Triticum durum Metallothionein

ISOLATION OF THE GENE AND STRUCTURAL CHARACTERIZATION OF THE PROTEIN USING SOLUTION SCATTERING AND MOLECULAR MODELING*

Kivanc Bilecen‡§, Umit H. Ozturk‡¶, Adil D. Duru‡, Tolga Sutlu‡, Maxim V. Petoukhov‖,
Dimitri I. Svergun‖, Michel H. J. Koch‖, Ugur O. Sezerman‡, Ismail Cakmak‡,
and Zehra Sayers‡**

From the ‡Sabanci University, Faculty of Engineering and Natural Sciences, 34956, Orhanli, Tuzla, Istanbul, Turkey,
‖European Molecular Biology Laboratory, EMBL % DESY, Notkestrasse 85, Hamburg D-22607, Germany, and
¶TUBITAK, Research Institute for Genetic Engineering and Biotechnology, P. O. Box 21, 41470, Gebze, Kocaeli, Turkey

A novel gene sequence, with two exons and one intron, encoding a metallothionein (MT) has been identified in durum wheat *Triticum durum* cv. Balcali85 genomic DNA. Multiple alignment analyses on the cDNA and the translated protein sequences showed that *T. durum* MT (dMT) can be classified as a type 1 MT. dMT has three Cys-*X*-Cys motifs in each of the N- and C-terminal domains and a 42-residue-long hinge region devoid of cysteines. dMT was overexpressed in *Escherichia coli* as a fusion protein (GSTdMT), and bacteria expressing the fusion protein showed increased tolerance to cadmium in the growth medium compared with controls. Purified GSTdMT was characterized by SDS- and native-PAGE, size exclusion chromatography, and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. It was shown that the recombinant protein binds 4 ± 1 mol of cadmium/mol of protein and has a high tendency to form stable oligomeric structures. The structure of GSTdMT and dMT was investigated by synchrotron x-ray solution scattering and computational methods. X-ray scattering measurements indicated a strong tendency for GSTdMT to form dimers and trimers in solution and yielded structural models that were compatible with a stable dimeric form in which dMT had an extended conformation. Results of homology modeling and *ab initio* solution scattering approaches produced an elongated dMT structure with a long central hinge region. The predicted model and those obtained from x-ray scattering are in agreement and suggest that dMT may be involved in functions other than metal detoxification.

Metallothioneins (MTs)[1] constitute a superfamily of ubiquitously expressed low molecular mass (6–7 kDa), cysteine-rich proteins lacking aromatic amino acids (1). After first being identified as a cadmium-binding protein in horse kidney (2), MTs were found in a wide range of organisms from plants to fungi, and their high capacity to bind metals (*e.g.* copper, cadmium, zinc, mercury, and silver) in metal-thiolate clusters was demonstrated (3, 4). Several functional roles have been attributed to MTs, including heavy metal detoxification, zinc and copper homeostasis (5, 6), scavenging of reactive oxygen species (7), regulation of metalloenzymes and transcription factors (5), involvement in metabolism of metallo-drugs and alkylating agents, response to stress conditions, and the potential involvement at inflammatory sites and in apoptosis (8).

Historically, MTs have been classified in three groups according to their sequence similarities. Class I consists of those with high sequence homology to mammalian MTs; class II includes all that do not display significant similarity to class I, and class III consists of phytochelatins, enzymatically synthesized peptides with a poly(γ-glutamylcysteinyl)glycine structure found in plants (9). In recent years computational analyses resulted in a more detailed classification, where the MT superfamily was divided into families, classes, and subclasses based on the location and distribution of the cysteine residues (10). Mammalian MTs (class I) contain 20 highly conserved cysteines that are arranged in the form C*X*C (where C is a cysteine residue and *X* indicates any other amino acid) clustered in the N- and C- terminal regions. A total of seven divalent metals are bound in these two clusters ($(Me(II))_3Cys_9$ and $(Me(II))_4Cys_{11}$), and the metal binding domains are connected by a variable length hinge region devoid of cysteines (3). In contrast, a typical class II representative, yeast MT, contains 12 cysteines distributed throughout the sequence with about eight Cu(I) ions bound in a single cluster (11, 12).

The first MT identified in plants was the zinc-binding Ec protein from wheat embryos (13), and since then more than 50 MT-like sequences have been reported from a variety of species (14). Completion of the *Arabidopsis* genome sequence analysis has led to the identification of an entire suite of MT genes in a higher plant suggesting gene amplification during evolution

[1] The abbreviations used are: MT, metallothionein, *dmt*; *T. durum* metallothionein gene; *amt*, *T. aestivum* metallothionein gene; RT, reverse transcriptase; GST, glutathione *S*-transferase; dMT, *T. durum* metallothionein; GSTdMT, fusion protein of GST and dMT, MALDI-TOF-MS, matrix-assisted laser desorption/ionization time of flight mass spectrometry; Ec, Early cysteine labeled; MM, molecular mass; MOPS, 4-morpholinepropanesulfonic acid.

(15). Plant (class II) MTs have been classified into four types based on the amino acid sequences (15). Type 1 includes sequences with C*X*C clusters at the N and C termini. Type 2 MTs have CC, C*X*C, and C*XX*C motifs in the N-terminal and C*X*C motif in the C-terminal domain, and type 3 proteins are small with cysteine clusters at their N and C termini. Type 4 includes the wheat Ec protein as well as maize and rice MTs, where C*X*C motifs are distributed throughout the entire sequence. In general, plant MTs are similar to their mammalian counterparts in displaying the C*X*C patterns but contain unusually long hinge regions.

The diversity of conditions inducing expression as well as variations in the sites of expression suggest that plant MTs may be involved in stress responses, programmed cell death, developmental regulation, and heavy metal metabolism (14–22). In contrast to identification of several *mt*-like genes in plants, there are few studies on purified proteins. Apart from Ec and *Arabidopsis thaliana* MTs (23), only the pea type 1 MT (*PsMT$_A$*) was expressed in *Escherichia coli* to characterize the protein and determine its metal-binding properties (24, 25).

Soil contamination with cadmium is observed in many agricultural areas because of the increased use of cadmium-containing fertilizers, application of contaminated sewage sludge, and atmospheric deposition of cadmium on crop and soil surfaces. These factors result in accumulation of the toxic heavy metal in crops and its transfer to humans via the food chain. Studies on different wheat cultivars indicated that cadmium accumulation is higher in grains of *Triticum durum* when compared with *Triticum aestivum* (26). Furthermore, transport of cadmium taken up through the roots to different parts of the plant is different in durum and bread wheat (27, 28). In the present study, the MT gene from *T. durum* cv. Balcali85 was characterized and compared with that of *T. aestivum* cv. Bezostaja. Cadmium-binding properties of *T. durum* MT and its possible role in detoxification of cadmium was investigated by overexpressing the wheat MT in *E. coli* as a GST fusion. Structural models of the recombinant protein were obtained from x-ray solution scattering patterns and compared with three-dimensional models developed using computational tools. Structural models suggest alternative functional roles for dMT.

EXPERIMENTAL PROCEDURES

Plant Materials and Growth Conditions—Seeds of both *T. durum* (*T. durum* cv. Balcali85 and *T. durum* cv. C-1252) and *T. aestivum* (*T. aestivum* cv. Bezostaja) were surface-sterilized for 20 min in 10% (w/v) H$_2$O$_2$, rinsed with distilled H$_2$O, and germinated in perlite moistened with saturated CaSO$_4$. After 4 days, seedlings were transferred to continuously aerated nutrient solutions, which were changed every 3 days, containing 0.88 mM K$_2$SO$_4$, 2.0 mM Ca(NO$_3$)$_2$, 0.25 mM KH$_2$PO$_4$, 1.0 mM MgSO$_4$, 0.1 mM KCl, 100 μM FeEDTA, 1 μM H$_3$BO$_3$, 0.5 μM MnSO$_4$, 0.2 μM CuSO$_4$ and grown with a 16:8-h light/dark regimen at 25.0 °C. Before total RNA isolation, 10 μM CdSO$_4$ was added to the seedlings grown in nutrient solution for 48 h.

PCR and Reverse Transcription-PCR—Plant genomic DNA was isolated from 100 mg of 12-day-old frozen tissue of both *T. aestivum* and *T. durum* using the Qiagen, DNeasy® plant mini kit. Amplification was carried out using PCR Core System II (Promega) with the following cycle conditions: 1-min wait at all temperatures (95.0, 55.0, and 72.0 °C) for 39 cycles followed by an additional final extension of 1 min. Primers were 5′-ATG TCT TGC AAC TGT GGA-3′ (F1) for upstream and 5′-ACA GTT GCA GGG GTT GCA-3′ (R1) for downstream, including a stop codon. Products were separated by electrophoresis on 1.5% agarose gels, purified using QIAquick® gel extraction kit (Qiagen), inserted into pGEM-TEasy vector (Promega), and propagated in *E. coli* grown in Luria Bertani (LB) medium containing 100 μg/ml ampicillin (LBA). Isolated constructs were sequenced at Seqlab GmbH (Goettingen, Germany).

Total RNA was isolated using RNeasy® plant mini kit (Qiagen), and poly(A)$^+$ RNA was isolated from total RNA using the Oligotex® mRNA mini kit (Qiagen). Results were monitored by electrophoresis in formaldehyde (2.2 M)-agarose (1.2%) gels in MOPS buffer (0.02 M MOPS, 2 mM CH$_3$COONa, 1 mM EDTA), pH 7.0 (29).

