

Characterization of the UDP-glucose pyrophosphorylase gene from the marine red alga *Gracilaria gracilis*

Arturo O. Lluisma^{1,2} & Mark A. Ragan^{1,3,*}

¹Institute for Marine Biosciences, National Research Council of Canada, Halifax, Nova Scotia, B3H 3Z1 Canada ²Present address: Marine Science Institute, University of the Philippines, 1101 Quezon City, Philippines ³Canadian Institute for Advanced Research, Program in Evolutionary Biology

(*Author for correspondence; phone +1 902 426 1674; fax +1 902-426-9413; e-mail mark.ragan@nrc.ca)

Received 16 October 1998; revised 20 October 1998; accepted 21 October 1998

Key words: UDPglucose pyrophosphorylase, polysaccharide biosynthesis, red algae

Abstract

UDP-glucose pyrophosphorylase (UGPase) is a key enzyme in carbohydrate metabolism, particularly polysaccharide biosynthesis, in red algae. In this report, we characterize at the genomic and cDNA levels the putative UGPase gene of the agarophytic red alga *Gracilaria gracilis*. The gene is single-copy, devoid of introns, and produces two kinds of transcripts that differ in size by 332 basepairs. The large and small transcripts appear to utilize distinct polyadenylation signals. The putative protein has 495 amino acids, and is about 50% identical in sequence to its homologs in plants, animals and fungi. Sequencing of the genomic clone revealed that another gene, potentially encoding a DNA helicase and containing a 76 bp-intron near its 3' end, occurs 376 bp downstream of the UGPase gene.

Introduction

UDP-glucose (UDPGlc) plays a key role in carbohydrate metabolism in red algae (Rhodophyceae). Its main biosynthetic function is as precursor of UDP-galactose (UDPGal), the D-galactosyl donor in the biosynthesis of galactans and of floridoside (α -D-galactopyranosyl- $(1 \rightarrow 2)$ -glycerol) (Su & Hassid, 1962; Kremer & Kirst, 1981; Manley & Burns, 1991). Galactans, such as the commercially important agarans and carrageenans, are the most abundant component of the cell walls in most red algae (Craigie, 1990), whereas floridoside is a key photosynthetic product that functions as a short-term low-molecularweight carbohydrate reserve (Kremer, 1978; Macler, 1986). There is some evidence that UDPGlc also serves as a minor glucosyl donor in the biosynthesis of floridean starch (Nagashima et al., 1971).

The key enzyme for the biosynthesis of UDP-Glc is UDP-glucose pyrophosphorylase (UGPase; EC 2.7.7.9), for which UTP and Glc1P are substrates. Plant, animal, and eubacterial UGPases are well characterized at the protein and gene levels (e.g., Turnquist & Hansen, 1973; Kleczkowski, 1994). The occurrence of this enzyme in red algae was first demonstrated in the early 1960s when UDPGlc and other metabolites of galactan biosynthesis were isolated from *Porphyra perforata* (Su & Hassid, 1962). More recently, Manley and Burns (1991) demonstrated UGPase activity in the red alga *Pterocladia capillacea*. However, characterization of red algal UGPases at the protein and gene levels has not yet been accomplished.

Herein we report the cloning of the nuclear gene encoding UGPase from the agarophytic marine red alga *Gracilaria gracilis* (Stackh.) Steentoft, Irvine & Farnham, and describe some features of its cDNA and genomic sequences.

Materials and methods

Construction and screening of genomic libraries and sequencing of clones

DNA was extracted from *G. gracilis* ('grass' strain) as described by Zhou & Ragan (1993), partially digested with *Sau*3AI, and ligated to the Lambda-DASH II vector (Stratagene, LaJolla, CA). The recombinant phage was packaged using the Gigapack III Gold Packaging Extract (Stratagene) and propagated in *Escherichia coli* XL1-Blue MRA (P2) strain (Stratagene).

A homologous UGPase probe was produced by a PCR-mediated approach. Degenerate primers were designed based on highly conserved regions of UG-Pases, as revealed by a multiple alignment of UGPase sequences obtained from the NCBI protein database. With genomic DNA from G. gracilis as template, PCR amplification with one pair of primers ('A', 5'-GGNGGN[TC]T[GT]GGNAC-[GT][AT]C[GT]ATGGG-3', corresponding to the highly conserved amino acid sequence GGLGTSMG, and 'B', 5'-[GA]T[TC]NCC[GA]TG[GT]CCNGG-[GT]GG[GA]TACCA-3', corresponding to the highly conserved amino acid sequence WYPPGHGD) yielded a product that was confirmed by subsequent sequencing to encode part of a potential UGPase gene. Using this PCR product as a probe, we screened the genomic library of G. gracilis we have prepared and the G. gracilis genomic library constructed by Zhou & Ragan (1994). Positive clones were purified using standard protocols, and subjected to automated sequencing on an ABI 373 'stretch' sequencer (Perkin Elmer – Applied Biosystems, Foster City, CA) using the manufacturer's protocols.

3' RACE

mRNA was extracted from laboratory-grown *G. gracilis* using the Invitrogen FastTrack 2.0 kit (Invitrogen, Carlsbad, CA). The 3' end of the UGPase transcript was reverse-transcribed using the Pharmacia T-primed Ready-To-Go kit (Pharmacia Biotech, Uppsala), and PCR-amplified using primer f2b (positions 746–767, Figure 1) as the gene-specific primer, together with an anchor primer (5'-ATTCGCGGCCGCAGGAATT-3'). As two PCR products were found on the gel, these were separately reamplified by PCR using the same primers. The PCR products were desalted by centrifugation through Centricon-100 concentrators (Amicon, Beverly, MA) with 2 mL distilled water, and used directly for sequencing.

