA digested with *Eco* RI, *Hin*d III, *Bam* HI and *Kpn* I (lanes E, H, B and K in Fig. 5A and B) with gene specific (3′ trailor- and intron-specific) and generic (coding) DNA probes. All

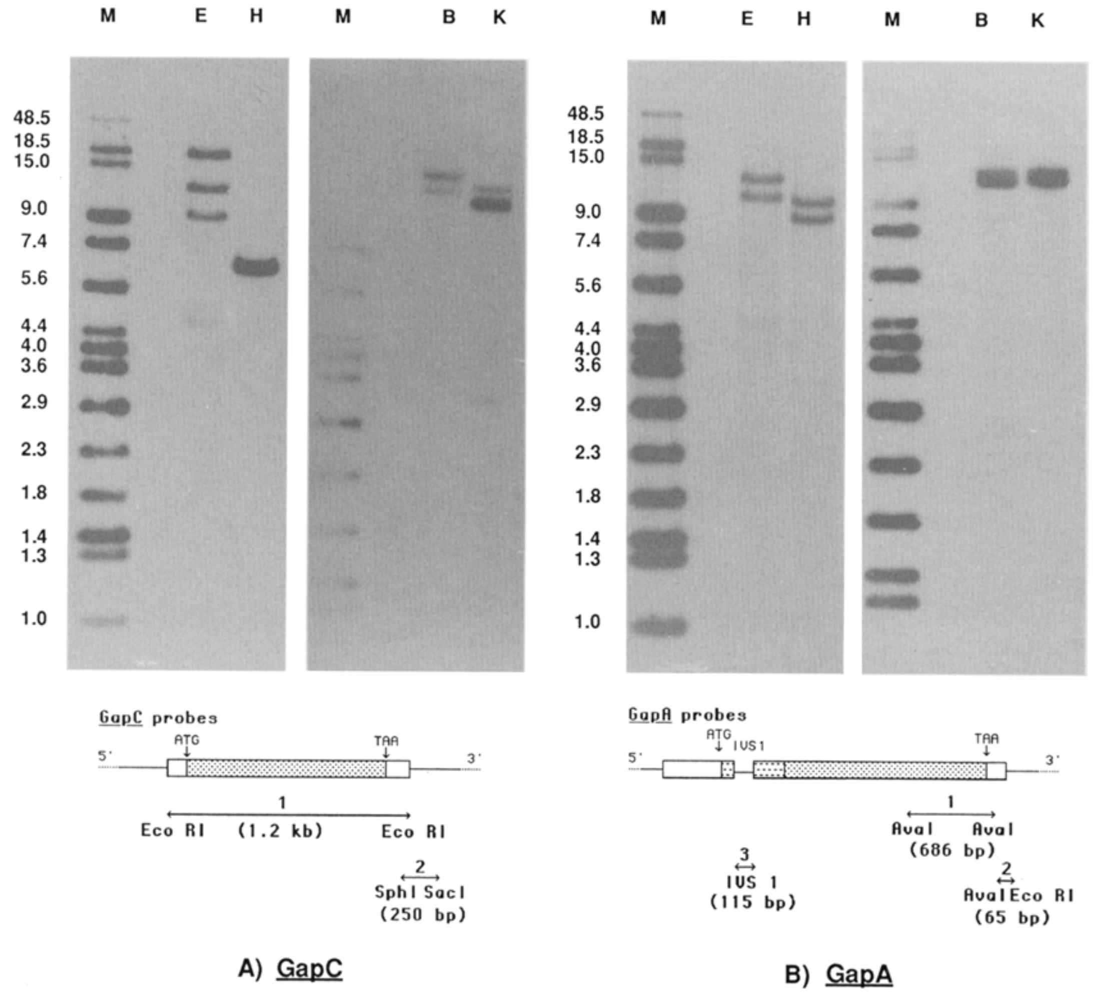

Fig. 5. Counting of genes *GapC* and *GapA* by genomic Southern blotting. Each lane contains 10 μg of genomic DNA digested with *Eco* RI, *Hin*d III, *Bam* HI and *Kpn* I (lanes E, H, B and K in panels A and B). The digests were electrophoresed on a 0.7% agarose gel and the separated fragments were blotted onto a nylon membrane. For each gene the same hybridization patterns were obtained if filters were hybridized with cDNA-coding fragments or with specific probes encoding 3′ trailor and intronic sequences. Schematic drawings of the probes used in each case are shown at the bottom of panels A and B. Note that the *Eco* RI sites at the borders of probes *GapC* and *GapA* are derived from the linkers used for cDNA cloning [33].

