Gene Expression Induced by Copper Stress in the Diatom
*Thalassiosira pseudonana*

Aubrey K. Davis, Mark Hildebrand, and Brian Palenik*

**Marine Biology Research Division, Scripps Institution of Oceanography, University of California—San Diego, La Jolla, California 92093**

Received 10 February 2006/Accepted 26 April 2006

Copper (Cu) is an essential micronutrient required as a redox cofactor in a number of enzymes, including cytochrome oxidase, plastocyanin, and Cu/Zn superoxide dismutase. However, due to the redox chemistry of Cu, it is a potent toxin at elevated concentrations, and organisms utilize homeostatic mechanisms to tightly control both the intracellular concentration and activity of Cu (26).

One means of Cu detoxification includes the synthesis of metal-binding ligands. The primary types of described metal-binding ligands are metallothioneins and phytochelatins, which are cysteine-rich protein molecules found in the plant and animal kingdoms (9), with metallothioneins also occurring in the prokaryotic genus *Synechococcus* (36). The amino acid sequences of metallothioneins are gene encoded, while phytochelatins are enzymatically produced by phytochelatin synthase. Although both of these ligands have important roles in metal detoxification, additional functions have not been ruled out, including roles in essential metal ion homeostasis. (For a review of phytochelatins and metallothioneins, see reference 9.)

A number of transporter families contain members that specifically transport metals, such as the cation diffusion facilitator (CDF) family (40), a subgroup of P-type ATPase transporters referred to as P_{H^+}-type ATPases (initially named CpX-type ATPases) (4, 50), and certain ABC-type transporters (39, 53). Cu-transporting P-type ATPases have been identified in a wide variety of organisms and are important for Cu homeostasis. While most described Cu-transporting P-type ATPases are responsible for exporting Cu(I) (4), a possible role in metal uptake has also been suggested (35, 42). The action of Cu-specific transporters can ameliorate the effects of Cu stress by either excreting Cu from the cell or transporting it into compartments where it is sequestered away from sensitive cellular sites. Some prokaryotes (for a review of efflux-mediated metal resistance in prokaryotes, see reference 34) and the eukaryote *Candida albicans* (57) utilize dedicated extrusion transporters to remove excess Cu from the cytoplasm. Yeast (39), plants (56), and diatoms (33) sequester excess metals, such as cadmium, zinc, or copper, into vacuoles, and in *Schizosaccharomyces pombe* this mechanism relies on an ABC-type transporter localized to the vacuolar membrane (39). *Escherichia coli* utilizes a Cu-exporting P-type ATPase to transport excess Cu into the periplasmic space and away from the cytoplasm (45).

Eukaryotes can also down-regulate the synthesis of metal uptake transporters in response to excess Cu (27). In *Saccharomyces cerevisiae*, both the nutritional Cu sensor, Mac1p, and the toxic Cu sensor, Ace1p, are required for cell survival under toxic conditions (41). In this role, Mac1p is responsible for sensing excess Cu and down-regulating high-affinity Cu uptake transporters. Additionally, yeast proteolytically degrades high-affinity Cu uptake transporters at the plasma membrane when exposed to elevated levels of Cu (37). Down-regulation and proteolytic degradation of uptake transporters presumably prevent additional import of Cu into cells already experiencing toxic levels.

Cu stress has environmental ramifications. Cu pollution is increasing in coastal California waters (52) and is likely increasing in other coastal environments, which may affect the growth and species composition of phytoplankton at the base of aquatic food webs. Most abundant among these are diatoms, which are unicellular, eukaryotic microorganisms that are typically photosynthetic (47) and encased in a shell made of silica called a frustule. Diatoms are ecologically important because they are found throughout the world in both freshwater and marine systems, are major constituents of the base of aquatic food webs, are responsible for ~20% of global carbon fixation (5, 58), and are the dominant contributors to biosilicification (55).
Until recently, little molecular information about diatoms was available. However, the genome sequence of the marine diatom *Thalassiosira pseudonana* (5) is now publicly available (www.jgi.doe.gov/). *T. pseudonana* was originally isolated from a polluted estuary, and it is highly resistant to Cu and cadmium (7). Molecular mechanisms for this resistance are undescribed, but the presence of a genome sequence facilitates the study of Cu homeostasis and detoxification mechanisms in this unexplored unicellular eukaryotic system. In order to understand the cellular response of these ecologically important organisms to Cu, we utilized in this study a PCR-based subtractive cDNA approach to identify Cu-responsive genes.

**MATERIALS AND METHODS**

**Cell culture.** An axenic culture of *Thalassiosira pseudonana* (Hustedt) Hasle and Heimdal clone 3H (CCMP 1335) was obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton, Bigelow Laboratory for Ocean Sciences. Cultures were grown in f/2 medium (15) made with autoclaved, autoclaved, local seawater. f/2 vitamins and inorganic nutrients (15) were added. Cells were cultured at 18°C under continuous cool white fluorescent light at approximately 119 μmol quanta·m⁻²·s⁻¹. Sterility was monitored by occasional inoculation into tryptone-/enriched media to check for bacterial growth (3). Cells intended for nucleic acid extraction were batch cultured and growth was monitored by cell counts with a Petroff-Hauser counting chamber. Small-scale growth studies to test various stress conditions were performed by growing a 1-liter culture to early cell density of approximately 1.5×10⁵ cells·ml⁻¹. Eight liters of exponentially growing culture, and cells were incubated for 1 h. Cells subjected to Cu stress that were intended for real-time reverse transcriptase PCR (RT-PCR) analysis were grown in three 8-liter batch cultures to a cell density of 1.5×10⁵ cells·ml⁻¹. One culture was harvested as a control, and the other two cultures were incubated with 10 μM CuSO₄ for 1 and 24 h. Cultures were treated with cycloheximide to arrest mRNA translation for 1 h and harvested by filtration as described above.