RT-PCR was carried out using the One-step RT-PCR kit (Qiagen®) with reverse transcription at 50.0 °C for 30 min, followed by heat inactivation of enzymes for 15 min at 95.0 °C. cDNA was amplified using primers F1 and R1 during 40 cycles: with a 1-min wait at 94.0, 53.5, and 72 °C, followed by a final extension of 10 min at 72.0 °C. *mt* cDNAs were inserted into pGEM-Teasy vector (Promega) and propagated in *E. coli* grown in LBA. Isolated constructs were sequenced at Seqlab GmbH (Goettingen, Germany).

Protein Expression in E. coli and Purification of Recombinant GST-dMT and dMT—The GST fusion protein vector pGEX-4T-2 (Amersham Biosciences) was used for recombinant wheat MT expression in *E. coli* BL21(DE3). Two primers (F2, 5′-CTATGGAATTCCCCATGTCTTG-CAAC-3′, and R2, 5′-CTATGCTCGAGTTAACAGTTGCAGG-3′, containing the EcoRI and XhoI restriction sites, respectively) were designed to facilitate in-frame ligation with the vector. The construct pGEXdMT was initially propagated in *E. coli* XL1-Blue and grown in LBA for sequence analysis; later, protein expression was carried out in *E. coli* BL21 (DE3). Optimum induction was obtained with 0.75 mM isopropyl 1-thio-β-D-galactopyranoside at 37.0 °C with continuous shaking at 250 rpm for 5 h. Expression was monitored by analysis of cellular extracts by 12% SDS-PAGE (30).

Large scale purification of recombinant protein was carried out from 1.5-liter cultures grown up to $A_{600} = 1.2$ before induction in LBA containing 0.05 mM CdSO$_4$. Cells were pelleted 4 h after induction and lysed in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$), pH 7.3. About 80 ml of extract was loaded on a 5-ml GSTrap FF affinity column (Amersham Biosciences) following the manufacturer's instructions. The GSTrap column was connected to an ΔKTA™-FPLC system (Amersham Biosciences), and 50 mM Tris-HCl, pH 8.0, with 20 mM reduced glutathione was used for elution of GSTdMT at 1 ml/min collecting 1-ml fractions.

The purified protein was analyzed by 12% SDS-PAGE, 4–20% native PAGE, and Western blot analysis (29), and its absorption spectrum was determined between 200 and 800 nm (Shimadzu UV 3150 UV-Visible Optic Spectrophotometer). The concentration of the GSTdMT fusion product was determined using an extinction coefficient of $A_{280}^{1\%} = 20$ for GST. The metal content, amino acid sequence, and molecular mass of purified protein was analyzed by MALDI-TOF-MS (TOPLAB Munich, Germany).

Oligomeric states of GSTdMT and dMT were analyzed by size exclusion chromatography. GSTdMT and dMT were dialyzed against several changes of PBS, pH 7.3, or 50 mM Tris-HCl, pH 8.0, and 150 mM NaCl, and about 6–7 ml of protein at 2.5 mg/ml were loaded on HiLoad 26/60 Superdex 75 (Amersham Biosciences) size exclusion column run with the same buffers. Proteins were eluted at 2.5 ml/min, and 1-ml fractions were collected.

Metal Tolerance of Transformed E. coli—Tolerance of transformed and control *E. coli* to metals in the growth medium was tested in the concentration range 0.05 to 0.9 mM CdCl$_2$. Briefly, starter cultures (5 ml) of bacteria (transformed with pGEXdMT and untransformed controls) were grown overnight and used at 1:100 dilution for induction cultures. dMT expression was induced at $A_{600} = 1.0$ by addition of 0.75 mM isopropyl 1-thio-β-D-galactopyranoside with simultaneous addition of metals to the cultures. Bacterial growth was monitored up to 12 h by A_{600} measurements.

Molecular Modeling of dMT—The N-terminal (residues 1–18) and C-terminal (residues 61–75) metal binding domains were reconstructed by homology modeling. dMT metal binding α- and β-domains were aligned with MT sequences obtained from the protein data bank (31). Candidate structures for homology modeling were selected according to pairwise alignment scores and their similarities in cysteine distribution patterns relative to the target sequences. Sea urchin MT β-domain (Protein Data Bank code 1QJL) and rat liver MT β-domain (Protein Data Bank code 2MRT) were used as template structures for homology modeling of dMT α- and β-domains, respectively. Target structures were reconstructed using the DeepView/Swiss-PdbViewer version 3.7 (32). During modeling, cysteine-metal distances in the template structures were kept constant, and other amino acids were placed in the structure accordingly for both domains.

As the protein data bank did not contain any structure with a sequence that would be sufficiently similar to the complete hinge region (residues 19–60), the "3D-PSSM" fold recognition technique (33) and the "ROSETTA/I-site Server" (34) for *ab initio* protein structure determination were used to predict the structure of the hinge region. The folds, with low *E* values obtained from the 3D-PSSM server, were aligned with the dMT hinge region to predict nine structures that were checked for overall stereochemical quality using ERRAT (35). Five more predicted structures were obtained from the ROSETTA/I-site server.

Solution X-ray Scattering and Data Processing—The synchrotron radiation x-ray scattering data were collected on the X33 camera (36, 37) of the European Molecular Biology Laboratory (EMBL) on the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) using multiwire proportional chambers with delay line readout (38). The scattering patterns were recorded at a sample detector distance of 2.4 m covering the range of momentum transfer $0.15 < s < 3.5$ nm^{-1} ($s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 0.15$ nm the x-ray wavelength). Solutions of GSTdMT fusion protein in PBS buffer were measured at several concentrations between 1.6 and 9 mg/ml. Because of aggregation at higher concentrations, only the scattering curves measured at concentrations below 3 mg/ml without reduced glutathione were taken for further structure analysis. Bovine serum albumin was also measured as a molecular mass standard at 7 mg/ml in 50 mM Hepes, pH 8.0, and 150 mM NaCl. To monitor radiation damage during the scattering experiments, data were collected in 15 successive 1-min frames. No radiation damage could be detected either in the x-ray data or in the SDS-PAGE analyses carried out before and after irradiation. Data reduction, background subtraction, and correction for detector response followed standard procedures using the program PRIMUS (39).

The maximum particle dimension $D_{\max}$ was estimated using the orthogonal expansion program ORTOGNOM (40). The forward scattering $I(0)$ and the radius of gyration R_g were evaluated using the Guinier approximation (41), assuming that at very small angles ($s < 1.3/R_g$) the intensity is represented as $I(s) = I(0) \exp(-(sR_g)^{2/3})$. These parameters were also computed from the entire scattering pattern using the indirect transform package GNOM (42), which also provides the distance distribution function $p(r)$ of the particle. The molecular mass of the solute was evaluated by comparison of the forward scattering with that from a reference solution of bovine serum albumin (66 kDa). The accuracy of this method was limited by the uncertainty in the measured protein concentration used for data normalization. The excluded volume of the hydrated particle (Porod volume) was computed without model assumptions and independently of normalization from the shape scattering curve by using Equation 1 (43),

$$V = 2\pi^2 I(0) \bigg/ \int_0^\infty s^2 I(s)\,ds \qquad \text{(Eq. 1)}$$

This parameter yielded an independent estimate of the molecular mass because for globular proteins its value in nm^3 corresponded to about twice the molecular mass in kDa.

Low resolution models of the GSTdMT fusion protein were generated *ab initio* by the programs DAMMIN (44) and GASBOR (45). The program DAMMIN represented the particle as a collection of densely packed beads inside a sphere with diameter $D_{\max}$ where $M \gg 1$. Each bead belongs either to the particle or to the solvent, and the shape is represented by a binary string of length M. Starting from a random string, simulated annealing is employed to search for a compact model that fits the experimental data $I_{\exp}(s)$ to minimize discrepancy as shown in Equation 2,

$$\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{\exp}(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)}\right]^2 \qquad \text{(Eq. 2)}$$

where N is the number of experimental points; c is a scaling factor and $I_{\text{calc}}(s)$ and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer s_j, respectively. Prior to shape analysis by DAMMIN, a constant is subtracted from each data point to force the s^{-4} decay of the intensity at higher angles following the Porod's law (43) for homogeneous particles.

The program GASBOR represents a protein by an assembly of dummy residues and uses simulated annealing to build a locally "chain-compatible" dummy residue model inside the same search volume (45). Dummy residue modeling is able to fit higher resolution data without subtraction of a constant and generally provides more detailed models than those given by the other shape determination algorithms.

