Identification of Two Novel Dictyostelium discoideum Cysteine Proteinases That Carry N-Acetylglucosamine-1-P Modification*

Glaucia M. Souza†, John Hirai, Darshini P. Mehta, and Hudson H. Freeze§

From the La Jolla Cancer Research Foundation, La Jolla, California 92037

Dictyostelium discoideum makes multiple developmentally regulated lysosomal cysteine proteinases. One of these, a lysosomal enzyme called proteinase I, contains a cluster of GlcNAc-1-P-Ser residues. We call this phosphoglycosylation. To study its function, a cDNA library from vegetative cells was screened, and two novel cysteine proteinase clones were characterized (cprD and cprE). Each of them has highly conserved regions expected for cysteine proteinases, but unlike any other, each has a serine-rich domain containing three distinct motifs, poly-S, SGSG, and SGSG. cprD and cprE cDNAs were overexpressed in Dictyostelium and the active enzymes identified. cprD codes for a protein of approximately 36 kDa (CP4), which is recognized by monoclonal antibodies against GlcNAc-1-P and fucose. cprE corresponds to a 29-kDa protein, which is recognized by antibodies against GlcNAc-1-P. mRNA for both enzymes is present in the vegetative phase and increases during growth on bacteria but decreases throughout development. When the formation of the fruiting body is complete the mRNA for both messages is detected again but in very low levels. Having cloned cDNAs for proteins that carry GlcNAc-1-P should allow us to probe the function of the carbohydrate in these putative lysosomal enzymes.

Dictyostelium discoideum is an eukaryotic amoeba that grows as single cells, but when the bacterial food source is removed, the cells initiate a complex multicellular developmental program. Cells aggregate and differentiate into six different types and, in the end, 85% of them are converted into spores setting atop a cellular stalk (1). We are interested in studying the role of carbohydrate modifications in this organism (2). One of these is the addition of GlcNAc-1-P to serine residues, which has been well documented to occur on a cysteine proteinase called proteinase I found in vegetative cells (3–5). Although antibodies against GlcNAc-1-P recognize various proteins in the cells and in secretions of cells grown in axenic medium (6) the identity of these proteins is unknown. To study the function of GlcNAc-1-P on a defined protein, we decided to clone members of the cysteine proteinase family expressed in vegetative cells.

Previous studies in Dictyostelium identified two developmentally regulated members of this gene family, cprA (CP1) and cprB (CP2) (6–8), but none have been identified in vegetative cells. Since cysteine proteinases are highly conserved in all eukaryotes, we used the active site consensus sequence of cysteine proteinases and the cprA and cprB cDNAs to clone two novel vegetative cysteine proteinases, cprD and cprE. They have the predicted conserved regions but also have an unusual serine-rich domain not previously found in any known cysteine proteinase that could be the site of GlcNAc-1-P addition. The cDNA clones were overexpressed and the active enzymes were shown to have GlcNAc-1-P.

EXPERIMENTAL PROCEDURES

Materials—Radiolabeled oligonucleotides were purchased from DuPont NEN and ICN Biomedicals, Inc. The random primed labeling kit and quick spin Sephadex G-25 and G-50 columns were from Boehringer Mannheim. The messenger RNA isolation kit and the R408 interference-resistant helper phage was from Stratagene Inc., La Jolla, CA. The Sequenase DNA sequencing kit was from U. S. Biochemical Corp. The monoclonal antibody 835 against fucose was a kind gift from Chr. Christopher West (University of Florida College of Medicine). Goat anti-mouse antibody conjugated to alkaline phosphatase was from Promega. H-o-Val-Leu-Lys-p-nitroanilide was from Chromogenix. Genetecin (G418) was from Life Technologies, Inc. Restriction and modifying enzymes were from New England Biochemicals and Boehringer Mannheim. Nitrocellulose filters and prefilters for development were from Millipore. All other chemicals were from Sigma.

