MOLECULAR AND CELLULAR BIOLOGY, Aug. 1988, p. 3458–3466
0270-7306/88/083458-09$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 8, No. 8

Structural Characterization of *Dictyostelium discoideum* Prespore-Specific Gene D19 and of Its Product, Cell Surface Glycoprotein PsA

A. E. EARLY,[1] J. G. WILLIAMS,[1]* H. E. MEYER,[2] S. B. POR,[3]† E. SMITH,[3]
K. L. WILLIAMS,[3] AND A. A. GOOLEY[3]

Imperial Cancer Research Fund, Clare Hall Laboratories, Blanche Lane, South Mimms, Potters Bar, Hertfordshire, EN6 3LD, United Kingdom[1]; Medizinische Fakultat, Institut fur Physiologische Chemie, Ruhr-Universitat Bochum, D-4630 Bochum, Federal Republic of Germany[2]; and School of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia[3]

Received 8 February 1988/Accepted 20 May 1988

The *Dictyostelium discoideum* cell surface antigen PsA is a glycoprotein which first appears in the multicellular stage soon after tip formation and is selectively expressed on prespore cells. The D19 gene encodes an mRNA sequence which is highly enriched in prespore over prestalk cells in the slug stage. We have determined 81 amino acid residues of N-terminal sequence from immunoaffinity-purified PsA protein and shown this sequence to be identical to the predicted sequence of the D19 gene. There are several short repeat elements close to the C terminus, and unequal crossing-over within these is proposed to account for the size polymorphism observed in PsA protein isolated from different *D. discoideum* strains. The repeats are proline rich and show similarity to the C-terminal region of the *D. discoideum* cell adhesion molecule, contact sites A. The extreme C terminus, which is also homologous to contact sites A, is characteristic of proteins attached to the plasma membrane via a glycosyl-phosphatidylinositol link. We have marked the PsA gene by insertion of an oligonucleotide encoding an epitope of the human c-*myc* protein. A construct containing this gene and 990 base pairs of 5′-flanking region directed correct temporal and spatial mRNA accumulation. We found the marked PsA protein, detected with the human c-*myc* antibody, to be correctly localized on the surface of cells.

The identification of cell-type-specific markers is of central importance in the analysis of cellular differentiation and pattern formation. The ideal marker should be exclusively restricted to the cell lineage which it characterizes, and, if it is to be used to analyze the initial events in cellular differentiation, it should be expressed early in the ontogeny of the cell lineage. The anterior one-fifth of the migratory slug of *Dictyostelium discoideum* contains stalk cell precursors, while the posterior four-fifths consists predominantly of prespore cells. Genes, which are expressed in a cell-type-specific manner during development, are readily isolated (1, 17, 27), because prestalk- and prespore-cell-enriched populations can be separated by microdissection or by density gradient centrifugation of disaggregated slugs (33, 42). The D19 cDNA clone derives from a highly prespore-enriched mRNA obtained from density gradient-separated cells (1). The D19 gene has been defined as a class 2 prespore sequence on the basis of the time of appearance of its mRNA during development and its dependence on extracellular cyclic AMP for continued expression (7). It has frequently been used as a prespore-specific marker in studies investigating the requirements for cell-type-specific gene expression and cellular differentiation (17, 32, 36).

An alternative approach to generating cell-type-specific markers has been to raise monoclonal antibodies (MAbs). The MAb MUD1 detects the surface glycoprotein PsA (prespore-specific antigen) first appearing on the prespore cell surface early in the prespore pathway, soon after tip formation (19). The MUD1 MAb has often been used to specifically localize prespore cells and to quantitate prespore gene expression (18, 37). PsA was originally defined as a 30-kilodalton (kDa) prespore-specific cell surface protein carrying a unique epitope which is recognized by MUD1 (19). By using genetic, immunological, and biochemical criteria, it was demonstrated that PsA is encoded by a single locus, *pspA*, and that the gene product is a protein with a molecular weight polymorphism in several wild isolates (12). Another epitope on PsA is recognized by the MAb MUD50 (13), which, in addition to PsA, recognizes a diverse group of prespore proteins (S. Alexander, E. Smith, L. Davis, A. Gooley, S. B. Por, L. Browne, and K. L. Williams, Differentiation, in press). West and Loomis (44) have shown that the PsA (SP29) glycoprotein lacks a glycoconjugate(s) in *modB* mutants, and it has been shown that the MUD50 epitope is, or is part of, this carbohydrate (Alexander et al., in press).

