Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4693–4697, May 1992
Cell Biology

Primary structure and expression of a gamete lytic enzyme in *Chlamydomonas reinhardtii*: Similarity of functional domains to matrix metalloproteases

(collagenase family/extracellular matrix/periplasm/preproenzyme form/zinc binding site)

TETSU KINOSHITA*, HIDEYA FUKUZAWA†, TOMOO SHIMADA*, TATSUAKI SAITO*‡, AND YOSHIHIRO MATSUDA*§

*Department of Biology, Faculty of Science, Kobe University, Nada-ku, Kobe 657, Japan; and †Laboratory of Plant Molecular Biology, Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, Kyoto 606-01, Japan

Communicated by Joseph E. Varner, February 24, 1992 (received for review August 27, 1991)

ABSTRACT **A gamete lytic enzyme (GLE) of *Chlamydomonas reinhardtii* is a zinc metalloprotease and mediates digestion of the cell walls of the two mating-type gametes during mating as a necessary prelude to cell fusion. The nucleotide sequence analysis of a cDNA revealed that GLE is synthesized in a preproenzyme form, a 638-amino acid polypeptide (M_r, 69,824) with a 28-amino acid signal peptide, a 155-amino acid propolypeptide, and a 455-amino acid mature polypeptide (M_r, 49,633). A potential site for autocatalytic activation was contained within the propolypeptide and a zinc binding site found within the mature polypeptide; both sites were highly homologous to those in mammalian collagenase. A putative calcium binding site was present in the near C-terminal region of the mature GLE. Both propolypeptide and mature polypeptide had potential sites for asparagine-linked glycosylation, and the Arg-$(Pro)_3$ and Arg-$(Pro)_2$ motifs, which are known to exist in hydroxyproline-rich glycoproteins of the *Chlamydomonas* cell wall. Northern blot analysis revealed that steady-state levels of the 2.4-kilobase GLE mRNA increased during growth and mitotic cell division in the vegetative cell cycle and also increased markedly during gametogenesis under nitrogen-starved conditions.**

The controlled remodeling and breakdown of the cell's extracellular matrix (ECM) are important in such biological processes as growth, development, fertilization, and cell fusion in both animals and plants. These events are mediated by various types of ECM-degrading enzymes. In animals, metalloproteases such as collagenases, transins, and stromelysins degrade connective tissues, and this degradation is responsible for cell migration, metastasis, uterine involution, bone resorption, and wound healing (1–3). In the developing embryo of sea urchin, a hatching enzyme (HE), a collagenase homologue, is secreted at the late blastula stage and digests the fertilization envelope to release the ciliated embryo (4). Plant cells also have a unique type of ECM, the cell wall, and its disintegration by cell wall-degrading enzymes is essential for cell expansion, pollen and seed germination, daughter cell hatching, and sexual cell fusion (5, 6).

In the unicellular biflagellated alga *Chlamydomonas reinhardtii*, the mating-type plus (mt^+) and minus (mt^-) gametes shed their cell walls during mating as a necessary prelude to cell fusion. The shedding is caused by the activity of a gamete lytic enzyme (GLE).¶ Release of GLE is induced by the signal of flagellar agglutination between gametes of the opposite mating type (7) or by the exogenous presentation of dibutyryl-cAMP to gametes (8). This enzyme is secreted into the culture medium by the cells concurrently with release of the cell walls. GLE purified from the medium after mating is characterized as a zinc-containing metalloprotease with a molecular mass of 62 kDa on SDS/PAGE (9, 10). This enzyme acts specifically on the framework proteins of the cell wall (10–12) and also cleaves several model peptides at specific sites (13). GLE is stored in the periplasm of gametes until its release (14, 15). GLE in gametic cells has also been purified from cell homogenates in an active and soluble form (G-form) and has been shown to have the same properties as those of GLE in the mating medium (14). Furthermore, GLE is actually present in vegetative cells in an insoluble and inactive form (V-form) (7, 14). The V-form GLE is activated *in vitro* by sonicating the vegetative cell homogenates, and the activated enzyme has similar properties to those of GLE secreted into the medium (14). As vegetative cells differentiate into gametes under nitrogen-starved conditions, the V-form enzyme may convert to the G-form (14, 16, 17). The activity of the G-form GLE found in the gametic cell homogenates, however, is consistently lower than that found in the mating medium (14). It has been reported recently that GLE is also stored in gametic cells as an inactive, higher molecular mass precursor, which is activated during sexual signaling (15, 18).

To investigate the molecular details of synthesis, storage, and activation of GLE, we have isolated cDNA clones for GLE from a cDNA library of RNA from vegetative cells. In this article, we report the isolation and characterization of a full-length cDNA coding for this enzyme, show that GLE is synthesized in a preproenzyme form, discuss the predicted structure of the protein, and analyze the temporal pattern of GLE gene expression in the cell cycle and gametogenesis.|| Finally, GLE shows molecular similarities with ECM-degrading metalloproteases from animal sources, notably in the amino acid sequences that function in zinc binding and in autocatalytic activation.

MATERIALS AND METHODS

Cells and Culture Conditions. The SAG 11-32b (mt^+) strain was used for synchronous vegetative culture under a 12-hr light/12-hr dark regime (17). For preparation of gametes, synchronously grown vegetative cells at the beginning of the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ECM, extracellular matrix; G-form, gametic form; HE, hatching enzyme; HRGP, hydroxyproline-rich glycoprotein; GLE, gamete lytic enzyme; IPTG, isopropyl β-D-thiogalactopyranoside; V-form, vegetative form; ORF, open reading frame.