four restriction enzymes do not have recognition sites within the sequenced gene regions. For *GapC* (Fig. 5, panel A), one (*Hin*d III), two (*Bam* HI, *Kpn* I) and three (*Eco* RI) bands, respectively, are discovered with either the generic or gene-specific probe suggesting that there are three very similar *GapC* genes or alleles in *C. crispus*. The three *GapC* bands of the *Eco* RI pattern show equal intensities as one would expect for three separate equimolar *GapC* genes. The single strongly hybridizing *Hin*d III band probably represents three different but equally sized genomic fragments, while the intensity differences between the separate bands of the *Bam* HI and *Kpn* I doublets (panel A, lanes B and K) can be interpreted in terms of three genomic fragments, two of which are identical in size. Two faint bands of about 5 and 3 kb can be seen in lanes E and K, respectively, hybridizing weakly with the *GapC* probe. Whether or not these bands represent an additional divergent *GapC* (pseudo)gene cannot be decided yet. For *GapA*, doublets with equal band intensities are found for all four restriction enzymes with generic and gene-specific probes indicating the presence of two separate but very similar *GapA* genes or alleles.

GapC *and* GapA *are differentially expressed in* Chondrus crispus *protoplasts*

The *C. crispus* cDNA library used to isolate *GapC* and *GapA* clones [33, 59] was constructed from poly(A)$^+$ mRNA prepared from protoplasts of light-grown gametophytes. Protoplasts were used to reduce the contamination of nucleic acid preparations with cell wall polysaccharides. When this cDNA library was screened with homologous probes the frequencies obtained for clones *GapC* and *GapA* differed at least by one order of magnitude (*GapC*>>*GapA*). Since higher plants grown in the light contain transcript levels of *GapA* at least as high or higher than those of *GapC* we suspected that the low representation of *GapA* transcripts in *C. crispus* protoplasts may be a specific stress phenomenon. Therefore, we isolated poly(A)$^+$ mRNA from intact gametophyte tissue and protoplasts by using a new RNA/DNA purification method (see Materials and methods and [3]) which efficiently eliminates cell wall polysaccharides from rhodophyte extracts.

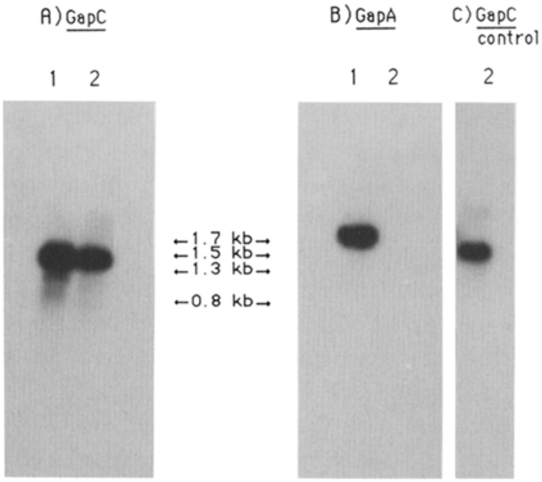

Fig. 6. RNA blot analysis of *GapC* (A) and *GapA* (B) transcripts from intact gametophytes (lane 1) and protoplasts (lane 2) of *Chondrus crispus*. In gametophytes both genes are expressed to the same extent giving rise to transcripts of 1.7 kb (*GapC*) and 1.5 kb (*GapA*), respectively. In protoplasts only *GapC* transcripts can be detected. C. Control hybridization with probe *GapC* of filter (B) previously hybridized with *GapA*.

In Fig. 6 hybridization patterns of RNA gel blots are shown using poly(A)$^+$ mRNA from gametophytes and protoplasts, respectively. For gametophytes a strong hybridization signal is observed for either gene probe at ca. 1.5 kb (*GapC*) and 1.7 kb (*GapA*) respectively (compare lanes 1 in Fig. 6A and B), suggesting that both genes are expressed to a similar extent. In protoplasts appreciable transcript levels can only be detected for *GapC* but not for *GapA* (compare lanes 2 in Fig. 6A and B, respectively). A control hybridization with the *GapC* probe of filter B previously hybridized with *GapA* clearly demonstrates that the absence of the *GapA* signal in lane 2 (B) is specific and not due to a blotting artifact (see Fig. 6C, lane 2).