**Incubations to induce cell stress.** For cDNA subtraction, *T. pseudonana* was grown in three 8-liter batch cultures in polycarbonate bottles to a density of 3×10⁶ cells·ml⁻¹. Eight liters of exponentially growing culture, and cells were incubated for 1 h. Cells subjected to Cu stress that were intended for RT-PCR analysis were grown in three 8-liter batch cultures to a cell density of approximately 1.5×10⁵ cells·ml⁻¹. One culture was harvested as a control, and the other two cultures were incubated with 10 μM CuSO₄ for 1 and 24 h. Cultures were treated with cycloheximide to arrest mRNA translation at a final concentration of 20 μg/ml at 10 min before cells were harvested by centrifugation at 10,000 × g for 10 min.

**Metal titration incubations were performed as follows.** Eight-liter batch cultures were grown to early exponential phase and then aliquoting 20 ml into 50-ml glass tubes to which various concentrations of Cu, H₂O₂, or Cd were added. Growth was monitored by measuring chlorophyll fluorescence on a Turner Designs fluorometer (model 9-8AU-000, with a red-sensitive photomultiplier) using filters with an excitation wavelength of ~450 nm and an emission wavelength of ~670 nm. Cells were also performed to confirm trends observed by measuring chlorophyll fluorescence on a Turner Designs fluorometer (model 9-8AU-000, with a red-sensitive photomultiplier) using filters with an excitation wavelength of ~450 nm and an emission wavelength of ~670 nm. Cells were grown in three 8-liter batch cultures in polycarbonate bottles to a density of approximately 1.5×10⁵ cells·ml⁻¹. Eight liters of exponentially growing culture, and cells were incubated for 1 h. Cells subjected to Cu stress that were intended for real-time reverse transcriptase PCR (RT-PCR) analysis were grown in three 8-liter batch cultures to a cell density of 1.5×10⁵ cells·ml⁻¹. One culture was harvested as a control, and the other two cultures were incubated with 10 μM CuSO₄ for 1 and 24 h. Cultures were treated with cycloheximide to arrest mRNA translation at a final concentration of 20 μg/ml at 10 min before cells were harvested by centrifugation at 10,000 × g for 10 min.

**RNA purification and reverse transcription.** Total RNA was extracted with Tri Reagent (Sigma) as previously described (18). Contaminating genomic DNA was removed from RNA intended for RT-PCR by using the RNaseasy Mini Kit (QIAGEN) in conjunction with the RNase-free DNase Set (QIAGEN). After DNase treatment, mRNA was purified from total RNA extracted from cells exposed to Cu for 1 and 24 h with the MicroPoly(A) Purist mRNA Purification Kit (Ambion) according to the manufacturer’s recommendations. mRNA was not purified from cells incubated with H₂O₂ or increasing concentrations of Cu or Cd; instead, total RNA was used for reverse transcription and RT-PCR, yielding comparable results. RNA was quantified with a fluorometric assay using the Ribogreen RNA Quantitation Assay Kit (Molecular Probes, Inc.) in a 96-well microplate format according to the manufacturer’s recommendations. Fluorescence as a result of excitation at 532 nm was measured with a Typhoon 9410 instrument (Amersham Pharmacia Biotech) using the 526SP emission filter. An equivalent amount of total RNA or mRNA from each time point or metal concentration in individual incubation experiments was reverse transcribed with Superscript II RNase H⁻ reverse transcriptase (Invitrogen) according to the manufacturer’s protocol. Reverse transcription was primed with oligo(dT)₁₀–₁₅ (Invitrogen) and carried out in the presence of RNasin (Promega) RNase inhibitor. We confirmed that the RNA was free of contaminating genomic DNA by PCR using the resulting cDNA as a template and primers that encompassed an intron.

**cDNA subtraction.** cDNA subtraction was performed with the PCR-Select cDNA Subtraction Kit (Clontech) as described by the manufacturer. Forward and reverse subtraction libraries were designed to identify cDNAs whose expression was increased or decreased, respectively, by Cu exposure. Briefly, forward subtraction was performed by converting mRNA populations from control and Cu-exposed cells to double-stranded cDNA and digestion with RsaI to yield blunt ends. cDNA from Cu-exposed cells, termed the “tester” cDNA population, was divided into two portions that were ligated with different adapters. cDNA from control cells, or “driver” cDNA, was added in excess and allowed to hybridize with each of the adapter-ligated tester populations. A second hybridization was then performed between the two primary hybridization samples, allowing cDNAs unique to the tester population, and with different adapters, to hybridize. After hybridizations were complete, the overhanging adapters were filled in with Taq DNA polymerase (Roche). Only cDNAs unique to the tester population had different adapters at each end allowing exponential PCR amplification. cDNAs having only one adapter, or the same adapter at each end, would amplify linearly, or form a secondary structure and not amplify, respectively. The reverse subtraction was performed in the same way except that cDNA from control cells was used as the driver population.

**cDNA screening by autoradiography.** Colony lifts of subtracted, cloned cDNAs onto Hybond N filters (Amersham) were performed as described previously (48) except that DNA was cross-linked to the filters by UV exposure with a UV Stratalinker 1800 instrument (Strategene). Filters were stripped (0.1 N NaOH) for 30 min at 65°C to remove residual bacterial debris and then neutralized with saturated NaCl solution (SNC) [0.1 N Tris·HCl (pH 7.5), 0.1 N NaOH, 0.1 N NaCl, 0.1 N SNC (pH 7), 0.1 N NaCl, 0.015 M sodium citrate, pH 7] for 30 min at 35°C. Filters were then rinsed with water and prehydrated in hybridization solution (6× SSC, 0.25% nonfat dried milk). All hybridizations were performed at 55°C.