The reconstructions were performed without symmetry restrictions and also assuming a point symmetry group P2 as the results suggested that the GSTdMT fusion protein is a dimer in solution. The crystallographic model of dimeric *Schistosoma japonicum* glutathione *S*-transferase (GST) was taken from the protein data bank (31), code 1GTA (46), for molecular modeling. A simulated annealing protocol was employed to construct dMT strands as interconnected chains of dummy residues (47) attached to the C termini of a GST dimer minimizing discrepancy between the scattering curve calculated from the entire
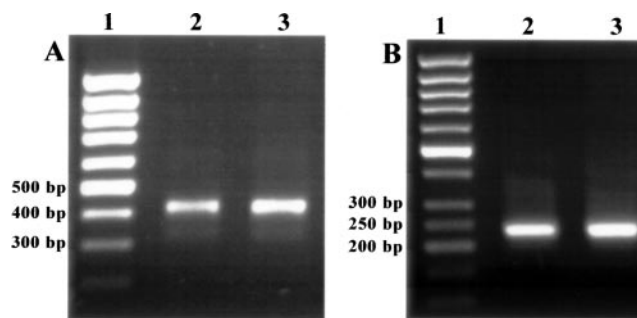
FIG. 1. **Identification of *T. durum mt* gene.** *A*, the genomic *mt* bands resolved by agarose gel electrophoresis were about 450 bp long for *T. durum* (*lane 2*) and *T. aestivum* (*lane 3*). *B*, cDNA bands appeared at positions corresponding 250 bp. DNA molecular mass markers are shown in *lane 1*.

fusion protein and the experimental scattering data. The procedure was constrained by imposing P2 symmetry. An additional attempt at rigid body modeling was performed based on the tentative model of dMT connected to C termini of GST dimer via 12-amino acid long dummy residues linkers. The theoretical scattering pattern $I(s)$ was calculated from the available high resolution coordinates of the portion with known structure and from the portion with unknown structure represented as dummy residues using spherical harmonics as shown in Equation 3,

$$I(s) = 2\pi^2 N^2 \left\{ \sum_{l=0}^\infty \sum_{m=-l}^l \left| \sum_n^l A^n{}_{lm}(s) + \sum_k B^k{}_{lm}(s) \right|^2 \right\} \qquad \text{(Eq. 3)}$$

Here the complex functions $A^n{}_{lm}(s)$ are the partial scattering amplitudes of the subunits in the given positions and orientations that depend on the scattering amplitudes in the reference positions calculated using the program CRYSOL (48) and on three rotational and three translational parameters. $B^k{}_{lm}(s)$ values are the partial amplitudes of dummy residues comprising the unknown parts, calculated using the form factor of a dummy residue (45, 47). The summations run only over symmetry-independent subunits and dummy residues and over m and l values permitted by the selection rules for the spherical harmonics; n is the number of symmetry mates (for P2, m is even, $n = 2$). The algorithm is a particular application of the program BUNCH, which allows quaternary structure modeling of multisubunit proteins and deletion mutants against the SAXS data when only a limited part of high resolution structure of subunits is available.[2] The program searches the optimal positions and orientations of available high resolution models of subunits and the most probable conformations of interconnected chains composed of dummy residues representing the missing fragments, which are attached to the appropriate residues in high resolution structures, to fit the experimental scattering data of the entire construct.

Sequence data from this article have been deposited with the Gen-Bank™ data library under accession numbers AY688468 (*T. aestivum* cv. Bezostaja MT mRNA), AY688469 (*T. durum* cv. Balcali85 MT mRNA), AY688470 (*T. durum* cv. C1252 MT mRNA), AY688471 (*T. aestivum* cv. Bezostaja *mt* gene), and AY688472 (*T. durum* cv. Balcali85 *mt* gene).

RESULTS

T. durum Metallothionein Gene—The presence of an *mt* gene in *T. durum* genome was detected by PCR. The isolated *durum* genome was used as template for the F1 and R1 primers, based on the available *T. aestivum mt* cDNA sequence (GenBank™ accession number L11879) (17). In separate PCRs, *T. aestivum* genomic DNA was used as the positive control template. Reactions resulted in amplification of DNA fragments with about 450 bp (Fig. 1A), and sequencing results indicated that the *mt* genes of *T. durum* (*dmt*) and *T. aestivum* containing 416 and 399 nucleotides, respectively, are highly homologous (Fig. 2A). The major difference appeared to be a repetitive "TTTTA"

A
```
1                                                 50
ATGTCTTGCAACTGTGGATCCGGTTGCAGCTGCGGCTCAGACTGCAAGTG   dmt cv.Balcali
ATGTCTTGCAACTGTGGATCCGGTTGCAGCTGCGGCTCAGACTGCAAGTG   amt cv.Bezostaja


51                                                100
CGGGTATGGATGCTTTTTTTTATTTTTTATTTTTTTGATTGACGTTAATGGA  dmt cv.Balcali
CGGGTATGGATGTTTT------------TTTTTCAAT---CATTAATGGA    amt cv.Bezostaja
           *                    **    *

101                                               150
TGCTTCTCCTTGCAAAAATACTCTCGTGTCTATTTATCTACTACTTCATCT  dmt cv.Balcali
TGCTTCTCCTTGCAAAA-TACTCTCGTGCTATTTATGTACTACTTCATCC   amt cv.Bezostaja
                                  *           *    *

151                                               200
GGTGACCCCTTTAGCTGATCAAGCATCAGTTCCTTTTGTGGTGATGAACT   dmt cv.Balcali
GGTGATCCCTTTAGCTGATCAAGCATCAATTCCTATTGTGGTGATGAACG   amt cv.Bezostaja
     *                      *     *             *

201                                               250
ATTGATCTGATTCGCCGACCGGTTGCCTTCGCCCGTTGCAGGAAGATGTA  dmt cv.Balcali
ATTGATTTGATTCGCCGACCGGTTGCCTTCGCCCGTTGCAGGAAGATGTA  amt cv.Bezostaja
      *

251                                               300
CCCTGATCTGACGGAGCAGGGCAGTGCCGCGGCCCAGGTCGCCGCCGTGG  dmt cv.Balcali
CCCTGATCTGACGGAGCAGGGCAGCGCCGCGCCCAGGTCGCCGCCGTGG   amt cv.Bezostaja
                        *       *

301                                               350
TCGTCCTCGGCGTGGCGCCTGAGAACAAGGCGGGGCAGTTCGAGGTGGCC  dmt cv.Balcali
TCGTCCTCGGCGTGGCGCCTGAGAACAAGGCGGGGCAGTTCGAGGTGGCC  amt cv.Bezostaja

351                                               400
GCCGGCCAGTCCGGCGAGGGCTGCAGCTGCGGCGACAACTGCAAGTGCAA  dmt cv.Balcali
GCCGGGCAGTCCGGCGAGGGCTGCAGCTGCGGCGACAACTGCAAGTGCAA  amt cv.Bezostaja
     *

401       416
CCCCTGCAACTGTTAA                                    dmt cv.Balcali
CCCCTGCAACTGTTAA                                    amt cv.Bezostaja
```

B
```
1                                                 60
ATGTCTTGCAACTGTGGATCCGGTTGCAGCTGCGGCTCAGACTGCAAGTGCGGGAAGATG  amt [L11879]
ATGTCTTGCAACTGTGGATCCGGTTGCAGCTGCGGCTCAGACTGCAAGTGCGGGAAGATG  dmt cv.Balcali
ATGTCTTGCAACTGTGGATCCGGTTGCAGCTGCGGCTCAGACTGCAAGTGCGGGAAGATG  amt cv.Bezostaja

61                                                120
TACCCTGATCTGACGGAGCAGGGCAGTGCCGCGGCCCAGGTCGCCGCCGTGGTCGTCCTC  amt [L11879]
TACCCTGATCTGACGGAGCAGGGCAGTGCCGCGGCCCAGGTCGCCGCCGTGGTCGTCCTC  dmt cv.Balcali
TACCCTGATCTGACGGAGCAGGGCAGCGCCGCGCCCAGGTCGCCGCCGTGGTCGTCCTC   amt cv.Bezostaja
                          *       *

121                                               180
GGCGTGGCGCCTGAGAACAAGGCGGGGCAGTTCGAGGTGGCCGCCGCCGGCCAGTCCGGCGAG  amt [L11879]
GGCGTGGCGCCTGAGAACAAGGCGGGGCAGTTCGAGGTGGCCGCCGCCGGCCAGTCCGGCGAG  dmt cv.Balcali
GGCGTGGCGCCTGAGAACAAGGCGGGGCAGTTCGAGGTGGCCGCCGGGCAGTCCGGCGAG    amt cv.Bezostaja
                                              *

181                                          228
GGCTGCAGCTGCGGCGACAACTGCAAGTGCAACCCCTGCAACTGTTAA  amt [L11879]
GGCTGCAGCTGCGGCGACAACTGCAAGTGCAACCCCTGCAACTGTTAA  dmt cv.Balcali
GGCTGCAGCTGCGGCGACAACTGCAAGTGCAACCCCTGCAACTGTTAA  amt cv.Bezostaja
```
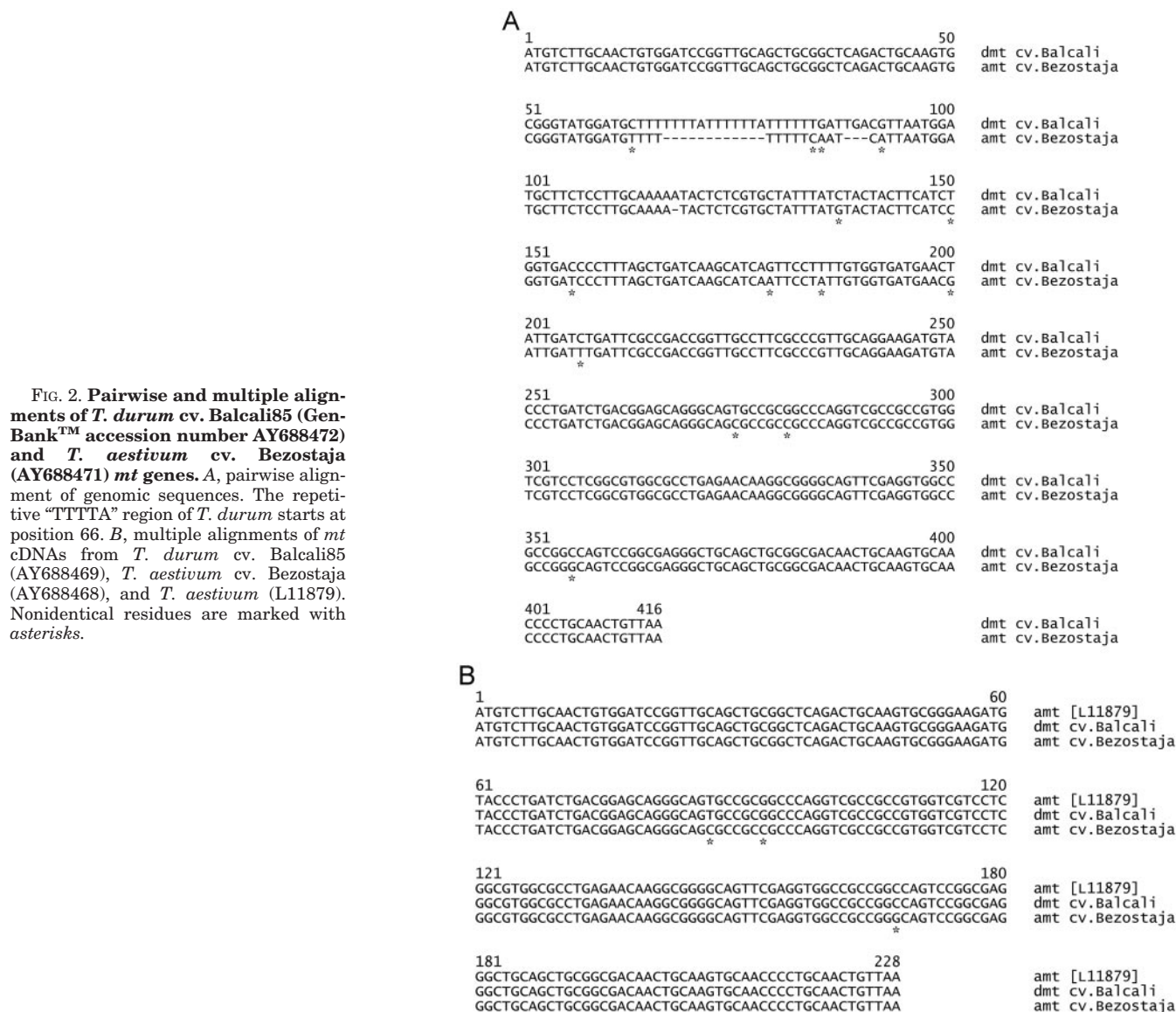
FIG. 2. **Pairwise and multiple alignments of *T. durum* cv. Balcali85 (Gen-Bank™ accession number AY688472) and *T. aestivum* cv. Bezostaja (AY688471) *mt* genes.** *A,* pairwise alignment of genomic sequences. The repetitive "TTTTA" region of *T. durum* starts at position 66. *B,* multiple alignments of *mt* cDNAs from *T. durum* cv. Balcali85 (AY688469), *T. aestivum* cv. Bezostaja (AY688468), and *T. aestivum* (L11879). Nonidentical residues are marked with *asterisks.*