Discussion

We have cloned and sequenced the first nuclear protein genes of red algae, those encoding cytosolic and chloroplast GAPDH of *C. crispus* (genes *GapC* and *GapA*, respectively). As shown by our previous phylogenetic analyses [33, 35] both genes are probably of eubacterial origin and became fixed in eukaryotic cells via independent intracellular symbioses leading to mitochondria (*GapC*) and chloroplasts (*GapA*), respectively. The data suggest that these endosymbiotic events happened before the separation of green plants, red algae, animals and fungi, so that trees based on *GapC* and *GapA* sequences, respectively, reflect the 'true' topologies of these major eukaryotic lineages. The branching patterns [33, 59] show that red algae separated from green plants at about the same time or somewhat later than fungi and animals. Although basically similar, nuclear genes and genomes of higher plants, fungi and animals show specific differences with respect to transcriptional control, pre-mRNA splicing and DNA methylation. As discussed below nuclear genes from red algae may have to offer some novelties in this respect.

Potential cis-*acting elements in the promoters of nuclear rhodophyte genes* (GapC *and* GapA *of* Chondrus crispus)

The basic features of transcriptional mechanisms are remarkably conserved in all eukaryotes [23]. Therefore, one would expect to find sequence motifs of universal *cis*-acting elements in nuclear genes of red algae. However, there are some interesting differences in this respect between the two genes *GapC* and *GapA* (see Fig. 1). The *GapC* gene is G + C-rich (see below), carries two potential TATA and CCAAT boxes each and multiple copies of GC boxes (GGGCGG, see [14]) in the 5′ upstream region. In addition, there are at least three upstream sequence elements related to plant-specific G-box motifs (GCCACGTGGC). In higher plants this palindromic sequence and related variations are found in several classes of promoters (e.g. *rbcS*, chalcone synthase, histone genes) where they interact with a family of DNA-binding proteins such as TAF-1, GBF and CG-1 [27, 30, 41]. As opposed to this, the only classic motif found in the *GapA* promoter is a single CCAAT box element, while a TATA box, known to specify the transcription start point in eukaryotes [14, 23], is absent in this gene. These differences with respect to classic motifs contrast with five '*Chondrus*-specific' sequence motifs shared by the two promoters. They are arranged in a similar way in both genes (Box A→Box B; Box 1→Box 2→Box 3, see Fig. 1) but their potential role in transcriptional regulation cannot yet be defined.

Conservation of a rhodophyte pre-mRNA intron with yeast-like splice junctions and a conspicuous secondary structure

There is now clear evidence [33, 35] that the nuclear genes *GapC* and *GapA* are descendants of two eubacterial genes which arose by a gene duplication in ancient eubacteria before the separation of the lineages leading to present day purple bacteria/mitochondria and cyanobacteria/chloroplasts, respectively. It is interesting in this context, that nuclear genes *GapC* and *GapA* contain introns in conserved positions [28, 32, 39, 47], suggesting that these introns may be archetypical relics which were present in the parental eubacterial gene before the *GapC*/*GapA* separation. It appears from this that genes in ancient eubacteria not only occurred in multiple copies [33, 35] but also contained intervening sequences prior to the evolutionary 'streamlining' of their genomes, in agreement with the 'intron early hypothesis' [13, 21].

Since at least five of the introns found in nuclear GAPDH genes may be as old as the GAPDH-coding sequences [28, 32], our finding of a single intron in the *GapA* gene of *C. crispus* (and none in the *GapC* gene, see Fig. 1) is somewhat surprising. It probably indicates that differential loss of introns has occurred in red algae as it has in other eukaryotic lineages (e.g. yeast [16]). The remaining intron in the *GapA* gene occupies a conserved position (see Fig. 3A) suggesting that it fulfills an essential function. We are presently testing the possibility of its implication in *GapA* transcriptional regulation in green plants and red algae.