**Probes were created from both the forward and reverse subtracted cDNAs.** Products of the secondary PCR performed with nested adapter primers and subtracted cDNAs as the template were digested with Rsal (New England Biolabs) to remove the adapter sequence. This material was then sequentially digested with Smal and Eagl (New England Biolabs) to ensure the removal of adapter sequence in the event that the Rsal site had been destroyed during ligation. Adapters were separated from cDNA with Sephadex S-400 HR Microspin Columns (Amersham Biosciences). Probes were generated by random priming in the presence of [³²P]dCTP with the Prime-It II kit from Stratagene according to the manufacturer’s recommendations. Subtracted cDNA was ligated with both forward and reverse subtracted cDNAs were first hybridized with the forward probe. Blots were then stripped as described above and hybridized with a probe made from the reverse cDNA library. Equivalent counts per minute of forward and reverse probes were added to the hybridization solution so that labeling intensity could be directly compared. Clone 76 was selected to generate a probe to identify how many of the forward subtracted clones corresponded to the same or similar genes. The insert from clone 76 was PCR amplified, and the adapters were removed by digestion with NotI and Eagl. DNA was excised from a 1% agarose gel and purified with the Rapid Gel Extraction System (Marligen Bioscience, Inc.). The probe was labeled by random priming as described above.

**After overnight hybridization with the probe, filters were washed twice at room temperature in 2× SSC and 0.1% sodium dodecyl sulfate for 5 min and then twice at 50°C for 40 min. Hybridization was imaged by exposing a general-purpose storage phosphor screen (Molecular Dynamics) to the radioactively labeled filters and subsequently scanning the screen with a Typhoon 9410 instrument (Amersham Pharmacia Biotech). The screen was exposed for equivalent amounts of time when forward and reverse subtracted cDNAs were used as probes so that labeling intensity could be compared. Densitometry of labeling intensity was performed with ImageQuant software (Molecular Dynamics).

**RT-PCR.** mRNA levels corresponding to genes identified in the forward subtracted cDNA library were quantified after cells were exposed to Cu, Cd, or H₂O₂ by performing RT-PCR with a Lightcycler (Roche) and either the Lightcycler DNA Master SYBR green I (Roche) or the Lightcycler DNA Master SYBR green I (Eurogentec) DNA amplification kit. For genes on scaffold 73 that were highly induced, a 1:10 dilution was performed on the cDNA from 1 h of Cu exposure.
TABLE 1. Oligonucleotides used in this study

| Primer | Orientation | Sequence (5' to 3') | Fragment size (bp) |
|--------|-------------|---------------------|-------------------|
| 1B     | Forward     | TCTCCGAGAATGCTCAAGCA | 310               |
| 1C     | Reverse     | ATCATTCTCACCATAAAGGAA | 318               |
| 2C     | Reverse     | GCCCTCACAGCAGTGTTTC | 318               |
| 3A     | Forward     | AGAAGATCAGTGGATATGG | 303               |
| 3B     | Reverse     | ACATCGCAGCTTCACTGG | 280               |
| 4A     | Forward     | ACTTGGCTCTACAGGGAATG | 278               |
| 4B     | Reverse     | AGCTTGGCTCTACAGGGAATG | 278           |
| 5A     | Forward     | TCTGTTATCCTTCGTACCT | 291               |
| 5B     | Reverse     | TCTGTTATCCTTCGTACCT | 291               |
| 6B     | Reverse     | TACCATAGCCAGCTTAACATT | 148               |
| 7A     | Forward     | CAAGAAGGAGCCCTACCATC | 229               |
| 7B     | Reverse     | CGGTGTAACCTGAGTTCCTG | 291               |
| 8A     | Forward     | TGGAGGTGTGTTGACTCCG | 278               |
| 8B     | Reverse     | TGGAGGTGTGTTGACTCCG | 278               |
| 9A     | Reverse     | AGGCTGTGATATAGGCAAGGC | 261               |
| 10A    | Forward     | CAGAAGTTTGAGAGGGATG | 261               |
| 10B    | Reverse     | CAGAAGTTTGAGAGGGATG | 261               |
| 11A    | Forward     | ACATGAAACTACATGAGG | 229               |
| 11B    | Forward     | ATGACAACAAGTCAGGAAG | 265               |
| 12A    | ?           | GGTGAGCCTGTGCACCATAGG | 119               |
| 12B    | ?           | GGTGAGCCTGTGCACCATAGG | 119               |
| 13A    | Forward     | AGGGTCTTCTTGTTAGGCTAG | 207               |
| 13B    | Reverse?    | GGGAATTTAACCCTTGTCAT | 207               |
| 14A    | Forward     | ACACCACCGGAGATCGAAGG | 261               |
| 14B    | Reverse     | ACACCACCGGAGATCGAAGG | 261               |
| 15A    | Forward     | GCCTATTTGAGAGACGAGGAA | 261        |
| 15B    | Reverse     | GCCTATTTGAGAGACGAGGAA | 261        |
| 16A    | Forward     | AGGCTGCACTATGAGGCTAGC | 280               |
| 16B    | Reverse     | AGGCTGCACTATGAGGCTAGC | 280               |
| 17A    | Forward     | AGCTGCTTAAGTGGGAGTGTG | 280               |
| 17B    | Reverse     | AGCTGCTTAAGTGGGAGTGTG | 280               |
| 18A    | Forward     | ACTGCTCCAATTGCTTGGTG | 322               |
| 18B    | Reverse     | ACTGCTCCAATTGCTTGGTG | 322               |
| 19A    | Forward     | TCCTGGTTACGATAGCTTCA | 277               |
| 19B    | Reverse     | TCCTGGTTACGATAGCTTCA | 277               |
| p150A  | Forward     | GTCCTGTCATCATCAGAT | 171 (283) |
| p150R  | Reverse     | GTCCTGTCATCATCAGAT | 171 (283) |
| p150-likeA | Forward     | TCTGTTGCTGAGCAGCTTCA | 334           |
| p150-likeB | Reverse     | TCTGTTGCTGAGCAGCTTCA | 334           |
| p150-likeA | Forward     | TCTGTTGCTGAGCAGCTTCA | 334           |
| p150-likeB | Reverse     | TCTGTTGCTGAGCAGCTTCA | 334           |
| NR1    | Forward     | CCTGGATATCTGTTGAGAAGT | 106               |
| NR2    | Reverse     | CCTGGATATCTGTTGAGAAGT | 106               |
| GPX1 F | Forward     | CAAAGGCGGTCAGGCTCTG | 192               |
| GPX1 R | Reverse     | CAAAGGCGGTCAGGCTCTG | 192               |
| GPX2 F | Forward     | TCCGGAGCAATTCCGCAGGGTG | 166               |
| GPX2 R | Reverse     | TCCGGAGCAATTCCGCAGGGTG | 166               |
| GPX-like F | Forward     | AAAATCTCCAAGAATACATC | 229               |
| GPX-like R | Reverse     | AAAATCTCCAAGAATACATC | 229               |
| Catalase F | Forward     | GTTGATGATTGTCCTGCTGAC | 229               |
| Catalase R | Reverse     | GTTGATGATTGTCCTGCTGAC | 229               |

* Genomic fragment sizes are in parentheses in cases in which the primers encompass an intron.