Southern analysis

Southern blotting experiments were carried out following standard protocols (Sambrook et al., 1989). The probe was synthesized and labeled with α -³²PdCTP using the Random Primed DNA labeling kit (Boehringer Mannheim Canada, Laval, Qué.), using a portion of the cloned DNA fragment (positions –33 to 424 in Figure 1; amplified by PCR) as template.

Results and discussion

Cloning of the G. gracilis UGPase gene

Using a homologous probe to screen a G. gracilis genomic library, two clones were initially recovered, each of which was found to contain a UGPase-coding region truncated near the 3'-end; one clone (λ 27) was sequenced. To obtain the sequence of the 3'-end and 3'-flanking region, another clone (\laulde{UGPb}) was isolated from another G. gracilis genomic library (Zhou & Ragan, 1994). The two clones were 100% identical in sequence throughout the 339-bp region of overlap (data not shown); and as UGPase is furthermore encoded by a single gene in G. gracilis (see below), λ UGPb almost certainly contains the same gene as $\lambda 27$. The final sequence, reconstructed from clones $\lambda 27$ and $\lambda UGPb$ (Figure 1), contains an ORF of 496 codons. Comparison of the deduced amino acid sequence with other UGPase sequences (Figure 2) shows significant sequence identity between the deduced protein and UGPases from Solanum tuberosum (50%), human (51%) and Saccharomyces cerevisiae (48%), suggesting strongly that the isolated gene encodes a UGPase. We designate this gene as *GgUGP*.

The 5' flanking region has proven recalcitrant to sequencing. Also proving difficult was the sequencing of the 5'-end of the transcript by the 5'-RACE technique, using methodology that has yielded the 5' ends of other *G. gracilis* genes (*GgGALT1* and *GgSBE1*; Lluisma & Ragan, 1998a).

Features of the gene

A number of ATG codons in-frame with the UGPaseencoding ORF are present just downstream of the only potential TATA box (Figure 1). In the absence of data on the location of the 5' end of the transcript, it is not possible to provide a definitive identification of the start codon. The first ATG codon downstream of the

tgcacccccccccgctccgctttgttcgcttcttgcaccccccgttccaccgcttcacccgtggaaccttctgctgctat	-60
cccgttcgcactgctgttcg <mark>tttatttaa</mark> gccaccatgatgccaaacggaaaaggagcactgatcgcgactccaggtctc M N R D S R S	22
tg caggact t caaggg cg t catgg a caag t ccg ccg ccg ccg ccg a gaag a g ct catg t catgg a ccag a t g c c g c g c g c g c g c g c g c g c	103
L Q D F K G V M D K S A A S T V A E K L T V M N Q M A ccaatgagctcgagaagatgaccgattctgagaccaccggcttcgtcggtgacggccgctacatgagcgaacgttcga	184
A N E L E K M T D S E T T G F V E L Y G K Y M S E K S aaaaggeegaaateaagtgggateteategaacageecagtgaaaacatgetgeaaaagtacgttgecaaageegg	265
Ccaccgacgaagaactcgcttcgctcctttccaaggtggctcaaggtcgggggggg	346
gcaaaggacccaagagcgtcatcgaagtgcgtgatgacaccacctcttggacctcattgttcagcagattggtcagctca C K G P K S V I E V R D D T T F I, D L I V O O I G O L	427
acaagaaccatccacggccaacgtccccctgcttctcatgaactctttcaacaccgactctgagaccgcaaagatcattc N K N H P T A N V P L L L M N S F N T D S E T A K I I	508
gcaagtaccaggataccagtgttaccctcaccaccttccagcaatctcgttaccccaggatcgtcaaggagtctctcgaac	589
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	670
actcgggattggttgatacccttcttgcgcagggtaaggagtacatctttgtctcaaacgttgacaatcttggcgccactg N S G L V D T L L A Q G K E Y I F V S N V D N L G A T	751
$\label{eq:constraint} \begin{array}{cccc} \underline{tcqactctcacattctc} \\ \hline cqatctcaacattctc} \\ V & D & L & N & I & L & K & N & V & V & D & R & E & V & E & Y & C & M & E & L & T & D & K & T & R & A & D \\ \end{array}$	832
tcaagggtggtaccattatttcatacgacggaaaggtgtctctgctcgaggtcgctcaggtccctgccaagtacgttgaag I K G G T I I S Y D G K V S L L E V A Q V P A K Y V E	913
agttcaagtctatttccaagtttaaggtgttcaacaccaataacatttgggtatcgttacgagcaatcaagcgcgtcatgc E F K S I S K F K V F N T N N I W V S L R A I K R V M	994
aatccggcgagatgaagcttgatatcattgtgaacaacaaggaagtcaagggtaccaaggtcatccagctcgaaagcgcta Q S G E M K L D I I V N N K E V K G T K V I Q L E S A	1075
ttggagcagccattggctacttcaacaatgcgtgcggtgtcaacgttcctcgttccctgttcctgtcaagtcaacct	1156
ccgatctcatgctcattcagagcaacatgtacaaccttaaatctggctctctggttatgaaccctgcccgccatttacta	1237
S D L M L I Q S N M Y N L K S G S L V M N P A R Q F T caactccagtgattaagcttggaaaggagtttaa <u>gaaqqtcqctcaatacttgaac</u> gtctgggtagcatcccetgacattt	1318
TTPVIKLGKEFKKVAQYLERLGSIPDI	
tggagettgaceateteaetgteteaggtgatgtetaetttggggetaaeaetaetetgaaaggaaeegttategtggtag TGGAGETTGACEAETGTCTCAGGGGATGTCTACTTTGGGGGCTAACAETAEGTGAAGGAACCGTTATCGTGGTAG L E L D H L T V S G D V Y F G A N T T L K G T V I V V	1399
caaaccctgggaacaccatcatgattccagaaggctcagttctcgagaacaaggtcgttcttggttctctccatgtgattcccatgttcccatgtga	1480
ANPGNTIMIPEGSVLENKVVLGSLHVI	
cgc <u>attaaa</u> catgtcctggtaggctacggtcagtgcgcgtactatacattaagcatcttcatttggtag <u>gcccqctatcqc</u> CGC <u>ATTAAA</u> CATGTCCTGGTAGGCTACGGTCAGTGCGCGTACTAAAAAAAA	1561
r n agcatgtgc agcatgtgc sf5 > ATCCTGAAGCAATCGGTGAGCCCATCAGTTAGAATGTTGGTCCATGCCGTGGTTGAATCACGAATGG	1642
ataccgacattttcggcatgaaggaccattcatgacctcgtcctgcgaccgtactacgagcgag	1723
$cgaacaatttactgctacgacacgacgcgtagtgagttagcgaattagtttgtcgaagggcacctcaatctgatacgcgct\\ CGAACAATTTACTGCTACGACACGACGCGTAGTGAGTTAGCGAATTAGTTTGTCGAAGGGCACCTCAATCTGATACGCGCT\\ \label{eq:cgacacgacgcgtagtgagttagcgaattagtttgtcgaagggcacctcaatctgatacgcgct}$	1804
aatacacgaattc <u>aataaa</u> caatgcaagaactatagggtgtttcggttttgaacttaatc <u>ttactctqqtgcqqtccqqct</u> AATACACGAATTC <u>AATAAA</u> CAATGCAAGAACTATAGGGTGTTTCGGTTTTGAAAAAAAA	1885
$\underline{tccgatgaattcatagttttccggtggcttctcccccccgagacttagctaccgatggttgcgccatgtcaagcagaaa}$	1966
agtacgtgcttcagcaattgcggtaccgagatcgagatccaaggagtcaaccagaacaggaactgggcaatccattccggaag	2047 2128
<pre>ccctgcgcactgagccgcttggcgcattgcctattgtccgttttggtgacgattcagaagaagaactgttagatttctcct</pre>	2209
agcacagcattgtaccttggcaaaagactgcagaacaaaacttacatcgaaaacaaggaagtcgtcctcccgtatttggta cttttctgccagaaaccgcaggcgagcctttaacactctggattcggtgtattggaaa	2290 2348