‡Present address: Department of Biology, Washington University, St. Louis, MO 63130.
§To whom reprint requests should be addressed.

¶Although this enzyme has been called different names in different laboratories (lytic enzyme, autolysin, gamete wall autolysin, lysin, and g-lysin), we propose the name gamete lytic enzyme (GLE).
||The sequence reported in this paper has been deposited in the GenBank data base (accession no. D90503).

light period (L-0 cells) were transferred to nitrogen-free medium and incubated for ≈8 hr (17). The mt^- (137c strain) gametes were obtained from plate cultures (14).

N-Terminal Amino Acid Sequencing of GLE. GLE was purified from the medium of mating gametes by a published procedure (13). Purified GLE was subjected to SDS/PAGE (14), electrotransferred to poly(vinylidene difluoride) membranes (Millipore), and visualized with Coomassie brilliant blue staining. The N-terminal amino acid sequence of the blotted GLE was determined with a peptide sequencer (model 477A/120A; Applied Biosystems).

Screening and Nucleotide Sequencing Analysis of cDNA Clones. Total RNA was extracted from synchronously grown vegetative cells at 6 hr into the dark period by using guanidium isothiocyanate and subsequent ultracentrifugation, and poly(A)$^+$ RNA was selected as described (19). cDNA was synthesized by using an Amersham kit and cloned into λgt10. A 57-mer oligonucleotide probe (see *Results*) was synthesized with a DNA synthesizer (model 380B; Applied Biosystems) and end labeled with [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase. Plaque hybridization was carried out with Hybond N$^+$ nylon membranes (Amersham) as described (19). The cDNA inserts from selected λgt10 phages were subcloned into M13mp18 and M13mp19. The nucleotide sequence was determined by using a 7-deaza sequencing kit (Takara, Kyoto).

Construction of Fusion Protein and Immunoblotting. A 1847-base-pair (bp) *Pvu* II/*Eco*RI portion of the gtGLE4 cDNA (see *Results*) was cloned into the expression vector pGEX-2T (Pharmacia), which had been predigested with *Sma* I and *Eco*RI. The recombinant plasmid was maintained in *Escherichia coli* JM101. Cells were grown for 8 hr with ampicillin (50 μg/ml) and 1 mM isopropyl β-D-thiogalactopyranoside (IPTG), pelleted, dissolved in sample buffer for SDS/PAGE (9), and loaded on 13% gels. After gel electrophoresis, proteins were electrotransferred to Hybond ECL membranes (Amersham) and reacted with an antibody raised against GLE. This antibody was obtained by injecting the GLE glycoprotein band at 62 kDa on acrylamide gels (9) into a rabbit and was purified by protein A affinity chromatography. The binding of antibody to the blotted proteins was detected by the ECL Western blotting detection system (Amersham).

Northern Blot Hybridizations. Total RNA was fractionated on 1% formaldehyde/agarose gels and blotted onto nylon membranes. The cloned cDNA fragments were radiolabeled with [α-^{32}P]dCTP (Amersham) using random oligonucleotide primers (19). Prehybridization and hybridization with each cDNA probe were carried out as described (20). An RNA ladder (BRL) was used for size markers.

RESULTS

Cloning and Characterization of GLE cDNA. The mature GLE isolated from the medium of mating gametes contained the 20 N-terminal amino acid sequence Glu-Ile-Tyr-Ala-Gly-Lys-Pro-Ile-Asp-Leu-Arg-Thr-Ile-Val-Tyr-Ile-Met-Asp-Phe-Ser. Since the codon usage of *C. reinhardtii* nuclear genes is highly biased (21), we took advantage of the most probable codons to design a single long oligonucleotide probe, based on the N-terminal amino acid sequence. This 57-mer oligonucleotide probe 5′-GAAGTCCATGATGTAIACGATGGTGCG-CAGGTCGATGGGCTTGCCGGCGTAGATCTC-3′ (I at position 16 is inosine) was complementary to a putative mRNA sequence encoding the first 19 N-terminal amino acid residues of the mature GLE. The 20th amino acid (serine) was not used for design of the probe since its codon usage is relatively broad (21). We also constructed a λgt10 cDNA library with poly(A)$^+$ RNA from the synchronously grown vegetative cells at 6 hr into the dark period since the V-form GLE increases during the early part of the dark period (Y.M., Y. Ono, M. Koseki, and T. Saito, unpublished data). About 1×10^6 recombinants were screened with the oligonucleotide probe, and 6 independent cDNA clones were obtained. These clones had 1.8- to 2.4-kilobase (kb) inserts and had identical nucleotide sequences for ≈200 bases from the 3′ ends of the inserts, except for the length of poly(A). The cDNA clone having the longest insert, designated gtGLE4, was used for further analysis.