Another interesting question concerns the mechanism of pre-mRNA splicing in red algae. The 5′ and 3′ splice junctions and the branch site of the *C. crispus GapA* intron are compatible with those of pre-mRNA introns of higher plants and animals but resemble also the highly conserved consensus sequences of yeast introns. In addition, the *C. crispus GapA* intron folds into a conspicuous secondary structure (see Fig. 2). Specific secondary structures are characteristic features of group I and group II introns [29, 43] but are usually absent in pre-mRNA introns. In a first approach to clarify whether secondary structure may be a necessary prerequisite of intron splicing in red algae we are trying to establish whether or not the folding pattern found for the *GapA* intron is characteristic of pre-mRNA introns of *C. crispus* in general.

Genes GapC *and* GapA *of* Chondrus crispus *show differential G + C enrichment and no CpG suppression*

Short- and long-range fluctuations in G + C content are characteristic features of vertebrate genomes [2, 5] and have recently also been reported for genomes of monocot angiosperms [1, 39, 46, 50]. In monocot genes the G + C content in the degenerate third base position of codons varies between 40% for genes of storage proteins and almost 100% for highly expressed genes regulated by light or hormones [9]. In contrast, dicot genes have comparatively low and homogeneous G + C contents [9, 50] irrespective of whether they are strongly or weakly expressed. This raises the question of whether genomic G + C variation is an ancient feature of eukaryotic genomes or whether it evolved in more recent times in vertebrates and independently a second time in monocot angiosperms. Our finding that genes *GapC* and *GapA* and their flanking sequences of the red alga *C. crispus* differ greatly in G + C content (see Fig. 4) would argue in favour of the former alternative. This implies that the homogeneous G + C distribution found in dicot genomes is a comparatively recent acquisition which evolved in dicot angiosperms after their separation from monocots.

In maize the light-regulated *GapA* gene is almost saturated with G + C (97% in the third base position of codons) while the constitutively expressed *GapC* gene is moderately G + C-enriched (67% in the degenerate codon position) [9]. In *C. crispus*, the situation is reversed with *GapC* being more strongly biased than *GapA* (89% versus 73% G + C in the third base position of codons, see Fig. 4A). G + C enrichment in synonymous sites of codons may be a ‘neutral’ consequence of strong gene expression rather than a prerequisite as previously suggested [9]. However, G + C enrichment does not occur in the first and second base position of codons (see Fig. 4A, values in brackets) and, hence, does not influence the non-synonomous mutation rate in GAPDH genes on which phylogenetic inferences are based [33, 35, 38, 59].

Genomic DNA of vertebrates and higher plants is highly methylated at the 5 position of cytosines of CpG dinucleotides [22, 52]. A direct consequence of CpG methylation is CpG depletion [6, 18, 46] caused by spontaneous deamination of 5-methylcytosine to thymine and the failure of DNA repair mechanisms to recognize these mutations. According to the model of Bird [6] only so-called CpG islands of housekeeping genes, comprising about 1.5 kb of DNA surrounding the transcription start site, are protected against methylation and, hence, CpG depletion [1, 18, 19, 39, 46]. In vascular plants cytosine methylation also occurs at CpNpG trinucleotides [4, 22], however, CpNpG suppression can usually not be observed [4, 20] possibly because of a CpNpG specific repair mechanism. Methylation studies with the green alga *Chlamydomonas reinhardtii* [4] revealed little or no CpG methylation and only a limited amount of CpNpG methylation in this organism.

We have sequenced over 6 kb of relatively G + C-rich genomic DNA from the red alga *C. crispus* with more than 500 potential CpG methylation sites (Fig. 4). If these sequences were methylated one would expect to see CpG suppression at least in the 3′ parts of genes and the adjacent flanking sequences. Our analyses show that the frequencies of CpG doublets (and CpNpG triplets; see Fig. 4) in genes *GapC* and *GapA* and the surrounding genomic regions do not deviate significantly from their expected values suggesting that they are not methylated. This could mean that CpG methylation is absent in red algae as it seems to be in green algae [4]. Direct methylation studies will be required to confirm this conclusion.