Figure 1. The nucleotide sequence of GgUGP. The numbering scheme assigns +1 to the first position of the putative start codon (marked with 1). The putative TATA box is shown in underlined boldface. The conceptual translation is shown below the coding region; * marks the stop codon. The 3'-RACE revealed two polyadenylation sites; the putative polyadenylation signal (ATTAAA) of the first site, which overlaps with the termination codon, and that of the second site (AATAAA), are underlined. The sequences of the 3'-ends of the cDNAs are aligned with the genomic sequence and shown in uppercase letters. Positions of the sequencing primers f4b and f5 are underlined, as is the position of primer f2b used in the RTPCR (positions 746–767). The portion of the sequence representing the 3' end of a putative DNA helicase gene (encoded on the complementary strand) is italicized and underlined. Within the helicase gene region, the putative intron is shown in boldface.

	10	20	30	40	50	60	70	80	90	
Ggra :	:	MN	RDSRSLQDFKGV	MDKSAAST	AEKLTVMNQM	ANELEKMI	DSETTGFVEI	LYGRYMSERSK	KAEIKWDLI	: 71
Hsap		MSRFVQDLSKA	MSQDGASQFQE	VILQELELS	KKELEKILTT2	ATSHEYEHI	KKDLDGERKI	LYHRFLQE K	GPSVDWGKI	: 78
Btau : Hunul ·		MSREVQULSKA	MSQDGASQFQE	ADSKIDCI	KKELEKILTT?	APSHEFEHI	KKDLDGPRKI	LFHRDIQE K	GPSVDWGKI	: 78
Scer		MSTKKH	TKTHSTVAFESN	TNSVAASO	RUAVARLGEIS		ENERAGE SI	-VSRIESGE	AEQUEWSKI	: 53
Stub :			MATAATLSE	ADAEKLNN	KSAVAGLNOIS	5	DNEKSGEINI	VGRAUSGE	AOHIDWSKI	
Sscr :		MSRFVQDLSKA	MSQDGASQFQEV	IRQELELS	KKELEKILTT2	- ApsHefeht	KKDLDGERKI	FHRE OE K	GPSVDWGKI	: 78
Ddis :	MTDTATSKATV	ERPKLQSTGSL	HSLFKDVDLFSE	NDEELYPP	QHGARFAAPIH	DSTLLALGMK	PDELKAFOK	RHATINK	D-QIYTDEI	: 89
Cele :					****		-MSNDQLKSP	LREFFDR	Q	: 17
	100	110	100	190	140	150	1.00	1 7 0		
Gora :	EOESENMIOK	DTIPKPATD	EELASLIISKIIA	TRINGELET	140 SMCCKCDKSVI	120 120		170	TRO	. 140
Hsap :	ORPPEDSIOPY	EKIKARGLP	DNISSVINKIVV	VKLNGGLGT	SMGCKGPKSLI	GVRNENDED	LTVOOID	ų	NKSYNTD-	145
Btau :	QRPPEDSIQPY	EKIKARGLP	dnvssviin <u>ki</u> vv	VKLNGGLGT	SMGCKGPKSLI	GVRNENTFLD	LTVQQIE	н	LNKTYDTD-	: 156
Hvul :	QTPTDEVVVPY	DTTAPPPEDL-	DAMKALLDKLVV	LKLNGGLGT	TMGC <mark>TGP</mark> KSVI	EVRNGFTFLD	LIVIQIE	s	LNKKYGCS-	: 132
Scer :	KSPNPDEVVKY	EIISQQPEN	VSNLSKLAV	LKLNGGLGT	SMGCVGPKSVI	EVREGNTFLD	LSVRQIE	Y	LNRQYDSD-	: 152
Stub :	QTPTDEVVVPY	DKLAPLSEDP-	AETKNLLDKLVV	LKLNGGLGT	TMGC TGPKSVI	EVRNGLTFLD	LIVKQIE	A	LNAKFGCS-	: 135
Ddig :	QREPEDSI OPY	EKIKARGLP		VKLNGGLGT	SMGCKGPKSLI	GVRNE NFFLD	LTVQQIE	H	NKTYNTD-	: 156
Cele :	PDOSEKAHO		SOFUENORU	IKINGGLGN	TMCCSKAKSIV	ETAPGVI FUD		WOIDNEEWEN	INODYNVD-	: 169
GEIG .	a goniangi	JOKIPAVBI		TKENGGEGI	THOCSRANSLY		CANCER OF IT	WQIDNEE WKN	HTFEGNQEQ	: 103
	190 2	200 23	10 220	23	0 240	250	260	270	28	
Ggra :		NVPLLLMN	SFNTDSETAKII	RKYQDTSVT	LTTFQQSRYPR	IVKESLEPMP	LTHDHYAH	EDWYPPGHGD	FEQSIYNSG	: 226
Hsap :		VPLVLMN	SENTDEDIKKIL	QKYNHCRVK	I Y HENOSRY DR	INKESLRPVA	KDVSSSGEST	EAWYPPGHGD	IYASFYNSG	: 234
Hvul :			SENTULDIANTS	ON INHERMA	THITTNOSOVPR	INTESLEVA	KNVSISGENT SKCOTCK	EAWY PPGHGD	I YASE YNSG	: 234
Scer :		VPLLLMN	SENHOKDREHLI	KKYSANRIR	IRSENOSREPR	VYKDSLLPVP	TEYDSPL	DAWYPPGHGD	UPESLHVSC	. 207
Stub :		VPLLLMNS	SFNTHDDTLKIV	EKYANSNID	IHTFNQSQYPR	LVTEDFAPLP	CKGNSGK	DGWYPPGHCD'	VEPSLMNSG	: 210