To verify whether the gtGLE4 actually encodes the GLE polypeptide, the nucleotide sequence of the 2389-bp insert was determined and confirmed on both strands. The nucleotide sequence and the deduced amino acid sequence are shown in Fig. 1. The cDNA insert in gtGLE4 consisted of a 31-bp 5′ untranslated region, a 1914-bp coding region (an open reading frame; ORF), a 434-bp 3′ untranslated region, and a 10-bp poly(A) tail (Fig. 1). A conserved poly(A) signal sequence, TGTAA, was found 13 nucleotides upstream from the poly(A) site. The ORF could encode a protein of 638 amino acids with a derived molecular weight of 69,829. A nucleotide stretch that was complementary to the 57-mer synthetic oligonucleotide probe was localized within the ORF at positions 550–606, wherein 51 nucleotides matched. The deduced amino acid sequence (Fig. 1) had an identical sequence at positions 184–203 with the arrangement of the first 20 amino acids of the mature GLE, including the 20th amino acid (serine). Furthermore, a 434-amino acid portion of the mature polypeptide (positions 205–638), which was synthesized by the expression vector pGEX-2T as a fusion protein, was immunoreactive with anti-GLE antibody (Fig. 2). Thus, we are convinced that gtGLE4 encodes the GLE polypeptide.

We assumed that the first ATG in gtGLE4, which was numbered as +1 in Fig. 1, would be the initiation codon for translation. This conclusion was based on the following reasons: (*i*) It is the first in-frame ATG from the 5′ end. (*ii*) The length of the gtGLE4 insert is in good agreement with the size (2.4 kb) of GLE mRNA (see below). (*iii*) The nucleotide sequence ACGCC is found immediately upstream of the first ATG codon. This sequence matches the consensus sequence CC(A/G)CC, which is conserved in the upstream position of translation in eukaryotes (22).

Amino Acid Sequence Analysis of GLE Polypeptide. After the first methionine, there is an apparent signal sequence (23) of 28 amino acids as expected for a secreted protein, which comprises a short, basic amino acid region, followed by a hydrophobic core containing 14 hydrophobic amino acid residues such as alanine and leucine. Since the N-terminal residue in the mature GLE (glutamic acid) is localized at position 184 in the putative precursor polypeptide (Fig. 1), we assumed that GLE is synthesized in a preproenzyme form with a 28-amino acid signal peptide, a 155-amino acid propolypeptide, and a 455-amino acid mature polypeptide. The proGLE, after the removal of the signal peptide, has a predicted molecular weight of 67,035, and the mature enzyme, after removal of the propolypeptide, has a molecular weight of 49,633.

The proGLE comprised eight potential asparagine-linked glycosylation sites, Asn-Xaa-Thr/Ser (24); three of them were localized in the propolypeptide domain, and five were localized in the mature polypeptide domain (Fig. 1). In addition, there were two proline-rich regions, one within the propolypeptide (positions 48–62) and the other within the mature polypeptide (positions 270–283). These regions contained Arg-(Pro)$_3$ and Arg-(Pro)$_2$ sequences, which are hallmarks of hydroxyproline-rich glycoproteins (HRGPs) of the *Chlamydomonas* and *Volvox* cell walls (25–27).