sequence in *dmt,* which was later shown to be in the intron region. A comparison of the *dmt* and *amt* sequences with that of maize (the only other *mt* sequence available (49)) revealed about 67% sequence similarity (data not shown). Wheat *mt* gene products were analyzed by total mRNA isolation and *in vitro* synthesis of *mt* cDNA by RT-PCR, again using F1 and R1 primers. cDNA bands, about 235 bp long, were detected (Fig. 1*B*). Alignment of the cDNA sequences indicated that wheat *mt* genes have coding sequences of 53 bases at the 5′-end and 175 bases at the 3′-end with sequences bordering the introns following the GT/AG rule (Fig. 2*B*) (50). This 2 exons/1 intron structure is typical for type 1 plant MTs (31). The translated sequences for MTs from *T. durum* and *T. aestivum* (dMT and aMT, respectively) were 100% identical because of the degeneracy of the genetic code.

The wheat MT protein contains 75 amino acids of which 12 are cysteines (Fig. 3). These residues are clustered in groups of six in the N and C termini and are distributed with the pattern of C*X*C, where *X* is any amino acid other than cysteine. The two cysteine-containing domains are separated by a 42-residue-long hinge region assumed to be between residues 19 and 61, which is devoid of cysteines. These results further support that wheat MT is a type 1 MT (14). The translated dMT sequence was aligned with those of putative type 1 MT-like proteins from

wheat (17), maize (49), the heavy metal tolerant plant *Festuca rubra* (51), barley (52), and rice (53). Results (Fig. 3) show that the highest homology is observed with wheat (97%) followed by *Festuca rubra* (73%). Alignment of amino acid sequences of typical type 2 plant MTs are included to illustrate differences in the cysteine distributions and similarities in the hinge regions (Fig. 3).

Expression of GSTdMT and Metal Tolerance in E. coli— Effect of expression of the fusion protein GSTdMT on metal tolerance of *E. coli* BL21(DE3) was investigated by supplementing the growth medium with varying concentrations of $CdCl_2$.

Growth curves indicated that addition of $CdCl_2$ up to 0.05 mM had no significant effect, and in cells expressing the fusion protein, only GST and untransformed controls reached the same steady state level of about $3.0 \pm 0.5 \times 10^9$ cells (assuming $A_{600} = 1$ to correspond to about 8×10^8 cells) after 7 h. Differences in cadmium tolerance were observed at concentrations of 0.3 mM $CdCl_2$ and above. At 0.3 mM $CdCl_2$ (Fig. 4), control cell counts at 7 h were about 50% those with GSTdMT. At 0.6 mM $CdCl_2$, there was no significant growth in control cultures, whereas bacteria expressing GSTdMT reached about 50% of their normal level in 7 h, and inclusion of $CdCl_2$ at 0.9 mM was lethal for both types of cells (data not shown).
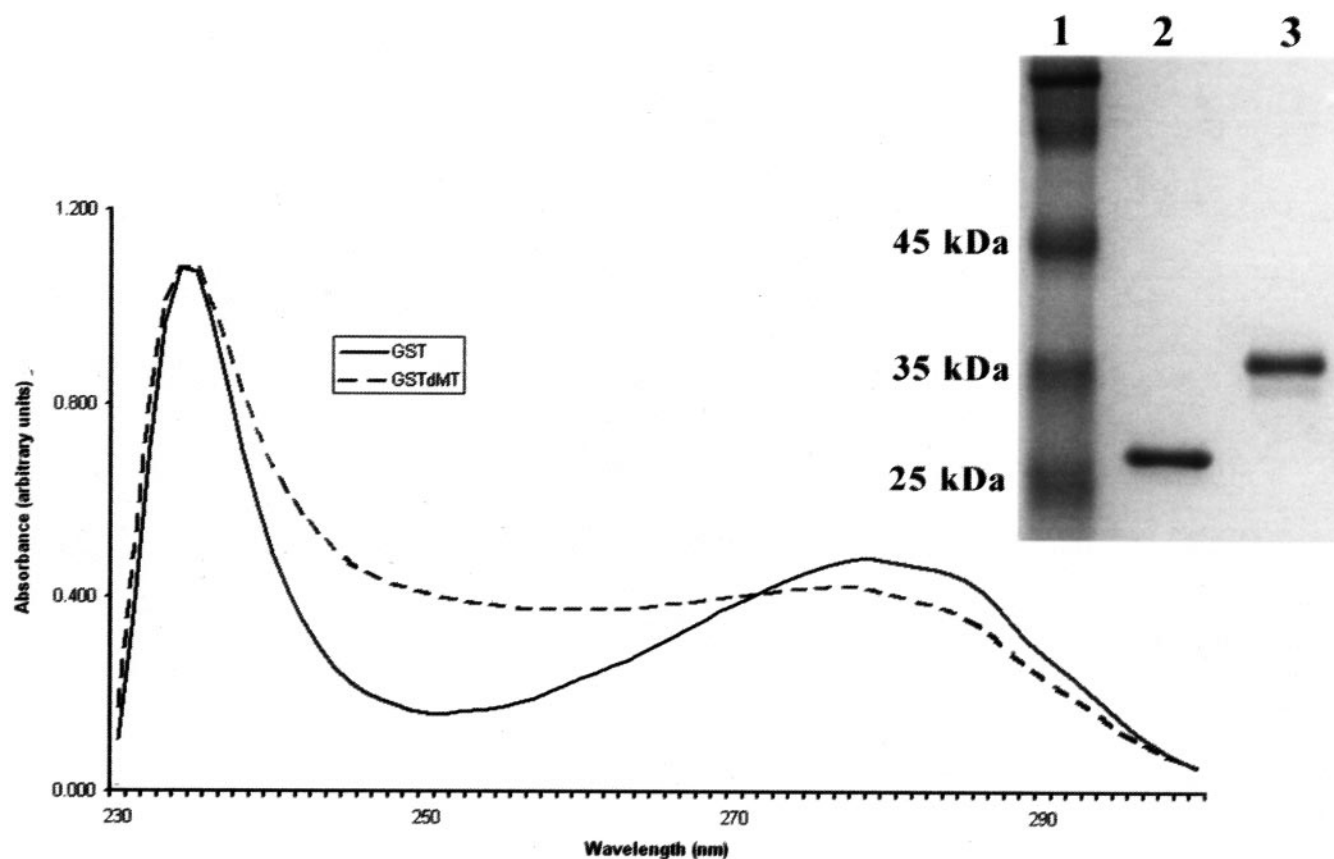
```
Type 1 Plant MTs
---------------

     !  !    !  !    !  !                                              !  !    !  !  !  !
MSCSCGSSCGCGSNCKCGKMYPDLDEQASTTTQ--AVVVVGVAHENKAG-QFEMASGE------GCKCGANCKCDPCNC   f. rubra[AAB70464]
MSCSCGSSCGCGSSCKCGKKYPDLEET-STAAQ--PTVVLGVAPEKKAAPEFVEAAAESGEAAHGCSCGSGCKCDPCNC  maize[AAB19992]
MSCSCGSSCSCGSNCSCGKKYPDLEEK-SSSTK--ATVVLGVAPEKKA--QQFEAAAESGETAHGCSCGSSCRCNPCNC  rice[S57768]
MSCSCGSSSCGCGSNCNCGKMYPDLEEK-SGATMQVTVIVLGVG---SAKVQFEEAA-EFGEAAHGCSCGANCKCNPCNC barley[SMBH1]
MSCNCGSGCSCGSDCKCGKMYPDLTEQGSAAAQVAAVVVLGVAPENKAG-QFEVAAGQSGE---GCSCGDNCKCNPCNC  wheat[AAA50846]
*** ***.* *** * *** **** * * :   ..:*:**.   . *  :   *: :    **.**  *.*:*:****

MSCNCGSGCSCGSDCKCGKMYPDLTEQGSAAAQVAAVVVLGVAPENKAG-QFEVAAGQSGE---GCSCGDNCKCNPCNC  wheat[durum]


Type 2 plant MTs
----------------

     !  !   !  !   !  !   !  !                                    !  !   !  !  !  !
MSCCGGNCGCGSSCKCGNGCGGCKMYPDLSYTESTTTETLVMGVAPVKAQFEGAEMGVPAENDGCKCGPNCSCNPCTCK  soybean[JQ2128]
MSCCGGSCGCGSGCKCGSGCGGCGMYPDL---ESTTTFTIIEGVAPMKN-YGVAEKATEGGN-GCKCGSNCTCDPCNC-  tomato[Q43513]
MSCCGGKCGCGSSCSCGSGCGGCGMYPDLSYSEMTTTETLIVGVAPQKTYFEGSEMGVAAEN-GCKCGSDCKCDPCTCK  kiwi[S48038]
******.***** *.** ***** *****   * *** *:: **** *   :  :*  .. .  * *****.:*.*:**.*
```