** Cue2 was amplified by primers 1B and 2C.**

*** Cue19 was amplified by primers 18B and 19A.***

Cloning, sequence determination, and analysis. PCR products resulting from RACE and cDNA subtraction were cloned into pCR 2.1-TOPO vector (Invitrogen) as recommended. RACE products were sequenced with BigDye Terminator Cycle Sequencing chemistry, version 3.1, on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Cloned, subected cDNAs were sequenced either this way or with D’Enamie ET Dye Terminator chemistry on a MegaBACE DNA Analysis System (Amersham Biosciences) according to the manufacturer’s instructions. The genome sequence of *T. pseudonana* is publically available at www.jgi.doe.gov. The browser window displays representations of genomic scaffolds and contiguous assembled DNA sequence, as well as gene models based on Grail and Genewise prediction software. Amino acid sequence alignments were performed with CLUSTAL W (54), and signal peptides were predicted with SignalP (6). PredictNLS (10) was used to identify potential nuclear localization signals (NLSs), and transmembrane domains were predicted with TMHMM (51). Sequence similarity searches were performed with BLAST (2), and conserved domains were identified with SMART (28).

Nucleotide sequence accession numbers. cDNA sequences of characterized Cue genes have been deposited in the GenBank database (accession numbers DQ115006 to DQ115015). See Table 3 for a list of the Cue genes and corresponding accession numbers.

RESULTS

Subtraction library results. A forward subtracted cDNA library was constructed to identify Cu-induced genes, yielding a total of 143 clones. Fifty-six of the clones either could not be sequenced or yielded poor-quality sequence. Many clones had the same adapter ligated to both the 5’ and 3’ ends, which should have suppressed PCR amplification due to the formation of a secondary structure (49), thereby eliminating these cDNA from the library. However, the subtraction process is imperfect. These cDNA fragments were present in the subtracted library, but due to undesirable amplification and were omitted from further analysis. Seven clones contained only vector sequence. In initial screening, we sequenced 39 correctly subtracted clones and determined that 25 were identical or nearly identical. One clone, 76, was selected to generate a probe to quantify the number of clones corresponding to the same or similar genes. In addition to the 25 sequenced clones, this probe strongly hybridized with an additional 41 clones (data not shown) which were not sequenced. Therefore, a total of 66 clones out of 80 correctly subtracted clones represented the same or similar genes. In total, 17 unique genes were identified in the forward subtracted cDNA library, and the number of sequenced clones representing each gene is shown in Table 2.

mRNA concentrations measured with RT-PCR. RT-PCR was used to measure mRNA levels in Cu-exposed cells to confirm that genes identified in the forward subtracted library were up-regulated and to observe expression patterns over time. mRNA levels for each gene after 1 and 24 h of Cu exposure were normalized to levels measured in control cells and displayed in Table 2 and Fig. 1A. Potential Cu-induced genes numbered 1 to 12, 14, 15, 18, and 19 were confirmed by RT-PCR to be induced by 1 h of Cu exposure (Table 2, Fig. 1A; also see Fig. 3) and designated “Cu-expressed genes” (Cue genes).

Cue genes located within an inverted repeat on scaffold 73 of the *T. pseudonana* genome. RT-PCR revealed that after 1 h of Cu exposure, mRNA concentrations corresponding to the Cue genes located within an inverted repeat on genomic scaffold 73 were between 500- and 2,000-fold higher than in control conditions (Table 2, Fig. 1). Replicate RT-PCRs were performed
for each gene, and the average result is shown in Fig. 1A. These genes, Cue1, Cue2, Cue6, and Cue7, were the ones that were numerically highly represented in the forward subtracted cDNA library (Table 2). An additional pair of genes located within this inverted repeat and designated Cue18 and Cue19 (Fig. 1B) were not represented in the subtracted library but were determined by RT-PCR to also be highly induced by Cu (Table 2, Fig. 1). After 24 h of Cu exposure, mRNA levels corresponding to the genes on scaffold 73 decreased to be between 4- and 25-fold higher than control levels. Although the coding sequences for these gene pairs were nearly identical, the 3′ UTRs for all four genes extended beyond the inverted repeat and therefore were distinct for each gene. Cue1, Cue2, Cue6, and Cue7 are related proteins, as determined by amino acid alignments (Fig. 2A), and collectively displayed 27% identity and 41% similarity. All of these proteins share a 73-amino-acid domain, and Cue6 and Cue7 are larger primarily due to a repeat of this domain. The domain contains a number of potential metal-binding residues, including six cysteines and four methionines as well as seven glutamic acid and seven aspartic acid residues.

Cue1 and Cue2 were predicted to contain an N-terminal signal sequence as well as a putative NLS (Fig. 2A). Cue6 and Cue7 were predicted to have a 23-amino-acid N-terminal signal sequence (Fig. 2A). Amino acid searches using BLAST failed to identify proteins with significant similarity (Table 3); therefore, all four of these proteins are novel. Cue1 and Cue2 contain weak homology to a domain (pfam07172, GRP) (E = 0.012) found in a protein family that includes glycine-rich proteins as well as proteins induced in response to various stresses. Statistically significant domains could not be identified in Cue6 or Cue7.