The deduced amino acid sequence of GLE was compared with those of other previously reported proteins filed in a NBRF protein data base. This search revealed that two blocks of amino acid sequence in GLE were highly homologous to those of proteins from the collagenase family (1–3). First, a 19-amino acid block at positions 388–406 in the GLE

```
 -31                                                                                      TTCCAGGTTCGGTGCCCCGCGCGAGAACGCC    -1

  +1  ATGTCGCTGGCGACGCGGCGCTTCGGCGCCGCAGCGGCACTTCTAGTCGCCGCATGCGTGCTATGCACAGCTCCTGCGTGGGCTCAGAATGAGACTACTGGCACCGGCATGGTCAAGACG   120
   1   M  S  L  A  T  R  R  F  G  A  A  A  A  L  L  V  A  A  C  V  L  C  T  A  P  A  W  A ^Q [N] E  T  T  G  T  G  M  V  K  T    40

 121  AAGAGCGCCTTCCGTTGGATTCGCCCACCTCCCGCCCGCCCGCCTCCCTTCCGCCGGCCTCCGCCTGCGCAGACCCCATATGTGCATAAGGTCGAGTACACAGAGCTGCAGATCCTGTGC   240
  41   K  S  A  F  R  W  I  R  P  P  P  A  R  P  P  P  F  R  R  P  P  P  A  Q  T  P  Y  V  H  K  V  E  Y  T  E  L  Q  I  L  C    80

 241  CCCCAGACCATCGACTCGGTGACTGGGTACCCCATGGACGACCCGAGATGCAACGTGCCCCGTGCCACCGTTGCTGCTGGCGAGGAGGCGCTGACCATTCGTAACGAGTTCGAGCTACTG   360
  81   P  Q  T  I  D  S  V  T  G  Y  P  M  D  D  P  R  C  N  V  P  R  A  T  V  A  A  G  E  E  A  L  T  I  R  N  E  F  E  L  L   120

 361  AATGGTGACGTGCTCAACGTGACCCTTGAGGAGGTGGACACGCCCGAGAACCCCAGCCGCCGCCGCCTTCTGTCCATCATTCGCGAGGAGCAGCGCACGGGCCGCGTCCTGCTGGCCACC   480
 121   N  G  D  V  L [N] V  T  L  E  E  V  D  T  P  E [N] P  S  R  R  R  L  L  S  I  I  R  E  E  Q  R  T  G  R  V  L  L  A  T   160

 481  AGCGCGGAGCTCCCGACCCCCACGTTCAGGTTGAAGAGCCTGAAGAGCATCCTTAAGGGTTCCCAGAAGGAGATTTACGCCGGCAAGCCCATTGACCTGCGCACGATTGTGTACATCATG   600
 161   S  A  E  L  P  T  P  T  F  R  L  K  S  L  K  S  I  L  K  G  S  Q  K ^E  I  Y  A  G  K  P  I  D  L  R  T  I  V  Y  I  M   200

 601  GACTTTAGCAGCTGCAAGCTGTCCGGCTGGTCGGCCCCCGCCACCCTGACCCCCGAGAAAGTCACGTCGGACATGCTGCGCGGCGCCAGTGCCCCCACCAATAACCTGGCTAACTACTAT   720
 201   D  F  S  S  C  K  L  S  G  W  S  A  P  A  T  L  T  P  E  K  V  T  S  D  M  L  R  G  A  S  A  P  T  N  N  L  A  N  Y  Y   240

 721  GGGGCCTGCTCGTACGAGAAGACCCTGTTCAATCCCGACAACTTCTTGGTGCTGGGCCCGGTGCCCGTGCCCTGCATTGGTGGCGTCACGCCGCCGCCCCGTCCTCCGCGCCCGCCGCGG   840
 241   G  A  C  S  Y  E  K  T  L  F  N  P  D  N  F  L  V  L  G  P  V  P  V  P  C  I  G  G  V  T  P  P  P  R  P  P  R  P  P  R   280

 841  CCGCCTCCGCGCGCCGGCTCCACCATTTCGTCCCTGTCGCGCCGTAACGACACCTACGACGACTGGTGGGATCTGTCCAAGTACTGCACTGCCTCGGAGCAGCAGGCTTGGGAGCGCGCC   960
 281   P  P  P  R  A  G  S  T  I  S  S  L  S  R  R [N] D  T  Y  D  D  W  W  D  L  S  K  Y  C  T  A  S  E  Q  Q  A  W  E  R  A   320

 961  GCTGAGGCCTACGCCCAGGCGATTGTCGCTCAGGACCCCAACAGCGCGACCGGCAAGAAGCTCCAGGGCATCCTGCAGTGGAGGGAGCGCCGGCGGAACATCTACATCCTGCCGCCCGGC  1080
 321   A  E  A  Y  A  Q  A  I  V  A  Q  D  P  N  S  A  T  G  K  K  L  Q  G  I  L  Q  W  R  E  R  R  R  N  I  Y  I  L  P  P  G   360

1081  GTCAAGTGCTCCTGGTCGGGCTACGCCGACGTGACCTGCACGTCGGCCACCTGCAGCGCCTACGTGCGCGGCTACTCGGACACCAACGCCATGCAGGTCATCATGCACGAGGCGATGCAC  1200
 361   V  K  C  S  W  S  G  Y  A  D  V  T  C  T  S  A  T  C  S  A  Y  V  R  G  Y  S  D  T  N  A  M  Q  V  I  M  H  E  A  M  H   400

1201  AACTACGGCCTCGAGCACGCCGGCCGCGGCACACTGGAGTACGGTGACGCCACCGACGTCATGGGCGACTTCAACAAGGCCGGCAAGGGCCTGCTGTGCCCCAACGCCCCCAACATGTAC  1320
 401   N  Y  G  L  E  H  A  G  R  G  T  L  E  Y  G  D  A  T  D  V  M  G  D  F  N  K  A  G  K  G  L  L  C  P  N  A  P  N  M  Y   440

1321  CGCATCGGCTGGGCCAAGCCCATCAACGAGCCCGGCGTCGCGCCCTTCCAGAATGCGACCGGCGCGTGGGGCAACCTCACAGCCGCCAACTTCACCACCGACCCCTGGATCAGGGGCCTG  1440
 441   R  I  G  W  A  K  P  I  N  E  P  G  V  A  P  F  Q [N] A  T  G  A  W  G [N] L  T  A  A [N] F  T  T  D  P  W  I  R  G  L   480

1441  GTCATCCCCGCCCAGGGCACCCGCGATGACAACATGATCGTGGTCAACGTGGGCGCTCAGAGCACCCGGGACGGGGCAATGAAGGCTACCGGCGCTCAGGCCTACTACTTCTCCTACCGC  1560
 481   V  I  P  A  Q  G  T  R  D  D  N  M  I  V  V  N  V  G  A  Q  S  T  R  D  G  A  M  K  A  T  G  A  Q  A  Y  Y  F  S  Y  R   520

1561  ATTAAGAACACCACCGCTGGCGGCTACGACTCTGGCCTGACCTTAGACTTCCATAAGAAGGTGCTGGTCCACGCCTACAACGGCATCCAGTCGGAGCGCGTGTTTGGCTTCAAGTCCAAC  1680
 521   I  K [N] T  T  A  G  G  Y  D  S  G  L  T  L  D  F  H  K  K  V  L  V  H  A  Y  N  G  I  Q  S  E  R  V  F  G  F  K  S  N   560

1681  CTGTTGGACTGGGGCCCGAACTTCCAATCCAGGAGCAACACCTGGACTTCACCCTTCCTGGCGTATAACAACGGTCTTGGTGGTGGCGTGAGGCTGGTGGTGCAGAGCACTTCTGACACG  1800
 561   L  L  D  W  G  P  N  F  Q  S  R  S  N  T  W  T  S  P  F  L  A  Y  N  N  G  L  G  G  G  V  R  L  V  V  Q  S  T  S  D  T   600

1801  CAGGCCGTCGTGGACATCTGCCGCATCAGTGAGAACGGCAAGGAGCTGTCGTGCGACGATGGTATTGACAACGACTGTGACGGCTTGCAGGATAATGAGGATCCCGACTGCCAGTAAAGC  1920
 601   Q  A  V  V  D  I  C  R  I  S  E  N  G  K  E  L  S  C  D  D  G  I  D  N  D  C  D  G  L  Q  D  N  E  D  P  D  C  Q  *   638

1921  CGTGCCCCAGACTCCCTGACTAGTGCGCATGCTGCTGCAGTAGTGTCTAGTTCATAAGACTAGTGGTTTTTGATATGGTCGTCATGCATTTCCTGTGCATGAGGGAGATGCAATTGCTTT  2040
2041  GTGCAAGTCAACCCTATGTCTGATACAACTGACGGCGGTCCTGTGCAAGCTGCGGTTGGCGTGACTGTTGTGTAGAGGTGGCCGGGTCCATCCAACTCACGCCCATGGGTGGTCAGTTGT  2160
2161  GACTGTTGCGGTGTTGCTGTTGATTTTGGTACATAGACTAAACAGGCTGCAAGGTCTTTCGTTCGTCTGGCTCGTAATATGGCGCGAACGCCCTTCAGCAGGGATTGCCTGTGTTTCTTA  2280
2281  TTCATTTGTGATCCGTCCATGCGCAACTGGCGGGATTTCCCTTGATATGGTTGTAACACCTATGGAACAAAAAAAAAA                                          2358
```