Sscr :		VPLVLMNS	SENTDEDTKKIL	OKYNHCRVK	IYTENQSRYPR	INKESLLPVA	KDVSYSGENT	EAWYPPGHGD	IYASFYNSG	: 234
Ddis :		VPLVIMNS	SYKTHNE TNKV I	ekykthkvs.	IKTFQQSMFPK	MYKDTLNLVPI	KPNTPMNP	KE <mark>WYPPG</mark> SGD	IFRSLORSG	: 245
Cele :	CRNPRKMCEAHI	1VD - TPLYLMNS	SFYTDEDTKKYL	AEKGYSN	VKTFVQSKCPR	LDAETKLPIEI	DENEDWGD	DAWCPPGHGN	IFQSLQNSG	: 191
	0 290	300	310	320	330	340	350	360	270	
Ggra :	LUTLIAO	TFVSNVDNLGA	TVDLNHIKNVV	DREV	CMPL ND RTRA	DIKCOULISM		MEAKYWEED	STSKARVA	: 315
Hsap :	LIDTFIEEGKEY	IFVSNIDNLGA	TVDLYILNHLI	NPPNGKRC	VMEVINKTRA	DVKGGTLTOY	GKLRLVEIA	OVPKAHVDEFI	SVSKFKIF	: 327
Btau :	LLDTFIGEGKEY	IFVSNIDNLGA	TVDLYILNHLM	NPPNGKPC	VMEVTNKTRA	DVKGGTL TOY	GKIRLVEHA	QVPKAHVDEF	SVSKFKIF	: 327
Hvul :	KLDTLLSQCKEY	VFVANSDNLGA	VI VDIKILNHLI	HNQN	<i>ic</i> mevtprtla	DVKGGTILI SYI	GRVQLLEIA	QVPDEHVDEFI	SIERFKIF	: 296
Scer :	EIDALIAQCREI	LFVSNGDNLGA	ATVDIKILNHMI	ETGA	IMELEDKTRA	DVKGGTL I SYI	GQVRILLEVA	QVPKEHIDEFI	NIRKETNE	: 316
Stub :	KUDALI AKEKDY	VEVANSDNLGA	AIVDIIKILNHLI.		OMEVIPKTLA	DVKGGTLISYI	EGKVQILLEIA	OVPDEHVNEF	SIEKFKIF	: 299
Ddis ·	LIDEFLAACKEV	TETSNUENLGS	AT VOID THE NHUM		CHINA MINIARA	DVKGGTLTQM	GKEREVEIA	OVERPHUDEF	SVSKEKTE	: 327
Cele :	VIDOLIADGREI	IFVSNIDNTGA	NTOLOTVOLMI	DKNVD	TMECTPKTOV		GRMMHLEMP	OVEAENUPDE	STRVERIE	· 280
										. 200
0	380	390	400	410	420	430	440	450	460	
Ggra :	NTNNI WVSLRAI	KRVMQSGEMKL	DIIVNNKI	SVKGTKV101	PSAICAAIGY	NNACGVNVP	(SRFILPVKST	SDLMLIQSNM	NHKSG-SI	: 403
Btau :	NTNNEWISLAAV	KRIOFONATON	EIIVNPKII	LDGGLNVIQI	THAVGAARS	ENSLGINVP	SRFLPVKITS SPFT DVKTT	SULLLVMSNLY	SINAG~SI	: 416 . 416
Hvul :	NTNNLWVNLKAI	KRLVDAEALKM	ÆIIPNPKI	VDGVKVLOI	DIAACAAIRF	EKAIGINVP	SRFLPVKAT	SDLLLVOSDL	TIVDG-YM	. 384
Scer :	NTNNLWINLKAV	KRLIESSNLEM	E I I PNQKTI TRI	OGHEINVLQI	FIACCAAIRH	DGAHGVVVP	SRFLPVKTC	SDILLIVKSDL	RLEHG-SI	: 408
Stub :	NTNNLWVNLSAI	KRLVEADALKM	ÆIIPNPKI	EVDGVKVLQI	елаасаанкг	DRAI GANVP	SRFLPVKAT:	SDLLLVQSDLY	TLTDEGYV	: 388
Sscr :	NTNNLWISLAAV	KRLQEQNAID	eiivnpkti	LDGGLNVIQI	DHAVCAARKS	ENSLGHNVP	SRFLPVKTT:	SDLLLVMSNLY	SLNAG-SL	: 416
Ddis :	NTNNIWVNLKSV	SNLIKEDKIDI	DWIVNYPI	LENHKAMVQI	DTPAGMGIQN	KNSVAIFVP	DRYRPIKST	SQLLVAQSNI E	QFDHG-QV	: 419
cere :	NTINNTYVINLKAV	KKLLPDHKS	SETUVIKK	TIRSREVLQI	PFSICCUKN	DNALCWHVE	KRERPVKNLO	JUL SURSTIC	DUDHS-TF	: 366
	470	480 4	90 500	51	.0 52	0 530) 54(0 550	1	
Ggra :	VMNPARQFTTTP	VIKLGKEFK-K	MAQYLERLGSI	DILELDHLI	VSGDVYFCAN	ITÜKGTVI VV#	NPGNTIMIP	EGSVLENKVVI	GSLHVIPH	: 495
Hsap :	TMSEKREFPTVP	LVKLGSSFT-K	VQDVLRRFESII	POMLELDHLI	VSGDVTFGKN	VSI.KGTVIII <i>P</i>	NHGDRIDIPI	PCAVLENKIVS	GNLRILDH	: 508
Btau :	TMSEKKEFPTVP	LVKLCSSBT-K	VODVLERFEST	DMLELDHLT	VSGDVIFGKN	VSILKETVITIA	NHGDRIDIPI	PCAVLENKIVS	GNLRILDH	: 508
Scer :	KLDPSSFGP-MP	Jacober Leik ~ K Likikeshok ~ K	WSGDNABT PHIL	KTVELDELT	TENVELOP		AGVALE IR	JGAVLENKDIE	GPEDI	: 473
Stub :	IRNPARSNPSNP	SIELGPERK-K	WANFLGREKS	SIIDIDSIK	VTGDVWEGSC	TLECKVTHAD	KSGVKUET DI	CAVIANKDUN	GPEDI	. 499
Sscr :	TMSEKREFPTVP	LVKLGSSFT-K	VODYLERFESI	DMLELDHLT	VSGDVTFGKN	SLKGTVIIIX	NHGDRIDIP	CAVLENKINS	GNLRILDH	: 508
Ddis :	KLNSKREGODVP	LVKLGEEFS-T	VSDYEKRFKSTI	DLLELDHLT	VSGDVYFGSR	TLKGTVINVA	NHGERVDIP	GVVLENKVLS	GTLRILDH	: 511
Cele :	KVHHNHELGAPP	VISIDPSIYNS	VEVVDLKFPHPI	VMDNCSEFA	VVGDVTFGKN	KLSGKVTVNG	KTESPGVVPI	GTVLKDQEYI	AE	: 453