Cue18 and Cue19 were characterized based solely on their gene models and were fully contained within the inverted repeat. Amino acid sequence alignment revealed 81% similarity and 80% identity between them (Fig. 2B). Their predicted coding sequences differ primarily due to a stretch of amino acids consisting of multiple repeats of proline-threonine-proline-leucine (PTPL) within Cue19 that are absent from Cue18. Both proteins are predicted to have identical 22-amino-acid N-terminal signal sequences. No proteins with similar sequences were identified by BLAST searches (Table 3). Cue18 and Cue19 were both enriched in potential metal-binding and acidic amino acids. Statistically significant domains were not identified.

Cue genes located within an inverted repeat on scaffold 23 (chromosome 9) of the T. pseudonana genome. Cue4 and Cue5 were represented by one clone in the forward subtracted cDNA library (Table 1). These genes are completely contained within an inverted repeat of 2,390 bp on genomic scaffold 23, located within chromosome 9, and were 100% identical at the amino acid level. Gene models for Cue4 and Cue5 were available: grail.23.58.1 and grail.23.68.1, respectively. However, full-length cDNA revealed that these models contained errors, including a miscalled start codon and an incorrectly designated intron which omitted part of the coding sequence from the gene model. The corrected amino acid sequence of Cue4 and Cue5 could not be assigned function based on BLAST analysis (Table 3). Due to the level of similarity of Cue4 and Cue5, it was not possible to develop unique primers and it could not be determined whether mRNA levels measured by RT-PCR represented one or both genes. Although also located within an inverted repeat, Cue6 and/or Cue5 had a substantially lower level of expression and a different expression pattern than genes located on scaffold 73. After a 1-h Cu treatment, mRNA

### Table 2. RT-PCR results

| Gene(s)       | No. of clones sequenced from subtracted cDNA library | Cu-induced fold change in mRNA concn | H₂O₂-induced fold change in mRNA concn |
|---------------|------------------------------------------------------|-------------------------------------|----------------------------------------|
|               |                                                      | 1 h | 24 h | 1 h | 24 h |
| Potential Cu-expressed genes† | 11 | 1,977 | 4 | 2.9 | 5.1 |
|               | 2 | 1,545 | 23 | 2 | 8.4 |
|               | 1 | 49 | 21 | 3.4 | 19.9 |
|               | 4, 5 | 1 | 1 | 2 | 0.4 | 5 |
|               | 6 | 1,826 | 11 | 2 | 0.9 |
|               | 7 | 1,249 | 7 | 1.1 | 0.7 |
|               | 8 | 1 | 46 | 5 | 1.5 | 1.1 |
|               | 9 | 1 | 4 | 1 | 2.7 | 2.1 |
|               | 10 | 1 | 7 | 2 | 3.7 | 3.2 |
|               | 11 | 2 | 1,101 | 203 | 2 | 2.3 |
|               | 12 | 1 | 8 | 3 | 1.4 | 2.2 |
|               | 13 | 1 | 0.4 | 0.4 | — |
|               | 14 | 1 | 2 | 2 | 1.8 | 2.6 |
|               | 15 | 2 | 13 | 2 | 1.5 | 5.1 |
|               | 16 | 1 | 0.9 | 0.1 | — |
|               | 17 | 1 | 0.6 | 0.8 | — |
|               | 18 | 0 | 514 | 16 | 0.8 | 4.6 |
|               | 19 | 0 | 772 | 25 | 0.8 | 11 |
| Cu-expressed genes‡ | 1 | 0 | 2 | 26 | 1.3 | 4.7 |
|               | 1 | 0 | 2 | 2 | 1.4 |
| Genes potentially involved in H₂O₂ degradation | 0 | 0.8 | 0.3 | 1.4 | 0.7 |
|               | 0 | 1.5 | 1.1 | 1.3 | 1.6 |
|               | 0 | 2.7 | 2.3 | 2.1 | 2.2 |
| Catalase | 0 | 2 | 2.4 | 0.5 | 0.6 |

† Details of 1-h forward subtracted cDNA library and RT-PCR results. —, not determined.
‡ RT-PCR results for genes previously shown to be induced by 24 h of Cu exposure.
levels of Cue4 and/or Cue5 remained close to those of control cells but displayed a longer-term response to Cu by increasing by twofold after 24 h (Table 2, Fig. 3).

Characterization of additional Cue genes for which full-length cDNA was acquired. Cue10 was represented in the forward subtracted cDNA library by a single clone. To confirm the gene model prediction (grail.22.30.1), full-length cDNA for Cue10 was acquired. Cue10 is not predicted to have an N-terminal signal sequence but does have one transmembrane domain. Additionally, particular amino acids, such as glycine (13.2%), serine (12.9%), aspartic acid (10.1%), histidine (7.7%) and methionine (6.8%), are highly represented in Cue10, frequently occurring in uninterrupted stretches. This gene could not be assigned a function based on amino acid similarity searches (Table 3); however, the large proportion of amino acids that are well suited to bind metals suggests that Cue10 may have a role in metal binding. Statistically significant domains could not be identified in Cue10; however, it is predicted to contain a coiled-coil region at the C terminal end.

mRNA levels for Cue10 were sevenfold higher than control cells after 1 h of Cu exposure, decreasing to twofold higher after 24 h (Table 2, Fig. 3).

Cue11 and Cue15 were not predicted in the genome and were characterized only after cDNA was obtained. Cue11 did not contain a predicted N-terminal signal sequence, but it did contain seven predicted transmembrane segments. Although BLAST analysis identified proteins with similarity to Cue11 (highest E value at 8e–57), the sequences were bacterial, putative, or uncharacterized membrane proteins to which no function was assigned (Table 3). Cue11 contains a domain of unknown function (pfam03988, DUF347) (E = 8.1e–5) that occurs in a family of bacterial membrane proteins. After 1 h of Cu exposure, Cue11 was highly induced, at 1,101-fold, and then decreased to 203-fold above control levels after 24 h (Table 2, Fig. 3).