FIG. 1. Nucleotide sequence of 2389-bp cDNA fragment in gtGLE4 and deduced amino acid sequence of *Chlamydomonas* GLE. Sequences of *Eco*RI linkers at both 5′ and 3′ ends are not included. Amino acid residues that are identical to the N-terminal peptide sequence of mature GLE are underlined. Junction between the signal peptide and the propolypeptide and that between the propolypeptide and the mature polypeptide are indicated by arrowheads. A conserved poly(A) signal sequence (TGTAA) is marked with a thick bar and a translation termination codon (TAA) is marked with a star. Potentially glycosylated asparagine residues are boxed.

mature polypeptide sequence (Fig. 3*A*) contains the His-Glu-Xaa-Xaa-His motif that corresponds to the zinc binding residues at the active site of a number of metalloproteases such as collagenase, stromelysin, gelatinase, HE from sea urchin, and thermolysin (1, 4, 28). Amino acid residues encompassing the motif are also highly conserved in all metalloproteases (Fig. 3*A*), which constitute the "zinc signature" (29). Second, an 18-amino acid block at positions 85–102 exhibits a high degree of homology (Fig. 3*B*). Especially, the Pro-Arg-Cys-Asn-Val-Pro motif at positions 95–100 is highly homologous to the conserved sequence Pro-Arg-Cys-Gly-Val-Pro in the mammalian collagenase family, which has been recently demonstrated to be critical in the autocatalytic conversion of the zymogens into active enzymes (30, 31). Finally, we found a 13-amino acid acidic motif

1 2

73 kDa

FIG. 2. Synthesis of a fusion protein immunoreactive with anti-GLE antibody. Cells of *E. coli* bearing recombinant plasmids were cultured for 8 hr in the absence (lane 1) or presence (lane 2) of 1 mM IPTG. Eight-microgram aliquots of total cell proteins were fractionated on SDS/polyacrylamide gels, electroblotted onto nitrocellulose membranes, and reacted with an antibody raised against the mature GLE. A strongly immunoreactive band at 73 kDa, which is expressed only by adding IPTG (lane 2), is approximately the sum of the molecular mass of glutathione *S*-transferase (25 kDa) and the 434-amino acid portion (47 kDa) of mature GLE.

```
A                                         *       *
GLE (Chlamydomonas)    T N A M Q V I M H E A M H N Y G L E H   406
Gelatinase (human)     Y S L F L V A A H E F G H A M G L E H   384
Stromelysin (human)    T N L F L V A A H E I G H S L G L F H   228
Collagenase (human)    T N L H R V A A H E L G H S L G L S H   229
HE (sea urchin)        T N L F Q V A A H E F G H S L G L Y H   293

B                                             +
GLE (Chlamydomonas)    D S V T G Y P M D D P R C N V P R A   102
Gelatinase (human)     D Q N T I E T M R K P R C G N P D A    79
Stromelysin (human)    D S D T L E V M R K P R C G V P D V    97
Collagenase (human)    D A E T L K V M K Q P R C G V P D V    97
HE (sea urchin)        D A D T A E L L S T P R C G V P D V   164

C
GLE (Chlamydomonas)    I D N D C D G L Q D N E D   634
TS (human)             Q D I D E D G H Q N N L D   854
TS (human)             H D D D N D G I P D D K D   890
Calmodulin 1 (rat)     F D K D G D G T I T T K E    31
Calmodulin 2 (rat)     V D A D G N G T I D F P E    67
Calmodulin 3 (rat)     F D K D G N G Y I S A A E   104
Calmodulin 4 (rat)     A N I D G D G Q V N Y E E   140
Parvalbumin (carp)     I D Q D K S G F I E E D E    62
Parvalbumin (carp)     G D S D G D G K I G V D E   101
   (Consensus)         ( O   O   O G     O     O)
```

FIG. 3. Comparison of the amino acid sequence of GLE with the collagenases and calcium binding proteins. Residues that are identical with those of *Chlamydomonas* GLE are boxed. (*A*) Sequence conservation of the putative zinc binding catalytic sites between GLE and members of the collagenase family. Asterisks denote the zinc binding ligands (two histidines). (*B*) Sequence homology of the putative autolytic activation sites in the propolypeptide domains of GLE and members of the collagenase family. The cysteine residue that binds to zinc is indicated by a plus sign. (*C*) Homology of the C-terminal acidic stretch of GLE to sequences of the calcium binding sites of calcium binding proteins. The consensus sequence is indicated in parentheses. The letter O in parentheses is an amino acid such as aspartic acid that contributes oxygen for calcium binding, and the letter G is a glycine residue. TS, thrombospondin 3A.