Figure 2. Alignment of UDP-glucose pyrophosphorylase (UGPase) sequences. The sequences were obtained from the NCBI database and aligned, together with the UGPase sequence from *G. gracilis*, using CLUSTAL W (Thompson et al. 1994) under its default parameters: pairwise alignments = slow (accurate), gap opening penalty = 10, gap extension penalty = 0.10 or 0.05, and the BLOSUM series (Henikoff & Henikoff, 1992) for scoring. Residues shown in white on a black background are conserved in at least 90% of the sequences. Ggra, *Gracilaria gracilis* UGPase, this paper (NCBI accession number AF100788); Hsap, *Homo sapiens* UGPase (731050); Btau, *Bos taurus* UGPase (731049); Hvul, *Hordeum vulgare* UGPase (2117937); Scer, *Saccharomyces cerevisiae* UGPase (UGP1, 1585157); Stub, *Solanum tuberosum* UGPase (322794); Sscr, *Sus scrofa* UGPase (1752677); Ddis, *Dictyostelium discoideum* UGPase (136738); Cele, *Caenorhabditis elegans* UGPase (1326259).

putative TATA box is in the right context as a translation initiation site, but is probably too proximate (6 bp). Another ATG codon, 30 bp downstream of the putative TATA box (Figure 1), is also in the right context for translation initiation; its 5'-flanking sequence (GCCATG, the start codon in bold) conforms to the canonical sequence, RCYATG, at translation initiation sites in red algal genes so far characterized (Zhou & Ragan, 1996). We thus provisionally designate this codon as the translation start site of the UGPase gene.