Cell Culture and Development Conditions—D. discoideum strains AX-4 and AX-2 were grown axenically in HL-5 or with Kdsibiæa aerogenes (9). After 50 h of growth, the bacteria is consumed and the plate appears clear. This is referred to as the clearing plate stage and corresponds to late vegetative growth. Development of 5 × 10⁶ cells from axenic cultures or from clearing plates was done on 47-mm black nitrocellulose filters resting on prefilters saturated with 20 mM phosphate buffer, pH 6.4, and took 24 h to complete.

cDNA Library Construction and Screening—Poly(A)⁺ mRNA was isolated from 5 × 10⁸ AX-4 cells from clearing plates. The cells were collected only from plates where no morphological development was visually observed. Since the first few hours of development occur without obvious morphological changes, the library could also have very early development cDNAs represented. Cells were washed twice in cold water treated with diethyl pyrocarbonate, and poly(A)⁺ mRNA was prepared using the messenger RNA isolation kit. A cDNA custom λ phage library was prepared by Stratagene in the EcoRI/XhoI site of Uni-ZAP XR vector. For screening, the library was plated on AX-2 cells, blotted into nitrocellulose filters, and hybridized with the different probes following the conditions advised by the manufacturer. 1 × 10⁶ phage plaques were screened using a 23-base mixed oligonucleotide corresponding to 8 conserved amino acids (IKNOQCGG) of cysteine proteinases as follows: 5’-AT(C/A)AA(A/G)TA(C/T)CA(G/A)GG(T/C/A/G)TG(T/C/A/G)CA(A/G)TG(T/C)CGG-3’ and 3’-TA(A/G)TT(T/C)CT(T/G)TA(G/A)TG(T/C)-

* This work was supported by National Institute of General Medical Sciences Grant 32485. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Supported by a postdoctoral fellowship from CNPq (Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, Brazil).

‡ To whom correspondence should be addressed: La Jolla Cancer Research Foundation, 10901 N. Torrey Pines Rd., La Jolla, CA, 92037. Tel.: 619-455-6480; Fax: 619-450-2101; E-mail: Hudson@jcrf.edu.

1 G. Souza and H. Freeze, unpublished results.

2 The abbreviations used are: p-NA, p-nitroaniline; bp, base pair(s); kb, kilobase(s); SSPE, saline/sodiumphosphate/EDTA; PAGE, polyacrylamide gel electrophoresis; N-Cbz, N-benzyloxycarbonyl; ON-p, p-nitrophenyl ester; DTT, dithiothreitol; E-64, trans-epoxysuccinyl-L-leucylamido(4-guanidino)butane; t-Boc, t-butoxycarbonyl; MCA, 7-amido-4-methylcoumarin.
CC(A/G/T)/CGT/T(AC/A)GC-5. The oligonucleotide mixture was radiolabeled using T4 polynucleotide kinase and [α-32P]ATP and purified by centrifugation through columns of Sephadex G-25. Three-tenths of 100 positive clones were reprocessed using a 775-bp BglII/SpeI fragment of CPR (from bases 340 to 1018) and a 405-bp HindIII fragment of CPR (from bases 338 to 743) as probes. These fragments exclude the 5′ end of the cDNAs, which correspond to unconserved regions of the precursor protein. The cDNA fragments were radiolabeled by random primed synthesis with [α-32P]ATP and [α-32P]dCTP and purified by centrifugation through columns of Sephadex G-50. Six positive clones were identified and sequenced. To sequence the pBluescript plasmid phage vector was used for sequencing. The full-length preparations were performed (10), and the double-stranded DNA was sequenced using the Sequenase DNA sequencing kit. Three clones corresponded to the same cDNA, and one of them was a full length of 1.1 kb (cprE). The other three clones corresponded to another cDNA, but none had an initiating methionine. Three different probes derived from the largest cDNA of this second group (1.4 kb) were then used to isolate a full-length clone: the 1.4-kb fragment itself, a 373-bp NvaI/StuI fragment corresponding to a serine-rich region not found in cprE, and a 206-bp EcoRI fragment corresponding to the 5′ end. 1 × 106 plaques were screened and 1% of these were positives. A polymerase chain reaction of phage particles from 200 plaques was then performed (10) using as primers the first 18 bases of the 1.4 kb fragment (5′-AGCGAAAAAATTAAAA-3′) and the T3 primer (5′-AATTAC- CCAATTAAACG-3′). Larger polymerase chain reaction products were rescreened with the 1.4-kb probe and sequenced, and all of them corresponded to a full-sized cDNA (cprD).