GapC *and* GapA *of* Chondrus crispus *are probably encoded by single loci composed of multiple alleles giving rise to RFLPs*

In higher eukaryotes genes *GapC* and *GapA* occur as gene families of variable sizes. The number of genes and pseudogenes encoding glycolytic GAPDH in vertebrates varies between a single copy in chicken [57], ten to 30 copies in man, hare, guinea pig and hamster, and over 200 copies in mouse and rat [24, 44]. For maize, multiple genes for both *GapC* and *GapA* have been described. The maize *GapC* gene family is composed of four functional members which differ significantly in sequence and which are expressed differentially under aerobic and anaerobic conditions [39, 49]. For maize *GapA* more than 10 separate genes can be distinguished by Southern blotting most of which are probably nonfunctional pseudogenes [46].

In *C. crispus*, *GapC* and *GapA* are encoded by three and two separate genes (alleles), respectively, of very similar primary structure. This is clearly indicated by our finding that patterns of genomic Southern blots are identical whether generic or gene specific probes are used under stringent hybridization conditions (Fig. 5A and B). In addition, in the case of *GapC* (Fig. 5A), the number and intensity of restriction fragments varies for different restriction enzymes suggesting that also the flanking sequences of *GapC* arè relatively conserved and polymorphic only for certain recognition sites. The most plausible explanation for these results is that *GapC* and *GapA* of *C. crispus* are encoded by single loci composed of multiple polymorphic alleles. The alternative explanation that both genes are encoded by multiple but highly conserved loci seems relatively unlikely. Since the genomic DNA was extracted from a mixture of different haploid *C. crispus* gametophytes (see Materials and methods), the different alleles detected by genomic Southern blotting would correspond to separate haploid individuals.

Differential gene expression in Chondrus crispus *protoplasts*

Our northern hybridizations clearly suggest that both genes, *GapC* and *GapA*, are vigorously expressed in light grown *C. crispus* gametophytes. In protoplasts *GapC* transcript levels seem to decrease only slightly, while *GapA* mRNAs fall below the limits of detection. This agrees with our previous finding [33] that *GapA* specific clones are very rare and at least 10 times less frequent than *GapC* clones in cDNA libraries of *C. crispus* protoplasts. Hence, the half-life of *GapA* transcripts seems rather short since they disappear almost completely during a 12 to 15 h period of protoplast preparation. These differential effects on nuclear gene expression suggest that the reaction of the cell to protoplast formation is specific and may predominantly affect those nuclear genes encoding plastid proteins. It has been shown for green plants that chloroplasts are strongly damaged upon protoplast isolation leading to loss of photosynthetic activity and to a block in the synthesis of nuclear encoded plastid proteins (see review by Galun [17]). Numerous gene expression studies with chloroplast deficient higher plants (mutants, herbicide treated seedlings) suggest that the transcriptional activity of plastid specific genes in the nucleus is dependent on intact chloroplasts and it has been speculated that a non-proteinaceaous chloroplast signal may be involved in this intercompartimental communication (for reviews see [42, 58]).

Acknowledgements

This work was financially supported by Deutsche Forschungsgemeinschaft (Ce 1/15-1) und Deutscher Akademischer Austauschdienst (PROCOPE). We thank Catherine Boyen (Roscoff) for her help in genomic DNA preparation and Michael Donath for providing the PCR amplified *GapA* intron probe.