FIG. 3. **Alignment of *T. durum* MT amino acid sequence with those of type 1 and type 2 plant MT sequences.** The main difference between type 1 and type 2 plant MTs is the presence of two adjacent cysteine residues (*!*) as 3rd and 4th residues in the type 2 sequence. Note the homology in the hinge region. ("*", identical residues; ":", highly conserved; and ".", weakly conserved groups).

FIG. 4. **Cadmium tolerance in *E. coli* expressing GSTdMT.** Comparison of the growth curves at 0.3 mM cadmium concentration reveals that *E. coli* BL21(DE3) expressing GSTdMT (▲) have higher tolerance to the presence of the metal than GST (■)-expressing cells.

These results demonstrated that GSTdMT could bind cadmium *in vivo*. As the presence of a metal ion facilitated correct folding and stabilized the recombinant fusion protein (12, 54), 0.05 mM cadmium was included in *E. coli* cultures grown for protein isolation.

Purification and Characterization of Recombinant GSTdMT and dMT—GSTdMT purified on glutathione-Sepharose affinity columns was about 90% homogeneous as determined from SDS-PAGE analysis and migrated at a position corresponding to 35 kDa compared with 25 kDa for GST (Fig. 5). The absorption spectrum of GSTdMT displays the characteristic charge transfer band between 240 and 260 nm, due to bound cadmium (Fig. 5). The recombinant protein yield was between 20 and 30 mg/liter culture, which is about half the amount obtained with GST alone.

Native PAGE analysis of recombinant GSTdMT showed the presence of high molecular mass species in the purified protein solutions (Fig. 6*A, inset*), and as these would interfere with x-ray solution scattering measurements, the protein was fractionated by size exclusion chromatography to monitor the dif-

ferent states of oligomerization. Assuming a molecular mass of about 34.5 kDa for GSTdMT, two peaks at positions corresponding to trimers and dimers were observed, and the separation was verified by native PAGE analyses (Fig. 6*A, inset I, lanes a* and *c*). SDS-PAGE patterns of these fractions were almost identical (Fig. 6*A, inset II*), and Western blots (Fig. 6*A, inset III*) verified that all observed bands corresponded to oligomers of GSTdMT. Analysis of peptide mass fingerprints by MALDI-TOF-MS after tryptic digestion verified the sequence of the recombinant protein. The molecular mass of GSTdMT, determined by MALDI-TOF-MS linear mode, was found to be about 34.42 kDa, and based on the difference between this value and the molecular mass calculated from the amino acid sequence (about 34 kDa), cadmium content was calculated to be 4 ± 1 cadmium atoms per protein molecule. SDS- and native-PAGE and Western blots showed that a lower molecular mass species appeared 1–2 days following purification of GSTdMT. Analyses of these samples by mass spectrometry revealed fragments with a molecular mass of 29.4 kDa and indicated cleavage of the recombinant protein at sites within dMT. The

FIG. 5. **Characterization of purified recombinant GSTdMT.** The absorption spectrum of GSTdMT (- - -) shows the charge transfer band between 240 and 260 nm due to bound cadmium that is lacking in GST spectrum (——). *Inset* shows SDS-PAGE analysis of purified recombinant proteins: GST (*lane 2*), GSTdMT (*lane 3*), and molecular mass marker (*lane 1*).

cleaved C-terminal fragments started with sequences VAAV and/or GQFE, which are within the designated hinge region.

We also attempted to isolate dMT by thrombin cleavage from the fusion protein, and we fractionated the purified protein for structural studies. Size exclusion chromatography of isolated and cleaved dMT yielded a major peak eluting at a position corresponding to ~23 kDa, which represents the dMT trimer (Fig. 6*B*). Even under strict conditions for minimizing air contact, isolated dMT appeared to form a wider range of oligomers than GSTdMT (Fig. 6*B*, *inset*). Because of this anomalous behavior, solution x-ray scattering data were collected only from GSTdMT preparations.

X-ray Solution Scattering Measurements on GSTdMT—Preliminary structural characterization of recombinant GSTdMT and dMT was carried out by solution x-ray scattering. The processed scattering pattern from the solution of GSTdMT fusion protein (Fig. 7) and the overall structural parameters (Table I) were used in molecular shape determinations. The estimated molecular mass of the protein (83 kDa) obtained from the ratio of the forward scattering and the protein concentration was somewhat larger than calculated for a GSTdMT dimer (~68.8 kDa). In contrast, the excluded volume of the hydrated particle or Porod volume, which is independent of protein concentration (43), corresponded to a molecular mass that was somewhat smaller than that expected for a GSTdMT dimer (Table I). Taken together these independent results suggested that the GSTdMT fusion protein in solution at concentrations below 3 mg/ml is a dimer.

Shape Determination—Several *ab initio* low resolution shapes of GSTdMT were restored by programs DAMMIN and GASBOR without symmetry constraints (P1) and, given that the fusion protein is probably dimeric in solution, also with P2

symmetry. Although all models had elongated shapes, they did not correspond to a unique solution; the models represent a globule with either one broad tail or two thinner tails (Fig. 8). The second type of shapes was typically generated by GASBOR in P2 and by DAMMIN in P2 with an additional requirement for the long axis to be perpendicular to the symmetry axis. This assumption was made for DAMMIN reconstructions only to impose an anisometry of the models compatible with that of the dimeric GST. Without this assumption, some DAMMIN reconstructions in P2 had only a single protuberance (Fig. 8, *upper left*). All models were compatible with the crystallographic dimer of GST (they could accommodate it within the globule) and provided good fits to the experimental data (Fig. 7). One or two thinner tails in the *ab initio* models could be considered as two contacting or separated dMT parts, respectively.