Another desirable feature of a marker is that it be well defined structurally and functionally. PsA is the best characterized cell-type-specific surface molecule of the multicellular stage. Here, we have determined the entire sequence of the D19 gene and have shown that it encodes the PsA protein, which was purified and partially sequenced.

MATERIALS AND METHODS

Strains and culture conditions. PsA was purified from the slug stage of three strains of *D. discoideum*: two wild isolates, WS576 and WS380B (9); and a mutant, HU2440, which carries the *modB501* mutation (44) in the genetic background of wild isolate DD61 (12; A. A. Gooley and K. L. Williams, manuscript in preparation).

Preparation of slug extracts and purification of PsA by immunoaffinity chromatography. Two methods were used to

* Corresponding author.
† Present address: St. Vincent's Hospital, Darlinghurst, Sydney, New South Wales, Australia.

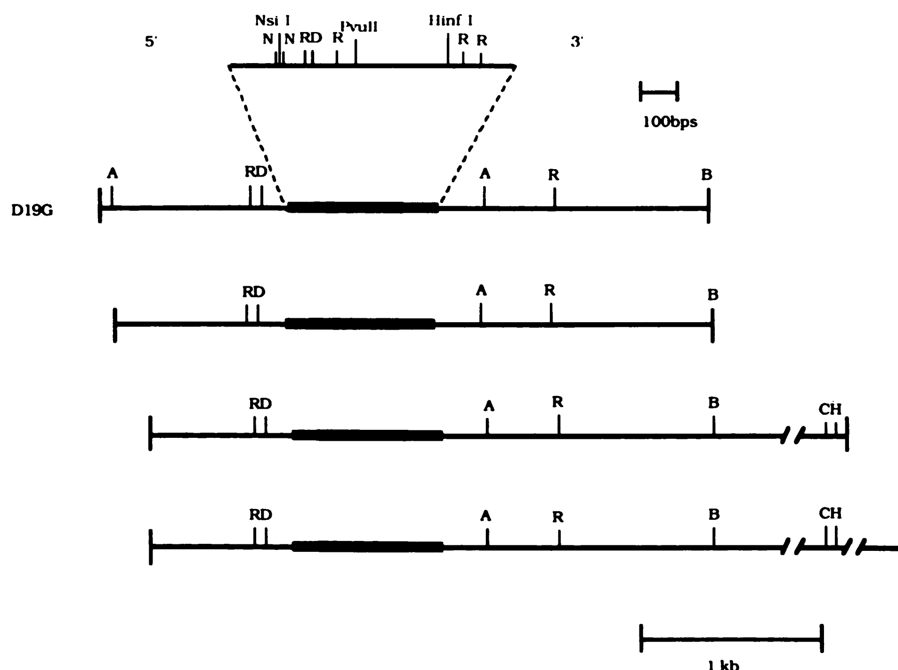

FIG. 1. Structure of the D19 genomic clones. The transcribed region ( ) contains the translated region ( ) and the 5′ and 3′ untranslated regions ( ). This part of the map has been expanded twofold relative to the intergenic DNA. Abbreviations for restriction sites are as follows: A, *Ava*II; B, *Bam*HI; C, *Cla*I; D, *Dde*I; H, *Hin*dIII; N, *Nde*I; R, *Rsa*I. D19G and three other genomic clones are shown.

prepare detergent extracts of slug cells. First, ~5 × 10^{10} slug cells were lysed in phosphate-buffered saline–0.1% (vol/vol) Nonidet P-40 (NP-40; Sigma Chemical Co.). The detergent extracts were passaged over two guard columns (20 ml of Sepharose 6B followed by 4 ml of CNBr-Sepharose 4B blocked with glycine) before chromatography on a MUD1 affinity column. Increased yields of PsA were obtained with a combination of detergents. Crude membranes were prepared from ~5 × 10^{10} slug cells by the Tween 40 (Sigma) detergent method used to purify Thy-1 from rat thymocytes (35). The membranes were then dissolved in 2% (wt/vol) sodium deoxycholate (BDH), and before binding the extract to the MUD1 affinity column, it was found necessary to add NP-40 to a final concentration of 1% (vol/vol) to restore antibody-antigen interaction. PsA was bound to a column of 1 ml of CNBr-Sepharose (Pharmacia) coupled to the MAb MUD1. Bound PsA was eluted from the column in 1-ml samples of 50 mM diethylamine (Sigma)–NaOH (pH 11.5)–0.1% (vol/vol) NP-40 and immediately adjusted to pH 8.0 with solid glycine. Fractions containing PsA, as detected with MUD1 in a modified dot-immunobinding assay (39), were pooled and lyophilized.