The amino acid sequence deduced from the genomic sequence of *GgUGP* shows significant sequence similarity with other UGPase sequences throughout its entire length (Figure 2); no region in the deduced sequence appears to be an 'insertion' relative to the other sequences, providing *a priori* evidence that *GgUGP* is devoid of introns. Other intron-free nuclear genes have been reported in red algae, e.g., genes for triose phosphate isomerase in *G. gracilis* (Zhou & Ragan, 1995c) and for GapC in *Chondrus crispus* (Liaud et al., 1993). In contrast, the UGPase gene from potato is interrupted by 19 introns (Borovkov et al., 1997), while that of *Dictyostelium discoideum* is interrupted by three introns (Ragheb & Dottin, 1987).

Reverse transcription and PCR amplification of the 3' end of GgUGP transcripts were carried out to determine the location of polyA sites and polyA signals. Two PCR products differing in size were obtained, and sequence was obtained from both using primer sf4b; the sequence of the 3' end of the smaller amplicon is shown in Figure 1. The longer transcript was identical throughout the region of overlap; the sequence of its 3'end was obtained using another primer, sf5 (Figure 1). Alignment of the sequences with the genomic sequence (Figure 1) confirms that there are two types of UGPase transcripts in G. gracilis, differing in size by 332 bp due to alternative polyadenylation sites. Each site apparently has its own polyA signal, AUUAAA for the shorter transcript, found 33 bp upstream of the polyA site; and AAUAAA for the larger one, 32 bp upstream of the other polyA site. The polyA signal for the shorter transcript overlaps with the termination codon, a situation observed with another G. gracilis gene, polyubiquitin (Zhou & Ragan, 1995b). Interestingly, the polyubiquitin gene also has an alternative polyA site with its own polyA signal, and, just like the UGPase transcript, the polyA signal for the longer polyubiquitin transcript is also AAUAAA. The presence of UAAA signal in both polyadenylation sites in *GgUGP* and in the polyubiquitin gene probably indicates the importance of this highly conserved polyadenylation signal in *G. gracilis*. The UAAA motif has been observed in all gene trancripts so far characterized from *G. gracilis* (see Zhou & Ragan, 1996). The relevance of alternative polyadenylation sites in the GgUGP transcript to the physiology of *G. gracilis* remains to be evaluated.

Features of the deduced protein

GgUGP as delineated above potentially encodes a 495-amino acid protein that shares a high level of sequence identity (around 50%) with UGPases from other eukaryotes (Figure 2). Affinity-labelling studies of potato (Kazuta et al., 1991) and bovine (Konishi et al., 1993) UGPases have identified conserved lysine residues (K336, K402, K444, K486, K488, numbered according to the alignment in Figure 2) that are at, or close to, the substrate-binding site. Three of these (K336, K402, K444) have been confirmed by site-directed mutagenesis to be important for enzyme activity (Katsube et al., 1991); K444 is involved in catalysis of pyrophosphorylation per se, while K336 and K402 are believed to be involved in binding pyrophosphate or α -D-glucose-1-phosphate. These residues are highly conserved among UGPases, including GgUGP. There are, however, highly conserved residues for which mutation does not adversely affect enzyme activity, at least in the human liver UG-Pase (Chang et al., 1996); examples of such mutations include C140S, H311R, W263S, R437H, R471Q, and R495H. One of these sites, the highly conserved H311, is substituted in GgUGP with N.

UGPase is single-copy in G. gracilis

Southern analysis was performed to determine copy number in the G. gracilis genome. The final wash was of moderate stringency (0.5X SSC/0.1% SDS, 65°C, 30 min, performed twice); at this stringency, multiple copies of other G. gracilis genes (GgGALT1 and GgSBE1, Lluisma & Ragan, 1998b) were clearly detected. However, in the case of UGPase, only one band was found per lane (data not shown), indicating that the UGPase gene is single-copy in G. gracilis. A second band (< 1 kb) observed after restriction with XhoI can be explained by the presence of a XhoI restriction site (CTCGAG, positions 112-117 in Figure 1) within the region spanned by the probe. UGPase from potato (Borovkov et al., 1996) has also been shown to be single-copy, although allelic isoforms have been described from this plant (Sowokinos et al., 1997).

G. S. M. D.	gracilis pombe musculus discoideum	FQYTESRVLKARLRFLAEKYQIREDDFLVFDVSFVLQSFAKVQCCARRNLTVLLL 624YQYTESRVLKARLEFLRDTYQIREADFLTFD
G.	gracilis	NRHQNGQ*AMRQAAQCAGRVIRNKNDYGIVIFADKRFTRAKLRSKLPKWIAQFLSVDSLDL
S.	pombe	AMRHAAQCLGRVLRGKDDHGIMVLADKRYGRSDKRTKLPKWIQQYITEGATNL
M.	musculus	AMRHAAQCVGRAIRGKTDYGLMVFADKRFARADKRGKLPRWIQEHLTDSNLNL
D.	discoideum	AMRTASQCVGRVIRGKSDYGIMIFADKRYNRLDKRNKLPQWILQFCQPQHLNL
G.	gracilis	DLGTAIAEARTFLLDMAQPSVAKSRVEEKPPENYEFIGSRTAPE*
S.	pombe	STDMSLALAKKFLRTMAQPFTASDQEGISWWSLDDLLIHQKKALKSAAIEQSKHEDEMDID
M.	musculus	TVDEGVQVAKYFLRQMAQPFHREDQLGLSLLSLEQLQSEETLQRIEQIAQQL*
D.	discoideum	STDMAISLSKTFLREMGQPFSREEQLGKSLWSLEHVEKQSTSKPPQQQNSAINSTITTSTT
G. S. M.	gracilis pombe musculus	VVET*