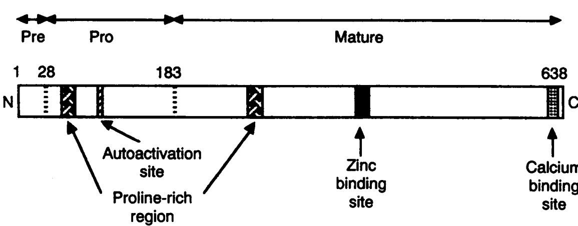

FIG. 4. Schematic representation of the preproenzyme structure of GLE. Characteristic regions are shown by marked areas. Numbers indicate positions of amino acid residues from the N terminus.

in the C-terminal region (positions 622–634) of the mature polypeptide (Fig. 3*C*), which resembles the calcium binding sites and thus an exposed region of β-turns found in several calcium binding proteins such as thrombospondin, calmodulin, and parvalbumin (32–35). The putative calcium binding sites of GLE polypeptide have the consensus amino acid alignments Asp-623, Asp-625, Asp-627, Gly-628, Asp-631, and Asp-634; the five aspartic acids may contribute oxygens for binding calcium and the glycine residue is a conserved amino acid in all of the calcium binding proteins (35).

The domain organization of GLE insert including the characteristic features described above is shown in Fig. 4.

Northern Blot Analysis of GLE Expression During Vegetative Cell Cycle and Gametogenesis. Since GLE exists in an inactive form in vegetative cells (14), we determined the steady-state levels of GLE mRNA during the vegetative cell cycle by Northern blot hybridization. As reported previously (17), cells in synchronous culture grow during the 12-hr light period, undergo mitosis and cytokinesis between 14 and 19 hr in the dark, and liberate daughter cells after ≈20 hr. Total RNA was isolated from cells collected every 2 hr throughout the 24-hr light/dark period, fractionated on denaturing agarose gels, and transferred to nylon membranes. The filters were hybridized with radiolabeled gtGLE4 insert. The results (Fig. 5*A*) revealed that the levels of GLE mRNA, which has a size of 2.4 kb, increase markedly between 6 and 12 hr, then decrease slightly, and increase again between 14 and 18 hr. The observed increase in the GLE mRNA level correlated with an extensive increase in cell size and volume during the cell cycle.

To examine changes in the steady-state levels of GLE mRNA during gametic differentiation, the synchronously grown vegetative cells at 0 hr in the light period (L-0 cells) were transferred to nitrogen-free medium and collected after 2, 4, 6, and 8 hr of incubation. As described (17), L-0 cells undergo gametic differentiation between 4 and 7 hr in nitrogen-free medium, during which neither growth nor cell division occurs. Since the amount of total RNA dropped by 50% after 2 hr in nitrogen-free medium because of degradation of rRNA (36), the Northern blot hybridization was performed

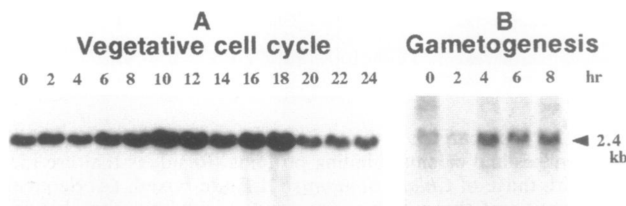

FIG. 5. Northern blot analysis of GLE transcript during vegetative cell cycle (*A*) and gametogenesis (*B*). (*A*) Total RNAs were extracted from 1–1.5 × 10^9 cells taken at the indicated stages of the vegetative cell cycle, and 30 μg of each sample was fractionated on denaturing agarose gels. After RNAs were blotted to nylon membranes, the filters were hybridized with the ^{32}P-labeled cDNA insert from gtGLE4. (*B*) Northern blot analysis was carried out by loading RNA from equal numbers of cells (0 hr, 30 μg; 2, 4, and 6 hr, 18 μg; 8 hr, 15 μg).

by loading total RNA from equal numbers of cells. The results (Fig. 5*B*) showed that the GLE mRNA is accumulated abundantly after 2 hr of induction.

DISCUSSION

The sequence data presented here provide structural and functional information that is concordant with previous biochemical and topographic data (9, 10, 13–18) on GLE as a cell wall-degrading metalloprotease of *C. reinhardtii*. Furthermore, analyses of conserved sequences reveal that GLE in a lower plant species shares several important features with the ECM-degrading animal metalloproteases, such as collagenase, stromelysin, transin, and sea urchin HE (1–4, 29–31). The common features are (*i*) the synthesis of polypeptides as an inactive preproenzyme with subsequent activation *in situ*, (*ii*) the conserved sequence of the autolytic activation site in their propolypeptide regions, and (*iii*) the conserved sequence of the putative zinc binding active site in their mature polypeptides.