DNA Preparation and Southern Blot—Dictyostelium DNA mini-preparations were performed as described (11). 50 μg of DNA were digested with restriction enzymes according to the manufacturer’s instructions, electrophoresed in 0.8% agarose gels, transferred into nitrocellulose filters, and prehybridized for 2 hours at 65°C in 5 × Denhardt’s solution, 4 × SSPE, 0.1% SDS, 50 μg/ml salmon sperm DNA (10). The filters were hybridized overnight at 55°C in the presence of 107 cpm of each probe (the complete cDNA of cprA, cprB, cprD, and cprE) and washed 3 times for 20 min with 1 × SSPE, 1% SDS and 3 times for 10 min with 0.1% SDS at room temperature. The filters were then dried and exposed to x-ray films (Kodak X-Omat) overnight.

Amino Acid Sequence Alignments—Amino acid sequence alignments were done using the GENEX WORKS program (Intelligenetics Inc.). ACTININ tertiary structure was displayed using Xfit Program (12).

Subcloning and Transformation Procedures—The 1.4-kb cDNA of CPR or the 1.3-kb cDNA of CPR E was subcloned into the BamHI/XhoI and BamHI/KpnI sites, respectively, of pCRII/Bluescript, which allows the expression under control of the actin 6 promoter. Cloning and DNA preparations followed standard procedures as described (10). Stable transformants overexpressing CPRD and CPR E were obtained using the calcium phosphate precipitate method (11). 2 × 107 cells of AX-4 (for CPRD) or AX-2 (for CPR E) were transfected with a calcium phosphate precipitate of plasmid DNA (20–23), plant cysteine proteinases (actinidin and papain) (24, 25), and two cysteine proteinases from Dictyostelium (CP1 and CP2) show that CPRD and CPR E have all the expected conserved regions and critical active site residues. The first 17 amino acids of both enzymes probably constitute a signal peptide (pre-region) since it is consistent with other typical signal sequences (26). Both CPRD and CPR E contain the pro-region (amino acids 18–111) in the N-terminal domain, which is not found in the mature form of known cysteine proteinases. Two potential N-linked glycosylation sites are found in both CPRD and CPR E.

An unusual feature of these deduced amino acid sequences is the presence of a serine-rich domain near the C terminus of both proteins. In CPRD it is 115 amino acids long and contains 60 serine residues (52%), while the same region contains 12 serine residues out of 24 amino acids (50%). Another Dictyostelium cysteine proteinase, CPR B, also contains an insert in this region (42 amino acids long), but its serine content is only 11%. Other cysteine proteinases typically have much shorter sequences (1–12 amino acids) in this region (Fig. 2). In CPRD, the serine residues seem to be distributed in three distinct repeated motifs: poly-S, SGSG, and SGSG. Serines in the insert from CPR B seem to follow the same pattern but with fewer repeats. The tertiary structures of cysteine proteinases actinidin, papain, and the human liver cathepsin B are known (27–29).

The similarity in the conserved regions of CPRD and CPR E to these cysteine proteinases suggests that they may have the same potential N-linked glycosylation sites are found in both CPRD and CPR E.

Southern Blot Analysis of cprD and cprE—The nucleotide sequence of CPRD and CPR E is depicted in Fig. 1, and from here the enzymes are referred to as cysteine proteinase 4 (CP4) and cysteine proteinase 5 (CP5), respectively. Sequence alignment of CP4 and CP5 (Fig. 2) with human cathepsins (B, H, L, and S) (20–23), plant cysteine proteinases (actinidin and papain) (24, 25), and two cysteine proteinases from Dictyostelium (CP1 and CP2) show that CPRD and CPR E have all the expected conserved regions and critical active site residues. The first 17 amino acids of both enzymes probably constitute a signal peptide (pre-region) since it is consistent with other typical signal sequences (26). Both CPRD and CPR E contain the pro-region (amino acids 18–111) in the N-terminal domain, which is not found in the mature form of known cysteine proteinases. Two potential N-linked glycosylation sites are found in both CPRD and CPR E.