D. discoideum TTTTTSTISETHLT*

Figure 3a. Alignment of the peptide fragment deduced from the portion of a putative DNA helicase gene found downstream of GgUGP (italicized portion in Figure 1) with the C-terminal portions of DNA helicase homologs from *Mus musculus* (NCBI accession number 2114484), *Dictyostelium discoideum* (RepD, 2058510), and *Schizosaccharomyces pombe* (RAD15, 5022); numbers at the start of sequences indicate positions in the original sequence of the first residue shown here. The 32-position insertion in the *Graciliaria gracilis* conceptually translated helicase sequence corresponds to a putative intron (Figure 3B and text) and as such would not actually occur in the protein. '-' indicates gaps inserted to achieve alignment; * indicates a position encoded by a stop codon. Alignment was produced using CLUSTAL W (Thompson et al. 1994).

An intron-containing gene occurs just downstream from GgUGP

We sequenced the region downstream from *GgUGP* and used the sequence to search the NCBI database (using BLASTX). The conceptual translation product of the sequence in this region has significant similarity with C-terminal portions of DNA helicases (Figure 3A), strongly suggesting that this region (italicized in Figure 1) represents the 3' portion of a potential DNA helicase gene (encoded on the complementary strand relative to GgUGP). Its stop codon is 376 bp downstream of the stop codon of GgUGP. The proximity of these two genes is yet another indication that close spacing of genes in the G. gracilis genome may not be uncommon. The proximity of genes encoding polyubiquitin and mitochondrial aconitase (Zhou & Ragan, 1995a), and of genes encoding GALT and a putative peptidyl-tRNA hydrolase (Lluisma & Ragan, 1998b), has already been observed.

A potential 96-bp, phase-0 intron was also found in the DNA helicase gene region. As indicated by the alignment (Figure 3A), the otherwise strong sequence similarity is interrupted by the presence in the conceptually translated *G. gracilis* sequence of an apparent insertion, 32 amino acids in length including an in-frame termination, that does not occur in DNA helicase proteins from *Schizosaccharomyces pombe, Dictyostelium discoideum*, or mouse. Inspection of the nucleotide sequence (Figure 3B) reveals that this apparent amino acid insertion in fact corresponds to a putative intron whose sequences at the potential 5' (AT:GTAAGT, where the colon indicates the presumptive exon-intron junction) and 3' (TAG:G) splice sites and at the potential branch site (CTAAC) conform perfectly to the canonical sequences for spliceosomal introns in red algal nuclear genes (Zhou & Ragan, 1996). Notably, this intron is positioned close to the 3' end of the gene; all other red algal genes characterized to date are devoid of introns at the 3' region.

Conclusion

From the red alga *G. gracilis* we have cloned a singlecopy nuclear gene that putatively encodes UGPase, a key enzyme of galactan biosynthesis. The characterization of this gene paves the way for more intensive studies of the biochemistry and molecular biology of UGPase and agar biosynthesis in this alga. The gene produces two transcripts that differ in length due to the presence of alternative polyadenylation sites; the possibility that this constitutes a mechanism for regulating ${\tt tttccaatacaccgaatccagagtgttaaaggctcgcctgcggtttctggcagaaaagtaccaaatacgggaggacga}$ F O Y T E S R V L K A R L R F L A E K Y O I R E D D cttccttgttttcgat**gt**aagttttgttctgcagtcttttgccaaggtacaatgctgtgctaggagaaatcta**ac**agt FLVFD AMROAAOCAGRV IRN K N D Y G I V I F A D K R F T R A K L R S K L P K M gattgcccagttcctgttgttgactccttggatctcgatctcggtaccgcaattgctgaagcacgtacttttctgct I A O F L S V D S L D L D L G T A I A E A R T F L L tgacatggcgcaaccatcggtagctaagtctcgagtggaggagaagccaccggaaaactatgaattcatcggaagccgD M A O P S V A K S R V E E K P P E N Y E F I G S R gaccgcaccagagtaa TAPE

Figure 3b. The nucleotide sequence of the 3'-portion of a putative gene encoding a DNA helicase homolog in *G. gracilis*; this sequence is complementary to the italicized portion in Figure 1 (positions 2348–1865). The conceptual translation is shown below the nucleotide sequence. The portion likely to be an intron is italicized; the dinucleotides conserved at the 5' and 3' ends (GT and AG, respectively) and at the branch site (AC) in red algal spliceosomal introns (Zhou & Ragan, 1996) are in bold and underlined.

gene expression remains to be investigated. The observation that another gene, potentially encoding a DNA helicase, is located just downstream of GgUGP adds to accumulating evidence that occurrence of closely spaced genes may not be unusual in the *G. gracilis* genome.

Acknowledgements

We thank the Marine Biology group of the NRC Institute for Marine Biosciences, especially Ms Colleen Murphy, for excellent technical assistance. AOL thanks the International Development Research Center of Canada for the PhD fellowship, and the University of the Philippines – Marine Science Institute for study leave. MAR thanks the Canadian Institute for Advanced Research, Program in Evolutionary Biology for fellowship support. Issued as NRCC no. 42277.

The DNA sequence reported herein has been deposited in GenBank under the accession number AF1007880.