References

1. Antequera F, Bird AP: Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J 7: 2295–2299 (1988).
2. Aota S, Ikemura T: Diversity in G + C content at the third base position of codons in vertebrate genes and its cause. Nucl Acids Res 14: 8129–8144 (1986).
3. Apt KE, Grossman AR: Characterization and transcript analysis of the major phycobiliprotein subunit genes from *Aglaothamnion neglectum* (Rhodophyta). Plant Mol Biol 21: 27–38 (1993).
4. Belanger FC, Hepburn AG: The evolution of CpNpG methylation in plants. J Mol Evol 30: 26–35 (1990).
5. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F: The mosaic genome of warm-blooded vertebrates. Science 228: 953–958 (1985).
6. Bird AP: CpG-rich islands and the function of DNA methylation. Nature 321: 209–213 (1986).
7. Branlant G, Branlant C: Nucleotide sequence of the *Escherichia coli gap* gene. Different evolutionary behaviour of the NAD + -binding domain and the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. Eur J Biochem 150: 61–66 (1985).
8. Brinkmann H, Cerff R, Salomon M, Soll J: Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach. J Plant Mol Biol 13: 81–94 (1989).
9. Brinkmann H, Martinez P, Martin W, Quigley F, Cerff R: Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde 3-phosphate dehydrogenase from maize. J Mol Evol 26: 320–328 (1987).
10. Cerff R: Separation and purification of NAD- and NADP-linked glyceraldehyde-3-phosphate dehydrogenases from higher plants. In: Edelmann M, Hallick RB, Chua NH (eds) Methods in Chloroplast Molecular Biology, pp. 683–694. Elsevier Biomedical Press, Amsterdam (1982).
11. Cerff R, Kloppstech K: Structural diversity and differential light control of mRNAs coding for angiosperm glyceraldehyde-3-phosphate dehydrogenases. Proc Natl Acad Sci USA 79: 7624–7628 (1982).
12. Doolittle RF, Feng DF, Anderson KL, Alberro MR: A naturally occuring horizontal gene transfer from a eukaryote to a prokaryote. J Mol Evol 31: 383–388 (1990).
13. Doolittle WF: The origin and function of intervening sequences in DNA: a review. Am Nat 130: 915–928 (1987).
14. Dynan WS, Tjian R: Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316: 774–778 (1985).
15. Feinberg AP, Vogelstein B: A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137: 266–267 (1984).
16. Fink GR: Pseudogenes in yeast. Cell 49: 5–6 (1987).
17. Galun E: Plant protoplasts as physiological tools. Annu Rev Plant Physiol 32: 237–266 (1981).
18. Gardiner-Garden M, Frommer M: CpG islands in vertebrate genomes. J Mol Biol 196: 261–282 (1987).
19. Gardiner-Garden M, Frommer M: Significant CpG-rich regions in angiosperm genes. J Mol Evol 34: 231–245 (1992).
20. Gardiner-Garden M, Sved JA, Frommer M: Methylation sites in angiosperm genes. J Mol Evol 34: 219–230 (1992).
21. Gilbert W: The exon theory of genes. Cold Spring Harbor Symp Quant Biol 52: 901–905 (1987).
22. Gruenbaum Y, Stein R, Cedar H, Razin A: Sequence specificity of methylation in higher plant DNA. Nature 292: 860–862 (1981).
23. Guarente L, Bermingham-McDonogh O: Conservation and evolution of transcriptional mechanisms in eukaryotes. Trends Genet 8: 27–32 (1992).
24. Hanauer A, Mandel JL: The glyceraldehyde 3-phosphate dehydrogenase gene family: structure of a human cDNA and of an X chromosome linked pseudogene; amazing complexity of the gene family in mouse: EMBO J 3: 2627–2633 (1984).
25. Hohn B: In vitro packaging of l and cosmid DNA. Meth Enzymol 68: 299 (1979).
26. Joshi CP: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl Acids Res 15: 6643–6653 (1987).
27. Katagiri F, Chua NH: Plant transcription factors: present knowledge and future challenges. Trends Genet 8: 22–27 (1992).
28. Kersanach R, Brinkmann H, Liaud M-F, Zhang D-X, Martin W, Cerff R: Five identical intron positions in ancient duplicated genes of eubacterial origin. Nature, in press (1993).
29. Krainer AR, Maniatis T: RNA splicing. In: Hames BD, Glover DM (eds) Transcription and Splicing, pp. 131–206. IRL Press, Oxford (1988).
30. Kuhlemeier C: Transcriptional and post-transcriptional regulation of gene expression in plants. Plant Mol Biol 19: 1–14 (1992).
31. Le Gall Y, Braud JP, Kloareg B: Protoplast production in *Chondrus crispus* gametophytes (Gigartinales, Rhodophyta). Plant Cell Rep 8: 582–585 (1990).