The predicted primary structure of GLE is composed of three domains (Fig. 4): a 28-amino acid signal peptide that presumably targets the premature enzyme to the periplasm (14) via the endoplasmic reticulum, a 155-amino acid propolypeptide that may be responsible for repressing the activity of GLE, and a 455-amino acid mature polypeptide that is the active form of GLE. The mature enzyme without carbohydrate has a predicted mass of 50 kDa, the value being lower by ≈12 kDa than that of the secreted GLE glycoprotein as determined by SDS/PAGE (9) [five potential asparagine-linked glycosylation sites are found in the mature polypeptide (Fig. 1), and there may be O-linked sugars as well]. The putative autocatalytic activation motif (Pro-Arg-Cys-Asn-Val-Pro; positions 95–100) lies within the propolypeptide, and the putative zinc binding motif (His-Glu-Ala-Met-His-Asn-Tyr-Gly-Leu-Glu-His; positions 396–406) is localized within the mature polypeptide. By analogy with collagenase family members (31), the cysteine residue at position 97 in the propolypeptide and zinc bound to the active site may form a complex, and the dissociation of that cysteine from the zinc atom may result in cleavage of the propolypeptide and conversion of latent proenzyme to active enzyme. It should be noted, however, that the putative sequence for autocatalytic activation is localized at nearly the center of propolypeptide in GLE, whereas it lies just upstream of the N termini of mature polypeptides in the collagenases. Perhaps, therefore, the presence of 81 amino acids between the putative autocatalytic activation site and Glu-184 (the N-terminal residue of mature GLE) accounts for the presence of an inactive precursor in gametes, which has a higher relative molecular mass than the active GLE by ≈2 kDa (15). A different mechanism for GLE activation has been proposed by Snell *et al.* (18).

We previously described *in vitro* activation of proGLE stored in vegetative cells (V-form enzyme) by a prior freeze/thaw treatment of the cells or sonication of the homogenates (14). The GLE activated *in vitro* has the identical molecular mass on SDS/PAGE (14) and N-terminal amino acid sequence (Y.M., Y. Ono, M. Koseki, and T. Saito, unpublished data) as the active GLE found in the mating medium or gametic cell homogenates (G-form enzyme). Such treatments may instigate conformational changes in the proGLE leading to activation. In human fibroblast collagenase, the latent enzyme can be activated by a variety of seemingly disparate means, and these treatments are thought to modify the conformation of latent enzyme leading to the dissociation of a cysteine–zinc complex, appearance of activity, and self-cleavage of the propolypeptide region (31).

A unique feature of the *Chlamydomonas* GLE is the presence of two proline-rich sequences (Fig. 4), which are very similar to those of HRGPs in the *Chlamydomonas* cell

wall (25, 27) and higher plant extensins (37). GLE is stored in the periplasm in both vegetative cells and gametes, and the inactive, V-form enzyme is found in the insoluble fraction of vegetative cell homogenates (14). Furthermore, the activated enzyme acts on only the inner wall (framework) proteins of the cell wall (10, 12). We speculate, therefore, that the proline-rich sequences present in both propolypeptide and mature polypeptide domains are related to the association of the proenzyme with the fibrous HRGP network of the cell wall, thereby allowing the enzyme to attack directly the cell wall proteins upon activation. The presence of a putative calcium binding site that is located very close to the C-terminal end (Fig. 4) may be related to triggering the activation of proenzyme and/or to association of the proenzyme with periplasmic components.

We have also demonstrated an increase in the level of GLE mRNA during the vegetative cell cycle and gametic differentiation by Northern blot analysis. When cells are cultured synchronously under a light/dark regime, the level of GLE mRNA (Fig. 5*A*) increases between 6 and 12 hr, during which each single cell enlarges severalfold, and between 14 and 18 hr, during which mitotic cell divisions occur and four to eight daughter cells are produced. The RNA increase is almost parallel with the increase of V-form enzyme in the synchronous culture (Y.M., Y. Ono, M. Koseki, and T. Saito, unpublished data). The significance of the presence of inactive GLE during the vegetative cell cycle is not clear at present. However, since the time of increase in the mRNA and protein levels for GLE is correlated with the time of an extensive expansion of the single cell wall during growth stage and the mother cell wall during cell division stage, some of the V-form enzyme might be activated temporally and used for cell wall loosening.

Our previous studies (14, 16, 17) suggest that the inactive, insoluble V-form enzyme converts to the active, soluble G-form enzyme during gametic differentiation under nitrogen-starved conditions. This conversion never occurs when temperature-sensitive mutants for gametic differentiation are starved for nitrogen at the restrictive temperature (38). It has also been reported that the activity level of GLE found in gametes is regularly $\approx$20% of that found in the mating medium (14), suggesting the existence of inhibitors or inactive precursors for GLE in gametic cells. Indeed, a recent study by Buchanan *et al.* (15) demonstrates the existence of an inactive, soluble precursor of GLE in the periplasm of gametes. Furthermore, immunoblotting analysis (39) shows the accumulation of this precursor polypeptide during gametogenesis. The present study also shows a dramatic increase in the level of GLE mRNA during gametic differentiation (Fig. 5*B*).

Taken together, these observations suggest the following. It appears that vegetative cells synthesize proGLE and store it in a locale that requires sonication for release (V-form) and that during gametogenesis this is shifted to a locale that allows release by homogenization alone (G-form). In both cases, homogenization effects at least a partial proGLE $\rightarrow$ GLE conversion; whether this occurs by native or by artificial means is not yet known. In addition, gametogenesis induces the enhanced expression of the GLE gene(s), and the newly synthesized proGLE is stored in the same locale as the G-form, both poised to be activated and released in response to the cAMP signal.