Rigid Body Modeling—To reduce the ambiguity and improve the resolution of the *ab initio* shape determination, further modeling was performed utilizing the high resolution crystal structure of the GST dimer. Two symmetric tails of 87 amino acids representing dMT strands with linkers were modeled at the C termini of the GST dimer to fit the experimental scattering data of GSTdMT fusion protein by the program BUNCH. Several reconstructions yielded virtually the same conformations of the added tails sticking out in opposite directions (Fig. 9) and provided good fits to the experimental data with $\chi = 0.76$ (Fig. 7). Other attempts at rigid body refinement based on the tentative model of dMT and those in P1 yielded somewhat worse fits (not shown) but preserved the overall shape with the main globule composed by two GST monomers and two noncontacting dMT tails (Fig. 9). Rigid body modeling against solution scattering therefore favors a noncontacting arrangement of dMTs in the GSTdMT fusion protein.
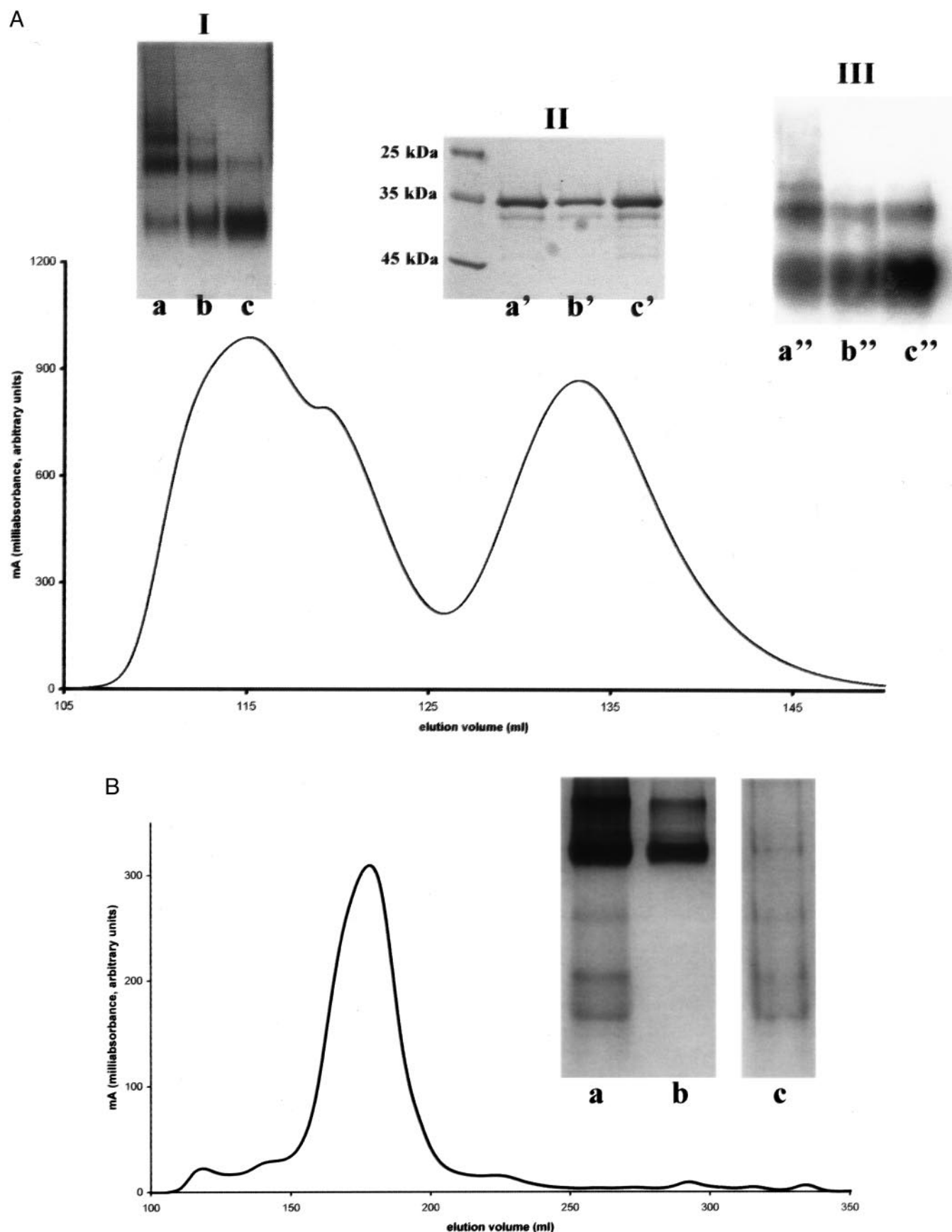
FIG. 6. **Oligomers of purified GSTdMT and dMT.** *A*, elution profile of GSTdMT from HiLoad 26/60 Superdex 75 column. *Insets*, lanes *a*, *a'*, *a''*, and lanes *c*, *c'*, and *c''* correspond to trimeric and dimeric structures eluting in the first and second peaks, respectively. The shoulder at the back of the first peak is analyzed in *lanes b*, *b'*, and *b''*. The native PAGE (*inset I*) and its Western blot (*inset III*) show the oligomeric structures. SDS-PAGE (*inset II*) and Western blot analyses verified that all bands originated from GSTdMT. *B*, elution profile of dMT from HiLoad 26/60 Superdex 75 column. The major peak corresponds to a molecular mass of 23.1 kDa indicating that the protein eluted from the column as a trimer. Results of native gel analyses of a thrombin-treated GSTdMT solution (*lane a*), isolated GST (*lane b*), and oligomers of dMT eluting in the major peak (*lane c*) are shown in the *inset*.

Structural Characterization of T. durum Metallothionein
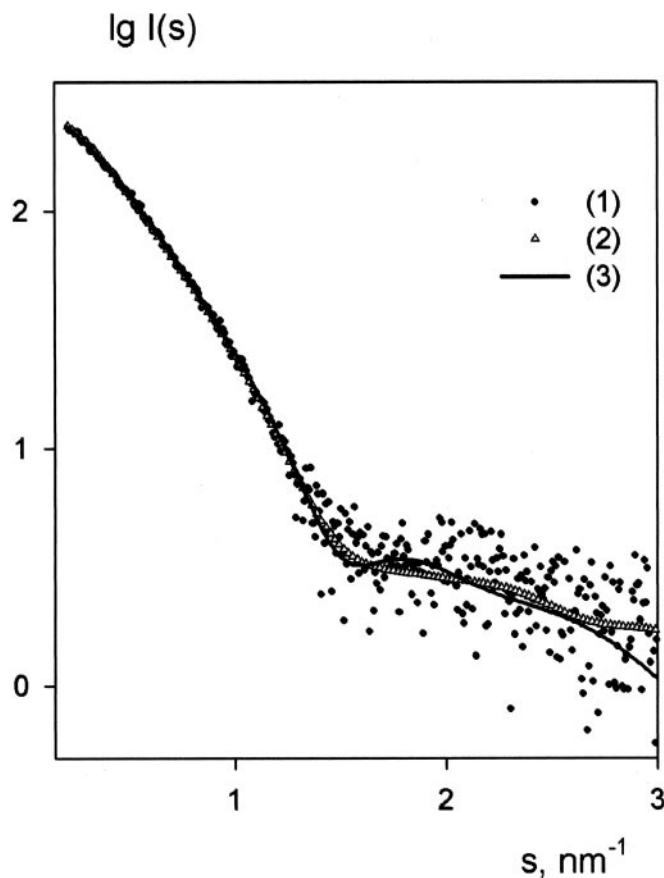
Fig. 7. Experimental and calculated scattering patterns from GSTdMT fusion protein. *1*, experimental data; *2*, scattering from the typical *ab initio* model; *3*, scattering from the best rigid body model with $\chi_S = 0.77$.

TABLE I
Summary of structural parameters of GSTdMT fusion protein computed from the scattering data

$R_g{}^a$	$D_{\max}{}^b$	MM^c	$MM_{\mathrm{mon}}{}^d$	$V_p{}^e$	$\chi_S{}^f$	χ^g
nm	*nm*	*kDa*	*kDa*	*nm³*		
3.2 ± 0.1	120	83 ± 10	34	110 ± 10	0.77	0.76

[a] Radius of gyration.

[b] Maximum particle dimension.

[c] Molecular mass.

[d] MM_{mon} is the MM of the monomer of GSTdMT fusion protein computed from the primary structure.

[e] Excluded Porod volume calculated from the scattering data.

[f] Discrepancy between the experimental data and computed scattering curves for the *ab initio* shape determination.

[g] Discrepancy between the experimental data and computed scattering curves for the rigid body model.

Molecular Modeling of dMT—dMT structure was modeled by considering three regions in the protein. The structures of the two metal binding domains were predicted using homology modeling, and *ab initio* approaches were used for the hinge region.

BLAST (55) sequence alignments for the metal binding domains gave hits mostly with MTs due to the similar distribution of cysteines (data not shown). Structures of the two metal-binding clusters were modeled separately, and sea urchin MT β-domain (Protein Data Bank code 1QJL_A) and rat liver MT β-domain (Protein Data Bank code 2MRT) were used as templates in the prediction of the structures of the dMT α- and β-domains, respectively (Fig. 10).

Standard programs such as PROCHECK and ERRAT could not be used for verification of the atomic coordinates of the domain models because the distances involved in the metal-
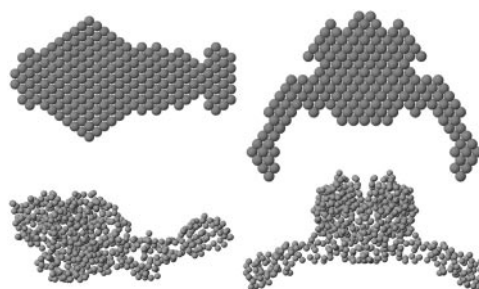

Fig. 8. Non-uniqueness of *ab initio* shape determination of GSTdMT fusion protein. *Ab initio* models with one protuberance (*left panel*) and two protuberances (*right panel*). Models calculated with programs DAMMIN (*top*) and GASBOR (*bottom*) are shown.

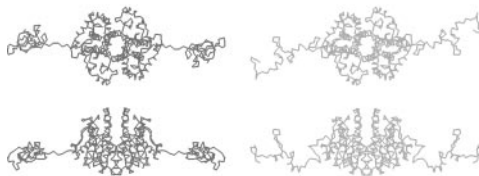**Fig. 9. Rigid body models of GSTdMT fusion protein (displayed as Cα traces).** *Left panel*, GST dimer with two 87-amino acid tails added. *Right panel*, GST dimer with two dMTs modeled as described in the text in symmetry related positions. In all panels, the bottom row is rotated clockwise by 90° around horizontal (*x*) axis.

thiolate clusters are not properly taken into account by these algorithms. Consequently, unsatisfactory results were also obtained when these algorithms are used with experimentally determined structures such as rat MT (56) and yeast MT (57). The surface electrostatic potential and molecular surface calculations gave highly charged surfaces in regions where residues Gly-8, Ser-10, and Ser-13 in the α-domain and Asn-7, Lys-9, and Asn-11 in the β-domain appeared to be surface-accessible after the metals were introduced into the model. The surface charge distribution in the rest of the model was neutral.