32. Liaud M-F, Zhang DX, Cerff R: Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate genes to the nucleus. Proc Natl Acad Sci USA 87: 8918–8922 (1990).
33. Liaud M-F, Valentin C, Martin W, Bouget FY, Kloareg B, Cerff R: The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrogenases from *Chondrus crispus*. J Mol Evol, in press (1993).
34. Maniatis T, Fritsch EF, Sambrook G: Construction of genomic libraries. In: Molecular Cloning: A Laboratory Manual, pp. 270–307. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
35. Martin W, Brinkmann H, Savona C, Cerff R: Evidence for a chimeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci USA 90: 8692–8696 (1993).
36. Martin W, Cerff R: Prokaryotic features of a nucleus-encoded enzyme. cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*). Eur J Biochem 159: 323–331 (1986).
37. Martin W, Gierl A, Saedler H: Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339: 46–48 (1989).
38. Martin W, Lydiate D, Brinkmann H, Forkmann G, Saedler H, Cerff R: Molecular phylogenies in angiosperm evolution. Mol Biol Evol 10: 140–162 (1993).
39. Martinez P, Martin W, Cerff R: Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Biol 208: 551–565 (1989).
40. Michels PAM, Marchand M, Kohl L, Allert S, Wierenga RK, Opperdoes FR: The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distant evolutionary relationship. Eur J Biochem 198: 421–428 (1991).
41. Oeda K, Salinas J, Chua NH: A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J 10: 1793–1802 (1991).
42. Oelmüller R: Photooxidative destruction of chloroplasts and its effect on nuclear gene expression and extraplastidic enzyme levels. Photochem Photobiol 49: 229–239 (1989).
43. Perlman PS, Peebles CL, Daniels C: Different types of introns and splicing mechanisms. In: Stone EM, Schwartz RJ (eds) Intervening Sequences in Evolution and Development, pp. 112–161. Oxford University Press, Oxford (1990).
44. Piechaczyk M, Blanchard JM, Sabouty SRE, Dani C, Marty L, Jeanteur P: Unusual abundance of glyceraldehyde 3-phosphate dehydrogenase pseudogenes in vertebrate genomes. Nature 312: 469–471 (1984).
45. Prabhala G, Rosenberg GH, Käufer NF: Architectural features of pre-mRNA introns in the fission yeast *Schizosaccharomyces pombe*. Yeast 8: 171–182 (1992).
46. Quigley F, Brinkmann H, Martin W, Cerff R: Strong functional GC pressure in a light regulated maize gene encoding subunit GAPA of chloroplast glyceraldehyde-3-phosphate dehydrogenase: implications for the evolution of GAPA pseudogenes. J Mol Evol 29: 412–421 (1989).
47. Quigley F, Martin W, Cerff R: Intron conservation across the prokaryote-eukaryote boundary: Structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. Proc Natl Acad Sci USA 85: 2672–2676 (1988).
48. Ruby SW, Abelson J: Pre-mRNA splicing in yeast. Trends Genet 7: 79–85 (1991).
49. Russell DA, Sachs MM: The maize cytosolic glyceraldehyde-3-phosphate dehydrogenase gene family: organ-specific expression and genetic analysis. Mol Gen Genet 229: 219–228 (1991).
50. Salinas J, Matassi G, Montero LM, Bernardi G: Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. Nucl Acids Res 16: 4269–4285 (1988).
51. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
52. Shapiro HS: Distribution of purines and pyrimidines in deoxyribonucleic acids. In: Fasman GD (ed.) Handbook of Biochemistry and Molecular Biology, pp. 241–275. CRC Press, Cleveland, OH (1976).
53. Shih MC, Lazar G, Goodman HM: Evidence in favor of the symbiotic origin of chloroplasts: Primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenases. Cell 47: 73–80 (1986).
54. Shih MC, Lazar G, Goodman HM: Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate- dehydrogenase from *Arabidopsis thaliana* (Corrigendum). Gene 119: 317–319 (1992).
55. Sinibaldi RM, Mettler IJ: Intron splicing and intron-mediated enhanced expression in monocots. Progr Nucl Acid Res Mol Biol 42: 229–257 (1992).
56. Smith TL: Disparate evolution of yeasts and filamentous fungi indicated by phylogenetic analysis of glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci USA 86: 7063–7066 (1989).
57. Stone EM, Rothblum KN, Alevy MC, Kuo TM, Schwartz RJ: Complete sequence of the chicken glyceraldehyde-3-phosphate dehydrogenase gene. Proc Natl Acad Sci USA 82: 1628–1632 (1985).
58. Taylor WC: Regulatory interactions between nuclear and plastid genomes. Annu Rev Plant Physiol Plant Mol Biol 40: 211–233 (1989).
59. Liaud MF, Valentin C, Bouget FY, Kloareg B, Cerff R: Molecular phylogeny of red algae as revealed by nuclear genes encoding chloroplast and cytosol specific proteins. In: Sato S, Ishida M, Ishikawa H (eds) Endocytobiology V, pp. 357–361. Tübingen University Press (1993).