We thank Drs. U. W. Goodenough, P. J. Ferris, and J. P. Woessner for critical reading of this manuscript and for invaluable suggestions. This work was supported in part by a research grant from the Ministry of Education, Science and Culture of Japan, from the Naito Foundation, and from the Kobe University Research Networks to Y.M.

1. Goldberg, G. I., Wilhelm, S. M., Kronberger, A., Bauer, E. A., Grant, G. A. & Eisen, A. Z. (1986) *J. Biol. Chem.* **261**, 6600–6605.
2. Wilhelm, S. M., Collier, I. E., Kronberger, A., Eisen, A. Z., Marmer, B. L., Grant, G. A., Bauer, E. A. & Goldberg, G. I. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6725–6729.
3. Collier, I. E., Wilhelm, S. M., Eisen, A. Z., Marmer, B. L., Grant, G. A., Seltzer, J. L., Kronberger, A., He, C., Bauer, E. A. & Goldberg, G. I. (1988) *J. Biol. Chem.* **263**, 6579–6587.
4. Lepage, T. & Gache, C. (1990) *EMBO J.* **9**, 3003–3012.
5. Schlösser, U. W. (1981) in *Encyclopedia of Plant Physiology New Series*, eds. Tanner, W. & Loewus, F. A. (Springer, Berlin), Vol. 13, pp. 333–351.
6. Taiz, L. (1984) *Annu. Rev. Plant Physiol.* **35**, 585–657.
7. Claes, H. (1971) *Arch. Mikrobiol.* **78**, 180–188.
8. Pasquale, S. M. & Goodenough, U. W. (1987) *J. Cell Biol.* **105**, 2279–2292.
9. Matsuda, Y., Yamasaki, A., Saito, T. & Yamaguchi, T. (1984) *FEBS Lett.* **166**, 293–297.
10. Matsuda, Y., Saito, T., Yamaguchi, T. & Kawase, H. (1985) *J. Biol. Chem.* **260**, 6373–6377.
11. Goodenough, U. W. & Heuser, J. E. (1985) *J. Cell Biol.* **101**, 1550–1568.
12. Imam, S. H. & Snell, W. J. (1988) *J. Cell Biol.* **106**, 2211–2221.
13. Matsuda, Y., Uzaki, T., Iwasawa, N., Tanaka, T. & Saito, T. (1990) *Plant Cell Physiol.* **31**, 717–720.
14. Matsuda, Y., Saito, T., Yamaguchi, T., Koseki, M. & Hayashi, K. (1987) *J. Cell Biol.* **104**, 321–329.
15. Buchanan, M. J., Imam, S. H., Eskue, W. A. & Snell, W. J. (1989) *J. Cell Biol.* **108**, 199–207.
16. Saito, T., Tsubo, Y. & Matsuda, Y. (1988) *Curr. Genet.* **14**, 59–63.
17. Matsuda, Y., Saito, T., Koseki, M. & Shimada, T. (1990) *Plant Physiol. (Life Sci. Adv.)* **9**, 1–6.
18. Snell, W. J., Eskue, W. A. & Buchanan, M. J. (1989) *J. Cell Biol.* **109**, 1689–1694.
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
20. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
21. Rochaix, J.-D. (1987) *FEMS Microbiol. Rev.* **46**, 13–34.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
23. von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105.
24. Hubbard, S. C. & Ivatt, R. J. (1981) *Annu. Rev. Biochem.* **50**, 555–583.
25. Woessner, J. P. & Goodenough, U. W. (1989) *Plant Cell* **1**, 901–911.
26. Ertl, H., Mengele, R., Wenzl, S., Engel, J. & Sumper, M. (1989) *J. Cell Biol.* **109**, 3493–3501.
27. Adair, W. S. & Apt, K. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7355–7359.
28. Matthews, B. W., Weaver, L. H. & Kester, W. R. (1974) *J. Biol. Chem.* **249**, 8030–8044.
29. Jongeneel, C. V., Bouvier, J. & Bairoch, A. (1989) *FEBS Lett.* **242**, 211–214.
30. Sanchez-Lopez, R., Nicholson, R., Gesnel, M.-C., Matrisian, L. M. & Breathnach, R. (1988) *J. Biol. Chem.* **263**, 11892–11899.
31. Springman, E. B., Angleton, E. L., Birkedal-Hansen, H. & Van Wart, H. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 364–368.
32. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251–276.
33. Kretsinger, R. H. & Nockolds, C. E. (1973) *J. Biol. Chem.* **248**, 3313–3326.
34. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. J. (1985) *Nature (London)* **315**, 37–40.
35. Lawler, J. & Hynes, R. O. (1986) *J. Cell Biol.* **103**, 1635–1648.
36. Siersma, P. W. & Chiang, K.-S. (1971) *J. Mol. Biol.* **58**, 167–185.
37. Varner, J. E. & Lin, L.-S. (1989) *Cell* **56**, 231–239.
38. Saito, T. & Matsuda, Y. (1991) *Curr. Genet.* **19**, 65–71.
39. Waffenschmidt, S., Kuhne, W. & Jaenicke, L. (1989) *Bot. Acta* **102**, 73–79.