The structure of the hinge region selected on the basis of the 3D-PSSM and ROSETTA/I-site results and verified using ERRAT (Fig. 11) gave a Ramachandran plot (not shown) indicating that most residues were in low energy regions. Predictions clustered in two main groups that had similar main features but differed in β-content. Prediction 1, which corresponded to the best prediction of the ROSETTA and the third best hit given by the 3D-PSSM (Fig. 11), was chosen for the hinge region of the final model for the structure of dMT. This model of the hinge region represented a more flexible and exposed structure, in accordance with observations that the hinge regions of wheat MTs are readily available for proteolytic attack (see above). This structure corresponds to a fold that is very similar to that of DNA-binding proteins (PSSM founder code clfp0a) and may point to a functional similarity based on the three-dimensional structure.

The three separately modeled regions were combined considering the total energy of the protein and using "prediction 1" (Fig. 11) to represent the structure of the hinge region. The final three-dimensional model (Fig. 12), one of the possible structures in agreement with the experimental data, gives a dumbbell-shaped molecule with the two metal binding domains emerging at opposite poles. As expected, metal centers appeared as protected globular structures, whereas the hinge region appears to be exposed to solution.

DISCUSSION

MTs are cysteine-rich metal-binding proteins found in all eukaryotes and many prokaryotes and appear to be involved in metal homeostasis, metal detoxification, and responses to oxida-

Fig. 10. **Homology modeling of dMT α- and β-domains.** Target structures sea urchin β-domain (Protein Data Bank code 1QJL_A) and rat liver β-domain (Protein Data Bank code 2MRT) are shown together with the resultant models for dMT-α- and -β-domains. Cysteine side chains (*sticks*) coordinating the cadmium atoms (*red spheres*) are shown. Pairwise sequence alignments for each domain indicate the similarities in the amino acid sequence.

Fig. 11. **Structure prediction of dMT hinge region.** Three predicted structures and their ERRAT structure verification scores are shown on the *left panel*. Two structures on the *right panel* are examples for the DNA-binding "winged-helix superfamily" proteins. This family includes mostly repressors and/or transcription activators.

tive stress (3). Mutations of MT genes in mammalian systems aiming at determining the function of the proteins have often failed to induce strong sensitivity toward metals and were inconclusive (58). Similar functional assignments were also extended to plant MTs, but attempts to confirm such roles *in situ* were ambiguous for most plant MTs and pointed to expression of *mt* genes under the influence of several other stress factors besides metals (59, 60). Expression of plant MTs in bacterial systems showed some protection against metal toxicity but with variable specificity. In plants, MT gene expression appears to be involved in developmental regulation, correlated with high concentrations

of metals in the environment, and with numerous stress responses (15). It is not clear, however, whether the protein takes part solely in metal homeostasis mechanisms or is expressed as a general stress-response protein.

To probe the involvement of MTs in metal binding in wheat, we studied one *mt* gene from *T. durum* and compared it with that from *T. aestivum*. We characterized the GSTdMT expressed in *E. coli* for determination of structural parameters and developed structural models to probe the function. The *dmt* gene was searched in the *T. durum* genome using the *T. aestivum* sequence (17).

FIG. 12. **The predicted structure of dMT.** Cadmium-binding (*blue spheres*) metal centers at each pole of the dumbbell-shaped molecule are depicted in *ball* and *stick* representation with the extended hinge region highlighted in ribbon representation.

Genomic sequences of *dmt* and *amt* show a high level of identity except for a 10-bp AT-rich region in the intron part. A detailed comparison of *dmt* and *amt* genomic sequences with other plant species could be carried out only with maize (49) because of the lack of data in the literature. This comparison indicates a high degree of homology with the main differences occurring in the intron region (data not shown). The cDNA sequences of *amt* and *dmt* are 99% identical with each other and 96% identical with L11879, leading to identical protein sequences when third codon degeneracy is taken into account. Assignment of one intron and two exon regions after alignment of the cDNA sequences indicated that durum MT is a type 1 plant MT. This was further supported by the location and motif of cysteine residues in the translated protein sequence. The assignment of the two metal binding domains from the cDNA sequence also highlighted an unusually long hinge region, which is mostly seen in plants but not in other organisms.

Expression of GSTdMT Enhances Cadmium Tolerance in E. coli—The recombinant durum MT expressed in *E. coli* as a C-terminal GST fusion (GSTdMT) affected cadmium tolerance and resulted in 4-fold increase in the threshold for toxicity (Fig. 4). There are no previous experimental results reported on altered heavy metal tolerance of bacteria expressing plant MTs. In an earlier study, Evans *et al.* (25) had expressed a pea *mt* gene, *PsMT*$_A$, in *E. coli* as a GST fusion and studied the effect of expression on cadmium, zinc, and copper accumulation. They observed an 8-fold higher copper accumulation when compared with controls, but no significant effect on cadmium or zinc accumulation. These results imply a strong interaction of PsMT$_A$ with copper but not with cadmium and appear to be contradictory to our observations. This may be due to different mechanisms of metal accumulation and detoxification in the different modified *E. coli* strains used for expression; *E. coli* JM101 was used by Evans *et al.* (25), whereas BL21(DE3) was used in this study.

Characterization of GSTdMT and dMT—MTs are highly sensitive to oxidation during lengthy isolation procedures, and the long hinge region of plant MTs appears to be a target for proteolytic attack (24).

Direct isolation of MT has only been reported for wheat Ec protein (13) and *A. thaliana* MTs (41). Ec protein was isolated from wheat embryos as a 6-kDa zinc-binding protein and contained about 5 mol of zinc/mol of protein. Abundant accumulation of Ec mRNA in immature wheat embryos suggested that

Ec protein might function as a zinc storage protein in the early embryo and play a role in zinc homeostasis during embryogenesis (16). Fractionation of extracts from *A. thaliana* grown in nutrient solution, which contained additional 30 μM CuCl$_2$, resulted in isolation of low molecular mass copper-binding and cysteine-rich proteins designated as MT1, MT2, and MT3 with corresponding molecular masses of 8, 7, and 4.5 kDa. The copper/protein molar ratios were 7.29:1 for MT1, 5.5:1 for MT2, and 8.45:1 for MT3. It was reported that localization and induction of expression of MT1 and MT2 correlated with exposure to copper, suggesting their involvement in copper homeostasis (61).

Recombinant production of MTs helps to circumvent some of the problems associated with direct isolation, and expression as a GST fusion offers simple possibilities for purification, quantification, and detection. Kille *et al.* (62) had expressed the pea *mt* gene *PsMT*$_A$ in *E. coli,* and the cadmium/protein ratio was found to be 6.1:1 for protein isolated from bacteria grown in medium supplemented with of 0.3 mM cadmium. The purified protein was proteolytically cleaved from the exposed hinge region but was held together through interactions of cysteine-rich regions with metals. Tommey *et al.* (24) had also expressed PsMT$_A$ in *E. coli* as a GST fusion and investigated metal-protein interactions. Bacteria were grown in media supplemented with zinc, cadmium, or copper and for isolated protein strength of interactions appeared to decrease in the order copper, cadmium, and zinc, and metal content was found to be 6.1 mol of copper, 5.7 mol of cadmium, and 11.5 mol of zinc/mol of protein.

Expression of *T. durum* MT in *E. coli* resulted in purification of the fusion protein with a 34.4-kDa molecular mass. The dMT component that has an expected molecular mass of 7.4 kDa was found to contain 4 $\pm$ 1 cadmium atoms/mol of protein as determined using mass spectroscopy. This result is in agreement with those reported for MT2 directly isolated from *A. thaliana* and for recombinant PsMT$_A$, both of which show high sequence homology (over 80%) with dMT. Because *A. thaliana* results correspond to the isolated native protein, about 5 cadmium/mol of protein may be indicative of correct folding of the recombinant species. Results of analyses showed that cleavage that was observed with pea MT also occurred with dMT after 1–2 days of storage following purification.

When purified to homogeneity, the recombinant GSTdMT formed stable oligomers that were fractionated to dimers and trimers. A strong tendency to form oligomers was also observed with dMT that was cleaved off GST and purified separately. It was not possible to obtain monodisperse fractions of these oligomers at the relatively high protein concentrations (>3 mg/ml) required to obtain reliable solution scattering patterns. In contrast, the increase in scattering power resulting from the fusion with GST allowed us to obtain reliable scattering data from GSTdMT solutions at lower effective molar dMT concentrations.

Structure Prediction, X-ray Scattering Measurements, and Possible Functional Assignments—Proteins of the MT family bind metals in thiolate clusters in one or two structural domains, and these metal clusters have globular folds similar to those observed in rat liver MT (Protein Data Bank code 4MT2), yeast MT (Protein Data Bank code 1AQR), or sea urchin MT (Protein Data Bank code 1QJL). Most two-domain MTs (*e.g.* mammalian and crustacean) have a short (up to 10 amino acids) hinge region connecting the metal-containing clusters. The likely metal binding domains of dMT were identified by pairwise sequence alignment with known MT structures and were modeled with a high level of confidence using straightforward homology modeling (Fig. 10). Designation of the metal binding domains yielded a long connecting sequence between

the two regions. The structure of the hinge consisting of about 42, mainly hydrophobic, amino acids could be predicted only using *ab initio* approaches. The resulting structures could be grouped in two categories as follows: the first one characterized by large loop regions, and the second one by a high β-sheet content giving rise to more compact rigid structures. The criterion for selecting the candidate structures for the hinge region from the ROSETTA and 3D-PSSM prediction pool was the presence of similar folds in the results from both servers (Fig. 11).