Cue15 was predicted to have a 19-amino-acid N-terminal signal sequence. BLAST analysis identified an unnamed protein product (E = 1e–4) that is a “pathogenesis-related-1” homolog from Nicotiana tabacum as being most significantly similar to Cue15. Cue15 and the pathogenesis-related-1 protein both contain a conserved domain (pfam00188, SCP), which belongs to a family of extracellular domains with no known function. Cue15 was induced by 13-fold after 1 h of Cu exposure and decreased to 2-fold after 24 h (Table 2, Fig. 3).

Characterization of additional Cue genes for which full-length cDNA was not acquired. Cue3, Cue8, and Cue9 were represented in the genome by gene model numbers grail.35.48.1, grail.10.4.1, and grail.3.377.1, respectively. Although these gene models were incomplete, appearing to be truncated at their 5’ ends, they had reasonable similarity to genes of known function (Table 3); therefore, characterization of full-length cDNA was not pursued. These genes displayed a pattern of induction with Cu exposure similar to that of other Cue genes examined (Table 2, Fig. 3) in that they were more highly induced after 1 h
than 24 h. Cue3 was most similar to a serine protease from *Bdellovibrio bacteriovorus* (E = 2e–37) and contained a domain found in serine protease, trypsin family proteins (IPR001254).

The characterized portion of Cue8 contains the HSF-type DNA-binding domain (IPR000232) and a leucine zipper (IPR002158) and has significant similarity (E = 3e–12) to a HSTF from *Danio rerio* (zebrafish). Cue9 contains a RNase II domain (IPR001900) and was found to be significantly similar (E = 3e–88) to a hypothetical RNase II, RNB protein family member from *Schizosaccharomyces pombe* that is similar to the gene *dis3*.

Cue12 and Cue14 were not predicted in the genome. 5′ RACE was unsuccessful for Cue14; however, the 3′ end consisted of 730 bp, coding for 160 amino acids. BLAST analysis failed to identify amino acid sequences similar to this portion of the protein. Cue14 was induced by twofold at 1 h and remained at that level after 24 h of Cu exposure (Table 1, Fig. 3). Neither 5′ nor 3′ RACE was successful for Cue12. It was...
Possible to translate the short DNA fragment obtained from the subtracted library in two frames; however, BLAST analysis failed to identify homologous amino acid sequences for either frame. 

**Cue12** was induced by eightfold at 1 h and decreased to threefold after 24 h of Cu exposure (Table 1, Fig. 3).

**Genes down-regulated by copper stress.** RT-PCR results revealed that potential Cu-expressed genes 13, 16 and 17 were false positives and demonstrated a decrease in mRNA concentration relative to controls after Cu exposure (Table 2). To serve as an example of a gene down-regulated by Cu stress, a clone was examined from the reverse subtracted cDNA library. Based on sequence similarity, this clone was found to represent nitrate reductase (NR) and was confirmed by RT-PCR to be down-regulated after Cu exposure for 1 h (Fig. 3). After 24 h, NR expression returned to levels slightly above those measured in control cells (Fig. 3).

**Expression patterns: Cue genes and previously characterized, copper-induced proteins.** Most of the Cue genes were highly induced at 1 h and maintained an elevated but attenuating expression level after 24 h of exposure (Table 2, Fig. 1A, Fig. 3), which is consistent with a rapid and then attenuating metal stress response. We compared this response to that of p150, a protein known to accumulate on the cell surface of T. pseudonana after 24 h of Cu exposure, as well as with a p150-like gene identified in the genome (11). In contrast to most of the Cue genes characterized in this study, the expression pat-
terns of the p150 and p150-like genes (Table 2, Fig. 3) suggest that they are part of a longer-term response that may include secondary effects of Cu toxicity, such as cell cycle arrest (11).

**Behavior of Cue and other genes during stress induced by H$_2$O$_2$ exposure compared to Cu exposure.** To determine whether Cue genes responded to other cellular stresses, their behavior was examined in cells experiencing H$_2$O$_2$-induced stress for 1 and 24 h. We first determined H$_2$O$_2$ levels that induced stress in *T. pseudonana* (Fig. 4A). H$_2$O$_2$ concentrations at or below 0.05 mM had no effect on growth relative to controls (Fig. 4A), and concentrations at 0.5 mM (Fig. 4A) or above (data not shown) resulted in cell death. Growth was inhibited by 0.1 mM H$_2$O$_2$ for at least 24 h after addition; however, cells were able to achieve control levels after 48 h probably due to the destruction of H$_2$O$_2$ in the culture tubes. Because 0.1 mM H$_2$O$_2$ induced cell stress but not cell death, similar to the Cu treatment, this concentration was used to monitor the response of the Cue genes to H$_2$O$_2$-induced stress for up to 24 h. In general, the timing and magnitude of the response of most of the Cue genes to H$_2$O$_2$ stress is different from that for Cu stress (Table 2, Fig. 4B). For example, Cue genes located in the inverted repeat on scaffold 73, Cue11, and Cue8 were much more highly induced by 1 h of Cu exposure than by H$_2$O$_2$ exposure. The response of these and most other Cue genes attenuates after 24 h of Cu exposure; however, attenuation does not generally occur with H$_2$O$_2$ exposure (Fig. 4B).