References

- Borovkov AY, McClean PE, Secor GA (1997) Organization and transcription of the gene encoding potato UDP-glucose pyrophosphorylase. Gene 186: 293–297.
- Chang H-Y, Peng H-L, Chao YC, Duggleby RG (1996) The importance of conserved residues in human liver UDPglucose pyrophosphorylase. Eur. J. Biochem. 236: 723–728.
- Craigie JS (1990) Cell walls. In Cole KM, Sheath RG (eds), Biology of the Red Algae. Cambridge University Press, Cambridge. pp. 221–257.
- Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc. Natl Acad. Sci. USA 89: 10915– 10919.
- Katsube T, Kazuta Y, Tanizawa K, Fukui T (1991) Expression in *Escherichia coli* of UDP-glucose pyrophosphorylase cDNA from potato tuber and functional assessment of the five lysyl residues located that the substrate-binding site. Biochemistry 30: 8546– 8551.
- Kazuta Y, Omura Y, Tagaya M, Nakano K, Fukui T (1991) Identification of lysyl residues located at the substrate-binding site in UDP-glucose pyrophosphorylase from potato tuber: affinity labelling with uridine di- and triphosphopyridoxals. Biochemistry 30: 8541–8545.
- Kleczkowski LA (1994) Glucose activation and metabolism through UDP-glucose pyrophosphorylase in plants. Phytochemistry 37: 1507–1515.
- Konishi Y, Tanizawa K, Muroya A, Fukui T (1993) Molecular cloning, nucleotide sequencing, and affinity labelling of bovine liver UDP-glucose pyrophosphorylase. J. Biochem. 114: 61–68.
- Kremer BP (1978) Patterns of photoassimilatory products in Pacific Rhodophyceae. Can. J. Bot. 56: 1655–1659.
- Kremer BP, Kirst GO (1981) Biosynthesis of 2-O-D-glycerol-α-D-galactopyranoside (floridoside) in marine Rhodophyceae. Pl. Sci. Lett. 23: 349–357.
- Liaud M-F, Valentin C, Brandt U, Bouget F-Y, Kloareg B, Cerff R (1993) The GAPDH gene system of the red alga *Chondrus*

crispus: promotor structures, intron/exon organization, genomic complexity and differential expression of genes. Pl. Mol. Biol. 23: 981–994.

- Lluisma AO, Ragan MA (1998a) Cloning and characterization of a gene encoding starch-branching enzyme from the marine red alga *Gracilaria gracilis*. Curr. Genet. 34: 105–111.
- Lluisma AO, Ragan MA (1998b) Characterization of a galactose-1-phosphate uridylyltransferase gene from the marine red alga *Gracilaria gracilis*. Curr. Genet. 34: 112–119.
- Macler BA (1986) Regulation of carbon flow by nitrogen and light in the red alga, *Gelidium coulteri*. Pl. Physiol. 82: 136–141.
- Manley SL, Burns DJ (1991) Formation of nucleoside diphosphate monosaccharides (NDP-sugars) by the agarophyte *Pterocladia capillacea* (Rhodophyceae). J. Phycol. 27: 702–709.
- Nagashima H, Nakamura S, Nisizawa K, Hori T (1971) Enzymic synthesis of floridean starch in a red alga, *Serraticardia maxima*. Pl. Cell Physiol. 12: 243–253.
- Ragheb JA, Dottin RP (1987) Structure and sequence of a UDP glucose pyrophosphorylase gene of *Dictyostelium discoideum*. Nucl. Acids Res. 15: 3891–3906.
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A Laboratory Manual. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sowokinos JR, Thomas C, Burrel MM (1997) Pyrophosphorylases in potato V. Allelic polymorphism of UDP-glucose pyrophosphorylase in potato cultivars and its association with tuber resistance to sweetening in the cold. Pl. Physiol. 113: 511–517.
- Su J-C, Hassid WZ (1962) Carbohydrates and nucleotides in the red alga *Porphyra perforata* II. separation and identification of nucleotides. Biochemistry 1: 474–480.

- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22: 4673–4680.
- Turnquist RL, Hansen RG (1973) Uridine diphosphoryl glucose pyrophosophorylase. The Enzymes, 3rd Edn., vol 8, pp. 51–71.
- Zhou Y-H, Ragan MA (1993) cDNA cloning and characterization of the nuclear gene encoding chloroplast glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracilaria verrucosa*. Curr. Genet. 23: 483–489.
- Zhou Y-H, Ragan MA (1994) Cloning and characterization of the nuclear gene encoding plastid glyceraldehyde-3-phosphate dehydrogenase from the marine red alga *Gracilaria verrucosa*. Curr. Genet. 26: 79–86.
- Zhou Y-H, Ragan MA (1995a) Characterization of the nuclear gene encoding mitochondrial aconitase in the marine red alga *Gracilaria verrucosa*. Pl. Mol. Biol. 28: 635–646.
- Zhou Y-H, Ragan MA (1995b) Characterization of the polyubiquitin gene in the marine red alga *Gracilaria verrucosa*. Biochim. Biophys. Acta 1261: 215–222.
- Zhou Y-H, Ragan MA (1995c) Cloning and characterization of the nuclear gene and cDNAs for triosephosphate isomerase of the marine red alga *Gracilaria verrucosa*. Curr. Genet. 28: 317–323.
- Zhou Y-H, Ragan MA (1996) Nuclear-encoded protein-encoding genes of the agarophyte *Gracilaria verrucosa* (Gracilariales, Rhodophyta). In Lindstrom SC, Chapman DJ (eds), Proceedings of the XV International Seaweed Symposium, pp. 429–436.