X-ray patterns indicated that the protein dimerizes at GST sites, and the dMTs can be modeled as extended structures that are not in contact with each other (Fig. 9). This suggests, in agreement with the predicted structure (Fig. 12), that at low concentrations dMT is indeed an elongated monomer with two domains separated by an extended hinge region.

The strong tendency for purified dMT to form oligomers is indicated by the anomalous behavior of the isolated protein on size exclusion columns, by observation of multiple bands on native gels, and through detection of high molecular mass species in x-ray solution scattering measurements. Additionally, the fact that dMT hinge region is readily available for proteolytic attacks implicates the extended structure between the two metals centers. The cleavage sites at Gln (33) and Ala (48) appeared to be at the end of the first and at the beginning of the second predicted helical regions, respectively. These experimental observations support results of model calculations indicating that the hinge region has a structure that could promote formation of oligomers and may be a likely spot for protein/DNA or protein-protein interactions. The predicted structure for the hinge region shows an intriguing similarity with a fold family that includes DNA-binding proteins.

The long hinge region observed in *T. durum* is conserved among distinct plant species with some identical amino acids. Conservation of the sequence of the hinge region (Fig. 2*C*), which probably has no role in metal binding, may indicate involvement of such metallothioneins in mechanisms other than metal detoxification. Metal binding seems to be the result of an intrinsic property of MTs because of their cysteine-rich motifs but not necessarily their primary role *in situ*. The structural modeling also suggests an alternative/additional role that may involve interactions with DNA. Convergence of experimental results and modeling give confidence in the proposed structure, but the functional prediction clearly requires further work. DNA binding assays, RNA interference experiments, investigation of the upstream regions of *dmt* gene, and computational analyses of the phylogeny of the plant MTs are being carried out in order to determine the primary function of plant MTs.

REFERENCES

1. Coyle, P., Philcox, J. C., Carey, L. C., and Rofe, A. M. (2002) *Cell. Mol. Life Sci.* **59,** 627–647
2. Vallee, B. L. (1991) *Methods Enzymol.* **205,** 3–7
3. Vasak, M., and Hasler, D. W. (2000) *Curr. Opin. Chem. Biol.* **4,** 177–183
4. Kagi, J. H. (1991) *Methods Enzymol.* **205,** 613–626
5. Andrews, G. K. (2000) *Biochem. Pharmacol.* **59,** 95–104
6. Brouwer, M., Syring, R., and Hoexum, B. T. (2002) *J. Inorg. Biochem.* **88,** 228–239
7. Ebadi, M., Leuschen, M. P., el Refaey, H., Hamada, F. M., and Rojas, P. (1996) *Neurochem. Int.* **29,** 159–166
8. Vallee, B. L. (1995) *Neurochem. Int.* **27,** 23–33
9. Klaassen, C. D. (ed) (1999) *Metallothionein IV,* Birkhauser Verlag, Basel
10. Binz, P. A., and Kagi, H. R. (1997) in *Metallothionein IV* (Klaassen, C. D., ed) pp. 7–21, Birkhauser Verlag, Basel
11. Wright, C. F., McKenney, K., Hamer, D. H., Byrd, J., and Winge, D. R. (1987) *J. Biol. Chem.* **262,** 12912–12919
12. Sayers, Z., Brouillon, P., Svergun, D. I., Zielenkiewicz, P., and Koch, M. H. J. (1999) *Eur. J. Biochem.* **262,** 858–865
13. Lane, B., Kajioka, R., and Kennedy, T. (1987) *Biochem. Cell Biol.* **65,** 1001–1005
14. Rauser, W. E. (1999) *Cell Biochem. Biophys.* **31,** 19–48
15. Cobbett, C., and Goldsbrough, P. (2002) *Annu. Rev. Plant Biol.* **53,** 159–182
16. Kawashima, I., Kennedy, T. D., Chino, M., and Lane, B. G. (1992) *Eur. J. Biochem.* **209,** 971–976
17. Snowden, K. C., and Gardner, R. C. (1993) *Plant Physiol.* **103,** 855–861
18. de Miranda, J. R., Thomas, M. A., Thurman, D. A., and Tomsett, A. B. (1990) *FEBS Lett.* **260,** 277–280
19. Snowden, K. C., Richards, K. D., and Gardner, R. C. (1995) *Plant Physiol.* **107,** 341–348
20. Choi, D., Kim, H. M., Yun, H. K., Park, J. A., Kim, W. T., and Bok, S. H. (1996) *Plant Physiol.* **112,** 353–359
21. Butt, A., Mousley, C., Morris, K., Beynon, J., Can, C., Holub, E., Greenberg, J. T., and Buchanan-Wollaston, V. (1998) *Plant J.* **16,** 209–221
22. Dunaeva, M., and Adamska, I. (2001) *Eur. J. Biochem.* **268,** 5521–5529
23. Robinson, N. J., Wilson, J. R., and Turner, J. S. (1996) *Plant Mol. Biol.* **30,** 1169–1179
24. Tommey, A. M., Shi, J., Lindsay, W. P., Urwin, P. E., and Robinson, N. J. (1991) *FEBS Lett.* **292,** 48–52
25. Evans, K. M., Gatehouse, J. A., Lindsay, W. P., Shi, J., Tommey, A. M., and Robinson, N. J. (1992) *Plant Mol. Biol.* **20,** 1019–1028
26. Meyers, M. W., Fricke, F. L., Holmgren, G. G., Kubota, S. J., and Chaney, R. L. (1982) *Agronomy Abstracts*, The American Society of Agronomy, Madison, WI
27. Cakmak, I., and Welch, R. M. (2000) *J. Exp. Bot.* **51,** 221–226
28. Hart, J. J., Welch, R. M., Norvell, W. A., and Kochian, L. V. (2002) *Physiol. Plant.* **116,** 73–78
29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
30. Laemmli, U. K. (1970) *Nature* **227,** 680–685
31. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235–242
32. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714–2723
33. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. E. (2000) *J. Mol. Biol.* **299,** 499–520
34. Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. (1997) *J. Mol. Biol.* **268,** 209–225
35. Colovos, C., and Yeates, T. O. (1993) *Protein Sci.* **2,** 1511–1519
36. Koch, M. H. J., and Bordas, J. (1983) *Nucl. Instrum. Methods A* **208,** 461–469
37. Boulin, C. J., Kempf, R., Gabriel, A., and Koch, M. H. J. (1988) *Nucl. Instrum. Methods A* **269,** 312–320
38. Gabriel, A., and Dauvergne, F. (1982) *Nucl. Instrum. Methods A* **201,** 203–204
39. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) *J. Appl. Crystallogr.* **36,** 1277–1282
40. Svergun, D. I. (1993) *J. Appl. Crystallogr.* **26,** 258–267
41. Guinier, A. (1939) *Ann. Physique* **12,** 161–237
42. Svergun, D. I. (1992) *J. Appl. Crystallogr.* **25,** 495–503
43. Porod, G. (1982) in *Small-angle X-ray Scattering* (Glatter, O., and Kratky, O., eds) pp. 17–51, Academic Press, London
44. Svergun, D. I. (1999) *Biophys. J.* **76,** 2879–2886
45. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) *Biophys. J.* **80,** 2946–2953
46. McTigue, M. A., Williams, D. R., and Tainer, J. A. (1995) *J. Mol. Biol.* **246,** 21–27
47. Petoukhov, M. V., Eady, N. A. J., Brown, K. A., and Svergun, D. I. (2002) *Biophys. J.* **83,** 3113–3125
48. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) *J. Appl. Crystallogr.* **28,** 768–773
49. de Framond, A. J. (1991) *FEBS Lett.* **290,** 103–106
50. Mount, S. M. (1996) *Science* **271,** 1690–1692
51. Ma, M., Lau, P. S., Jia, Y. T., Tsang, W. K., Lam, S. K. S., Tam, N. F. Y., and Wong, Y. S. (2003) *Plant Science* **164,** 51–60
52. Okumura, N., Nishizawa, N. K., Umehara, Y., Ohata, T., Nakanishi, H., Yamaguchi, T., Chino, M., and Mori, S. (1994) *Plant Mol. Biol.* **25,** 705–719
53. Hsieh, H. M., Liu, W. K., and Huang, P. C. (1995) *Plant Mol. Biol.* **28,** 381–389
54. Yu, W. H., Cai, B., Gao, Y., Xie, Y., and Huang, Z. X. (2002) *J. Protein Chem.* **21,** 177–185
55. Altschul, S. F., Thomas, L. M., Alejandro, A. S., Jinghui, Z., Zheng, Z., Webb, M., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
56. Braun, W., Vasak, M., Robbins, A. H., Stout, C. D., Wagner, G., Kägi, J. H. R., and Wütrich, K. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 10124–10128
57. Peterson, C. W., Narula, S. S., and Armitage, I. (1996) *FEBS Lett.* **379,** 85–93
58. Palmiter, R. D. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 8428–8430
59. Kim, S., Ahn, I. P., and Lee, Y. H. (2001) *Mol. Plant-Microbe Interact.* **14,** 1340–1346
60. Hsieh, H. M., Liu, W. K., Chang, A., and Huang, P. C. (1996) *Plant Mol. Biol.* **32,** 525–529
61. Murphy, A., Zhou, J., Goldsbrough, P. B., and Taiz, L. (1997) *Plant Physiol.* **113,** 1293–1301
62. Kille, P., Winge, D. R., Harwood, J. L., and Kay, J. (1991) *FEBS Lett.* **295,** 171–175