An unusual feature of these deduced amino acid sequences is the presence of a serine-rich domain near the C terminus of both proteins. In CPRD it is 115 amino acids long and contains 60 serine residues (52%), while the same region contains 12 serine residues out of 24 amino acids (50%). Another Dictyostelium cysteine proteinase, CPR B, also contains an insert in this region (42 amino acids long), but its serine content is only 11%. Other cysteine proteinases typically have much shorter sequences (1–12 amino acids) in this region (Fig. 2). In CPRD, the serine residues seem to be distributed in three distinct repeated motifs: poly-S, SGSG, and SGSG. Serines in the insert from CPR B seem to follow the same pattern but with fewer repeats. The tertiary structures of cysteine proteinases actinidin, papain, and the human liver cathepsin B are known (27–29).

The similarity in the conserved regions of CPRD and CPR E to these cysteine proteinases suggests that they may have the same overall structure. Fig. 3 shows the tertiary structure of actinidin and the location of the serine-rich insert of CPRD and CPR E based on the inferred amino acid sequence homology and crystal structures. The insertion occurs at Gly-170 (actinidin), and in CPRD it comprises nearly one-third of the predicted size of the mature protein. As seen in Fig. 3, the insert lies on the opposite side of the protein away from the active site.

Southern Blot Analysis of cprD and cprE—Southern blots of Dictyostelium DNA samples digested with different enzymes were probed with the CPR A, CPR B, CPR D, and CPR E cDNAs and washed at high stringency (Fig. 4).

The results confirm that the
Fig. 1. Nucleotide and deduced amino acid sequences of cprD and cprE. Panels A and B correspond to cprD and cprE, respectively. The start of the polyadenylation signal, AATAAA, is underlined. The putative N terminus of the mature proteinase is boxed. Asterisks signify termination. The amino acids are indicated by the single letter code.
cDNA clones described here are different from \textit{cprA} and \textit{cprB} and that they are from \textit{Dictyostelium} and not from their bacterial food source. Some cross-reactivity occurs because there are common sequences among \textit{cprA}, \textit{cprB}, \textit{cprD}, and \textit{cprE}. This is shown by the alignment of faint bands of \textit{cprA} and \textit{cprB} when blots are probed with \textit{cprD} and \textit{cprE}. A clear example of this is the dark 5.3-kb band in the EcoRI/HindIII digestion of the blot probed with \textit{cprB}. The blots probed with the other cDNAs show faint 5.3-kb bands.

Analysis of \textit{cprD} and \textit{cprE} mRNA Levels during Growth and Development—Expression of the mRNA corresponding to \textit{cprD} and \textit{cprE} was analyzed in cells grown in axenic cultures (HL-5) or with bacteria (KA). AX-2 cells exponentially growing in HL-5 were plated on SM agar plates along with bacteria and collected after 44, 47, and 50 h of growth. These times correspond to log growth, beginning of clearing, and total clearing plates, respectively. Cells were then collected at 50 h and plated for synchronous development to analyze mRNA expression (Fig. 5A). Alternatively, cells were plated on filters directly from axenic cultures (Fig. 5C). As seen, mRNA levels corresponding to \textit{cprD} and \textit{cprE} increase during growth on bacteria up to clearing plates and decrease once development starts. Densitometer scanning of the autoradiograms show that the mRNA levels increase from 2- to 4-fold between 44 and 50 h of growth in bacteria. Cells prepared from axenic culture also display a decrease in mRNA expression during development. As seen in Fig. 5C, the mRNAs are not detected after 4 h of development until the formation of the fruiting body is complete, when very low levels of mRNA for both messages are again detected. The same pattern of expression is observed when mRNAs from AX-4 strain are analyzed (not shown).