The response of additional non-Cue genes to H$_2$O$_2$ stress was also examined. The p150 gene displayed a response over time (i.e., increasing mRNA level with length of exposure) similar to that with Cu stress but at a lower magnitude (Table 2, Fig. 4B). The p150-like gene was not highly induced by
either stressor, but levels appeared to be slightly higher after a 1-h exposure to H$_2$O$_2$ than a 24-h exposure (Table 2, Fig. 4B). Genes potentially involved in H$_2$O$_2$ degradation, including glutathione peroxidase gene 1 (GPX1) and GPX2, as well as a glutathione peroxidase homolog (GPX-like), and a catalase gene were annotated in the *T. pseudonana* genome (newV2.0.genewise.27.59.1, newV2.0.genewise.61.23.1, genewise.32.213.1, and grail.151.10.1, respectively) and were examined in an attempt to monitor oxidative stress at the transcription level (Table 2, Fig. 4B). The response of these genes to either stressor was minimal. GPX1 and GPX2 were either down-regulated or close to control levels upon exposure to either Cu or H$_2$O$_2$. The GPX-like gene appeared to be slightly induced (2- to 2.7-fold increase in mRNA) by both conditions of stress. The catalase gene was induced approximately twofold by Cu exposure but was down-regulated by H$_2$O$_2$ exposure. Multilevel (including posttranscriptional and posttranslational) regulation of antioxidant enzymes has been described for a number of organisms (8, 44), and the level of regulation of antioxidant genes in *T. pseudonana* is not known at this time, which may explain these results.

FIG. 5. Effects of Cu and Cd exposure on *T. pseudonana*. Growth of *T. pseudonana* exposed to various concentrations of Cu (A) and Cd (B) is shown. Cell growth was followed by measurement of chlorophyll fluorescence, and results are presented in relative fluorescence units (RFU). Results shown are of representative growth experiments. Cells grown in batch culture were aliquoted into various treatment conditions on day 0. Error bars represent standard deviations for replicate cultures and are contained within the corresponding symbol when not visible. Expression changes of a subset of *Cue* genes were monitored by measuring mRNA levels with RT-PCR after cells had been exposed for 1 h to either Cu (C) or Cd (D) at the concentrations indicated in the keys at right. Note that the y axes are logarithmic scales.
Responses of Cue genes to increasing concentrations of Cu and Cd. To further evaluate the specificity of the response of Cue genes to Cu, the dose-dependent behaviors of these genes to different concentrations of Cu and Cd were examined. Small-scale experiments were performed to determine the effects of various Cu or Cd concentrations on cell growth (Fig. 5A and B). Final Cu or Cd concentrations at or below 3 μM did not inhibit growth. However, 10 μM Cu halted cell growth and 30 μM Cu resulted in cell death (Fig. 5A). A concentration of 10 μM Cd inhibited growth compared to controls; however, 30 μM Cd was required to completely halt growth (Fig. 5B). The expression behaviors of select Cue genes were monitored in concentrations of Cu and Cd that did not result in cell death. For Cu, all six Cue genes within the inverted repeat on scaffold 73 were examined, as well as the highly Cu-induced Cue11 gene (Table 4, Fig. 5C). In addition to these genes, Cue3, Cue8, and Cue9 were monitored in Cd-exposed cells (Table 4, Fig. 5D).

Cue1 and Cue2 displayed a dose-dependent response to increasing Cu concentrations (Table 4, Fig. 5C); however, they were down-regulated by all but the lowest concentration of Cd, further suggesting their specificity for Cu. The pattern of induction of Cue6, Cue7, Cue11, Cue18, and Cue19 was slightly different in that mRNA levels were dependent on Cu concentration but only when cells were exposed to at least 3 μM Cu. All of the Cue genes examined were moderately induced in cells exposed to the lowest concentration of Cd (0.1 μM); however, mRNA levels did not consistently increase with increasing concentrations of Cd (Table 4, Fig. 5D), as was observed for Cu. In addition, the magnitude of the induction of these genes by Cd was substantially lower (Fig. 5D) than with Cu exposure (Fig. 5C). Although the two highest Cd concentrations tested (10 and 30 μM) inhibited cell growth, mRNA levels for most of the Cue genes at these concentrations were lower than in control cells. Cue3 was induced at all Cd concentrations without a consistent trend, with the highest induction being 10.4-fold at 3 μM Cd. Cue9 was induced at Cd concentrations at or below 3 μM, above which it was down-regulated.

**DISCUSSION**

Using subtractive hybridization, we identified 16 genes induced by 1 h of Cu stress. Two sets of them were located in inverted repeats. All of the genes located within the inverted repeat on scaffold 73 were functional and highly induced by Cu, did not contain introns, and encoded novel proteins (Table 3). Cue1, Cue2, Cue6, and Cue7 were related, based on amino acid alignments (Fig. 2A), and we refer to them as the Cue1 gene family. All of these proteins were predicted to contain N-terminal signal sequences (Fig. 2); however, Cue1 and Cue2 also contained a potential NLS (Fig. 2A), suggesting that members of this gene family may perform similar functions in different cellular compartments, such as the nucleus and chloroplast.

Many proteins that have the ability to sequester and transport heavy metals contain a CXXC metal-binding motif (where X is any amino acid) (12, 38), although this motif can also be found in some non-metal-binding domains. A repeated CXXC sequence is a key structural motif present in all metallothionein genes (9, 23, 24) and in plants occur in gene clusters, which have been characterized and novel metal-binding motifs. Metallothioneins are characterized as low-molecular-weight, cysteine-rich, metal-binding proteins that contain CXC clust-
metallothioneins failed to identify strict homologs; however, similarity searches may not be effective due to the diverse nature of class II metallothioneins (9, 21). Metallothionein homologs were not represented in the subtracted cDNA library. If true homologs are absent in \textit{T. pseudonana}, this suggests that the relatively small, cysteine-rich proteins from the \textit{Cue1} gene family could be functional homologs to the metallothioneins; however, direct proof of this function will require metal-binding studies.

Cadmium is the most effective inducer of phytochelatin synthesis in plants (14) and in the diatom \textit{T. weissflogii} (1). Phytochelatin synthase genes were identified in the \textit{T. pseudonana} genome (5) but were not represented in the subtracted cDNA library. In organisms equipped to synthesize both ligands, metallothiones are induced by cadmium exposure while metallothioneins are produced in response to Cu (30, 60), which may explain the absence of phytochelatin synthase genes in the subtracted library. Alternatively, it is possible that in \textit{T. pseudonana} phytochelatin synthase is constitutively expressed, as has been described for many organisms (9).