Overexpression of \textit{cprD} and \textit{cprE} in \textit{D. discoideum}—We decided to overexpress \textit{cprD} and \textit{cprE} in \textit{Dictyostelium} to investigate whether it codes for active cysteine proteinases modified by GlcNAc-1-P. The cDNAs were subcloned into pDNeo67 (13), an expression vector that contains the G418 resistance marker and where the cDNA is under the control of the \textit{Dictyostelium} actin 6 promoter. Transformants were isolated and analyzed for mRNA expression. As shown in Fig. 5B, transformed clones (CP4–6, CP4–25, and CP5–12) overexpress the mRNAs 5–10-fold, compared with control cells. Cells transfected with the pDNeo67 plasmid alone show the same pattern as non-transfected control cells.

Identification of the Overexpressed Proteins in the Transformed Cells—Cell lysates from control cells grown in HL-5 or on bacteria and from cDNA clones overexpressing \textit{cprD} (CP4–25 and
CP4–6) or cprE (CP5–12) were analyzed by Western blots using monoclonal antibodies against GlcNAc-1-P or fucose (30) and detected by a secondary antibody-alkaline phosphatase conjugate. As seen in Fig. 6, a protein band of approximately 36 kDa is enriched in the transformed cells (CP4–25 and CP4–6) as detected with antibodies against GlcNAc-1-P and fucose. A very faint 36-kDa band can also be seen in the control cells and in a preparation of a 38-kDa cysteine proteinase purified from cells grown on bacteria that is known to carry GlcNAc-1-P. It appears that CP4 in control cells is expressed in low amounts relative to the 38-kDa protein, which co-migrates with proteinase I (3–5). In cells overexpressing cprE (CP5–12) a 29-kDa band is increased as detected by the antibody against GlcNAc-1-P, but no increase was found with the antibody against fucose (not shown).

Cysteine Proteinase Activity Assays on Transformed Cells—To investigate if cells overexpressing CP4 and CP5 had an increased cysteine proteinase activity, total cell lysates of control or transformed cells grown in HL-5 were assayed for activity as described under “Experimental Procedures” using N-Cbz-L-Lys-ON-p or H-D-Val-Leu-Lys-p-NA as substrates. As seen in Fig. 7, cells that overexpress CP4 and CP5 have 2.5–3.6 times more cysteine proteinase activity with the substrate H-D-Val-Leu-Lys-p-NA and 50% more activity with the substrate N-Cbz-L-Lys-ON-p. In the presence of E-64, a specific cysteine proteinase inhibitor, activity is reduced to 20% of control. As a control, activity levels of two lysosomal enzymes, α-D-mannosidase and β-D-glucosidase, were analyzed using 4-methylumbelliferyl substrates, and no difference in activity was detected in the transformed cells when compared with control cells (not shown). To verify that the increase in cysteine proteinase activity was due to the overexpressed 36- and 29-kDa proteins, we performed activity assays of the proteins separated by SDS-PAGE using the fluorogenic substrate N-t-Boc-Val-Leu-Lys-7-MCA. Fig. 8 shows that the 36-kDa band has low activity levels in AX-4 cells, which increases in the two CP4 transformants analyzed, and that it is inhibited by E-64. Also, a faint band of 29 kDa is observed in control AX-2 cells and is increased in the CP5 transformants.

**DISCUSSION**

North and Cotter (31) have described cysteine protease activities in Dictyostelium throughout development and point out the complex and dynamic activity patterns seen in vegetative cells (31). A series of 4–5 different cysteine proteinase activity bands with apparent M, of 30–54 kDa is expressed depending upon whether cells are grown on bacteria or in axenic media (32, 33). Gustafson and co-workers (3, 4) reported a vegetative stage cysteine proteinase of 38 kDa, proteinase I, that contained up to 20% by weight GlcNAc-1-P linked to serine resi-
Novel Cysteine Proteinases

![Southern blots of cprA, cprB, cprD, and cprE](image)

**Fig. 4. Southern blots of cprA, cprB, cprD, and cprE.** Genomic DNA digested with BamHI, BglII, ClaI, and EcoRI/HindIII was electrophoresed in agarose gels and blotted into nylon. The blot was probed at high stringency (55°C) to the entire cDNAs of cprD and cprE. The same filters were reprobed with cprA and cprB. The molecular weight markers are indicated in kb.

The amounts of mRNA of cprD and cprE in developing Dictyostelium cultures were determined by Northern blot analysis. Total RNA was isolated from AX-2, AX-4, CP4–25, CP4–6, and CP5–12 axenic cultures. cprA and cprE were probed using specific cDNA probes corresponding to cprD and cprE. cprA, cprB, cprD, and cprE were cloned in Dictyostelium, but their relative expression closely resembles that of other typical eukaryotic cysteine proteinases. A partial sequence for another developmentally regulated cysteine proteinase (CP3) has also been identified; however, it does not encode a full-sized enzyme.

We are interested in studying the function of GlcNAc-1-P and the signals needed for its addition to proteins. Based on previous studies, we screened a vegetative cell cDNA library to look for typical cysteine proteinase genes, cprA (CP1) and cprB (CP2) (6, 7), that would have serine-rich regions (Fig. 1). We found two such cDNAs that could code for cysteine proteinases, cprD and cprE. mRNA for both genes is detected during vegetative growth and decreases with the start of development, reappearing in low levels when the fruiting body is formed (Fig. 5). This is in agreement with the observation that general cysteine proteinase activity slightly increases at the end of development (35). A surprising feature is that the amount of mRNA increases substantially at the end of vegetative growth. This is typical of the prestavation responsive genes (36) and occurs in parallel with a burst of cysteine proteinase activity seen at this time (35). The reason for this is unclear, but this may reflect an increased need for digesting bacteria or for increased protein turnover known to occur in development. When the cells start development, the protease may not be necessary and its mRNA levels decrease. This is consistent with the decrease seen in cysteine proteinase activity during development (37, 38). Southern blot analysis of cprA, cprB, cprD, and cprE shows that they are located in different genomic DNA fragments (Fig. 4). This was confirmed by mapping the genes in the Dictyostelium genome using yeast artificial chromosomes (YACs). cprD maps to chromosome 3 and cprE maps to the middle of chromosome 2 (39).

CP4 and CP5 have an unusual domain not present in the other previously studied cysteine proteinases. CP4 contains a 115-amino acid domain composed of 52% serine residues divided into three separate contiguous motifs, poly-S, SGSG, and SGSG. CP5 contains similar motifs within a 24-amino acid domain. The serine stretches probably evolved from a series of tandem duplications. CP5 appears to be the older version of the motifs before the onset of tandem duplications. The serine-rich inserts in both CP4 and CP5 appear to be located in a non-conserved region of other cysteine proteinases (Fig. 2). Although they are near the active site histidine residue, their location in space is expected to be away from the active site as shown in the three-dimensional structure on Fig. 3. It is possible, though, that the presence of the insert or of the putative carbohydrate chains may influence the activity of the enzymes, since the serine-rich domain is connected directly to the β-strand involved in the active site. This domain may serve special needs for CP4 and CP5 but is obviously not vital for activity since most cysteine proteinases are devoid of it.