A gene similar to Cu-transporting P-type ATPases in other organisms was identified in the \textit{T. pseudonana} genome (newV2.0.genewise.122.68.1); however, this gene was not represented in our subtracted cDNA library. While Cu-transporting P-type ATPases are present in most organisms and are involved in Cu homeostasis, the P-type ATPase homolog in \textit{T. pseudonana} has not been characterized; therefore, the level of gene regulation and direction of Cu transport by the resulting protein have yet to be determined. However, \textit{Cue11} was identified in our cDNA library, was highly Cu inducible, and was predicted to contain seven transmembrane segments. \textit{Cue11} is significantly similar (E = 6e–51) to COG4705, which represents an uncharacterized membrane-anchored protein conserved in bacteria. Since transporters typically contain multiple transmembrane segments, it is tempting to speculate that \textit{Cue11} could be a novel type of Cu or Cu chelate transporter.

The \textit{Cue} genes that had clear homologs were identified as a serine protease (\textit{Cue3}), a heat shock-like transcription factor (HSTF) (\textit{Cue8}), and an RNase (\textit{Cue9}), all of which may facilitate a cell’s response to changing conditions. Proteases are induced by cellular stress and are required to prevent damaged and abnormally formed proteins from accumulating within the cell (19), as well as to enable the cell to reduce the level of a protein if its presence becomes disadvantageous. Similarly, the induction of \textit{Cue8}, an HSTF, may allow the cell to induce the transcription of genes required to respond to conditions of cellular stress. \textit{Cue9} was similar to \textit{dis3}, which is essential for mitotic control (25) and is a component of the exosome, a complex consisting primarily of exoribonucleolytic proteins that is involved in maintaining correct RNA levels in eukaryotic cells (43). Induction of an exoribonucleolytic protein may facilitate changes in mRNA levels as a result of cellular stress.

Because \textit{Cue3} and \textit{Cue9} were also induced by H$_2$O$_2$ (Table 2, Fig. 4B) and most Cd concentrations examined (Table 4, Fig. 5D), they may represent general stress response genes. Although \textit{Cue4}, \textit{Cue5}, and \textit{Cue14} represent novel genes, they were similarly or more highly induced by H$_2$O$_2$ than by Cu at all times examined and may also represent genes that respond to general or oxidative stress. In contrast, the response of \textit{Cue8} to Cu was decidedly higher than to either H$_2$O$_2$ or Cd, suggesting that it may be a transcription factor involved in a Cu-specific response.

Evaluation of the responses of selected \textit{Cue} genes to stress induced by H$_2$O$_2$ revealed that most of the \textit{Cue} genes were much more highly induced by short-term Cu stress than by H$_2$O$_2$ stress. In general, the timing and magnitude of the response of the \textit{Cue} genes to H$_2$O$_2$ was different from that for Cu exposure (Fig. 4B). Some of the \textit{Cue} genes were similarly induced by both stressors after 24 h; however, the response to long-term exposure may include secondary effects. Overall, these results suggest that most of the \textit{Cue} genes respond specifically to Cu.

To further evaluate specificity, we tested the responses of selected \textit{Cue} genes to different concentrations of Cu and Cd. In the ascomycete \textit{N. crassa}, Cu metallothionein mRNA levels were strongly dependent on the Cu ion concentration in the medium, and these metallothioneins did not respond to other metals, including Cd, Co, Ni, and Zn (31). \textit{Cue1} and \textit{Cue2} responded to Cu in a dose-dependent manner, and \textit{Cue6}, \textit{Cue7}, \textit{Cue18}, \textit{Cue19}, and \textit{Cue11} were strongly induced at a Cu concentration that did not significantly affect cell growth (Fig. 5A and C), a response consistent with the idea that these gene products ameliorate the effects of high Cu levels. The levels of induction of \textit{Cue} genes examined in this titration experiment varied from previous results (Table 2, Fig. 1A, Fig. 3); however, these data were obtained from independent incubation experiments, and it is not unexpected that induction levels may vary due to slight differences between cultures, such as cell density upon metal exposure. In contrast to their strong Cu response, this subset of \textit{Cue} genes did not demonstrate a strong response or pattern of induction upon Cd exposure. While it cannot be concluded that this subset of \textit{Cue} genes is completely Cu specific, its involvement in a general stress response (with Cd and H$_2$O$_2$ used as examples) can be ruled out.

\section*{ACKNOWLEDGMENTS}

We thank Eric Allen, Jon Flowers, and Chris Willis for helpful advice and the University of California Toxic Substances Research and Training Program for providing a forum for discussing these research results at annual meetings and retreats. We thank anonymous reviewers for improving the manuscript. Funding for this research was provided by an Environmental Protection Agency STAR Exploratory Research Grant (R827107-01-0) and a National Science Foundation graduate fellowship. Instrumentation used here was purchased with support from a National Science Foundation MRI Grant (0115801).

\section*{REFERENCES}

1. Ahner, B., and F. M. M. Morel. 1995. Phytochelatin production in marine algae. 2. Induction by various metals. Limnol. Oceanogr. \textbf{40}:658–665.

2. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. \textbf{215}:403–410.

3. Andersen, R. A., S. L. Morton, and J. P. Sexton. 1997. Provasoli-Guillard National Center for Culture of Marine Phytoplankton. J. Phycol. \textbf{33}(Suppl.):1–75.

4. Argiello, J. M. 2003. Identification of ion-selectivity determinants in heavy-metal transport \textit{P}_{\text{ATP}}-\text{type ATPases}. J. Membr. Biol. \textbf{195}53–108.

5. Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. G. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brezeinski, B. K. Chaul, A. Chiowittti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadl, U. Heilstein, M. Hildebrand, B. D. Jenkins, J. Jurka, N. V. Kapitonov, N. Kruger, W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Oboznik, M. S. Parker, B. Palenek, G. J. Paizour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar. 2004. The genome of the diatom \textit{Thalassiosira pseudonana}: ecology, evolution, and metabolism. Science \textbf{306}:79–86.

6. Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides. Signal \textbf{3}3. J. Mol. Biol. \textbf{340}:783–795.