To show that cprD and cprE code for an enzyme that can carry GlcNAc-1-P, the cDNAs were overexpressed in axenically growing cells. This resulted in an average 3-fold increase in cysteine proteinase activity (Fig. 7), which corresponded to an increased activity band of 36 kDa (in CP4 transformants) or 29 kDa (in CP5 transformants) on SDS-PAGE (Fig. 8). Significantly, a monoclonal antibody against GlcNAc-1-P recognizes the same band in the transformants that is found in very low amounts in non-transformed cells (Fig. 6). CP4 transformants also show some additional bands ranging from 45 to 70 kDa, which are detected in the Western blots but not in the activity...
gels. We are currently unable to explain this effect, but they could possibly represent unprocessed forms of the enzyme due to its overexpression. The results also show that this antibody recognizes GlcNAc-1-P in the 38-kDa protease I purified from bacterially grown cells (Fig. 6) (5). Axenically grown cells also have a 38-kDa protein, but it seems to migrate at a slightly higher molecular weight, both in control and transfected cells. North and co-workers (32, 33) have shown that 38-kDa cysteine proteinases are present in cells grown axenically or in the presence of bacteria but that they have different biochemical properties. Different cysteine protease activity patterns are observed in vegetative cells depending on the nutrient availability. These interconversions may be due to differences in post-translational modifications (31–33), and the anomalous migration of the 38-kDa protein in the transformants may reflect altered glycosylation of this protein when CP4 is overexpressed. Resolving these issues will require additional experiments, but it is clear that CP4 (36 kDa) and protease I (38 kDa) are different. This was confirmed by partial amino acid sequencing of protease I (35), although both proteins showed similar amino acid compositions. It is possible that both are members of a closely related family modified by GlcNAc-1-P, since they at least partially co-purify. Both proteins also contain fucose, as shown by binding of another monoclonal antibody (Fig. 6). The location of the fucose residues is unknown, and they either seem to be absent in CP5 or the expression levels were not high enough to permit detection with this antibody even when the blots were overdeveloped. Based on the processing of other cysteine proteinases, the expected masses of CP4 and CP5 without any modifications would be 32,816 and 24,459 Da, respectively. Two potential N-linked glycosylation sites occur in both enzymes, but it is not known if they are actually used. CP4 migrates in SDS-PAGE as an approximate
36-kDa band and CP5 as a 29-kDa band, but additional experiments will be necessary to determine how much of this mass is contributed by either N-linked chains or by GlcNAc-1-P and fucose.

GlcNAc-1-P is most probably added to CP4 and CP5 in this newly characterized serine-rich domain. It is interesting to note that this domain has three distinct motifs, polyserine, SGSG, and SGSG. An enzyme activity that transfers GlcNAc-1-P to core proteins such as serylglcins (42, 43). Polyserine repeats are used as sites for the addition of glycosaminoglycan chains to core proteins such as serylglcins (42, 43). Polyserine repeats have recently been described in a secreted acid phosphatase from Leishmania, which is modified by a new class of phosphoserine-linked glycans, Man-1-P bound to serines (44, 45). The enzyme activity. Since these mutant cDNAs can be expressed by fucosylated and phosphorylated (49). It may be that these presporo vesicle proteins, some putative lysosomal proteins like CP4 and CP5, and other proteins yet to be identified share a property influenced or controlled by GlcNAc-1-P and/or fucose.

The cloning of these two novel cysteine proteinases will allow us to begin to determine the function of GlcNAc-1-P. By characterizing the sites of GlcNAc-1-P addition and creating mutations in these sites we may understand its potential role in targeting these cysteine proteinases to lysosomes or in affecting enzyme activity. Since these mutant cDNAs can be expressed in Dictyostelium, we can study the fate of the protein with altered glycosylation.

Acknowledgments—We thank Dr. A. Kuspa for the physical mapping of cprD and cprE into Dictyostelium chromosomes, Dr. J. Williams for the cprA and cprB clones, Dr. C. West for the antibody against fucose, and Dr. S. Kudo and Dr. M. Fukuda for technical advice and use of equipment in early phases of this work. Liying Wang provided excellent technical assistance. We are also grateful to Dr. R. Doolittle for advice on the analysis of the cysteine proteinase sequences.

References

1. Loomis, W. F. (1982) The Development of Dictyostelium discoides, Academic Press, New York.

2. Freeze, H. H. (1991) in Cell Surface Carbohydrates and Cell Development. (Fukuda, M., ed) pp. 285-317, CRC Press, Boca Raton, FL.

3. Gustafson, G. L., and Milner, A. L. (1979) Biochim. Biophys. Acta 667, 667-673

4. Gustafson, G. L., and Milner, A. L. (1980) J. Biol. Chem. 255, 7208-7210

5. Gustafson, G. L., and Thon, L. A. (1979) J. Biol. Chem. 254, 12471-12478

3 C. West and H. Freeze, unpublished results.