Purified PsA was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) by the method of Laemmli (22) with 12.5% (wt/vol) acrylamide in the resolving gel and electroblotted onto nitrocellulose by the method of either Towbin et al. (41) or Kyhse-Anderson (20).

Endoproteinase Glu-C (*Staphylococcus aureus* V8 protease) digestion. Protease digests of reduced and carboxymethylated PsA and peptide separation on high-pressure liquid chromatography were essentially as described by Meyer et al. (28). Cleavage with V8 protease (Miles Laboratories, Inc.) (5% wt/wt) was performed for 2 h (three times) at 37°C in 25 mM Tris hydrochloride–0.1 mM EDTA–0.025% (wt/vol) SDS (pH 7.8). The digestion mixture (4.5 nmol in 3.0 ml) was chromatographed on a 10-μm Vydac pH-stable reversed-phase column (4.6 by 250 mm) with a hexafluoroacetone (E. Merck AG)-containing solvent system. Solvent A contained (vol/vol) hexafluoroacetone-NH_3 (pH 8.6); solvent B contained 64% (vol/vol) acetonitrile–20% (vol/vol) propanol–0.03% (vol/vol) hexafluoroacetone-NH_3 (pH 8.6). A gradient ranging from 0 to 75% solvent B was applied at a flow rate of 1 ml/min.

Sequence analysis. The amino acid sequences of the N terminus and V8 peptides were determined by automated gas phase Edman degradation with on-line identification of the PTH-amino acids (16). Chemicals and solvents for automated sequence analysis were from Applied Biosystems.

Isolation and analysis of the D19G genomic clone. A genomic library, constructed by insertion of a partial *Sau*3A digest of *D. discoideum* DNA from strain AX2 into the *Bam*HI site of the plasmid pAT153, was screened with an internal *Nsi*I-*Hin*fI fragment from the D19 cDNA clone (1). The genomic clone (D19G) was sequenced by the method of Maxam and Gilbert (26) after labeling at the *Nde*I, *Rsa*I, and *Hin*fI sites (Fig. 1). Other cloning methods used were essentially as described by Maniatis et al. (25).

Insertion of a human c-*myc* epitope into D19G. The c-*myc* oligonucleotide is a 33-base-pair sequence which encodes an eight-amino-acid epitope of the human c-*myc* protein recognized by a MAb, 9E10 (10, 29). The oligonucleotide was synthesized with *D. discoideum* codon usage and inserted in frame into the unique *Pvu*II site in the D19G clone (coding strand: 5′-CT GAA GAA AAA TTA ATT TCA GAA GAA GAT CTC T-3′) (Fig. 1). The two strands of the oligonucleotide were annealed at 60°C for 1 h in 1× ligase buffer (50 mM Tris [pH 7.5], 10 mM $MgCl_2$, 10 mM dithiothreitol, 5 mM ATP) and then ligated to the linearized vector with the same buffer. A 50-fold molar excess of nonphosphorylated oligonucleotide over vector was used to ensure efficient insertion of a single copy of the epitope. The oligonucleotide

was designed so that a *Pvu*II site was regenerated at one end while at the other, a *Bgl*II site was created. This latter site was used in sequencing to verify correct orientation of the epitope. The c-*myc*-marked gene was excised from pAT153 by using the *Eco*RI site in the vector and a *Bam*HI site which was regenerated at the 3′ boundary of the insert. By using this fragment, the gene was cloned into the corresponding multilinker sites of the *D. discoideum* transformation vector, pB10TP2 (8).

DNA-mediated transformation and analysis of cellular differentiation. The *D. discoideum* strain AX2 (43) was used for DNA-mediated transformation by the modified protocol described by Early and Williams (8), which is based on the procedure of Nellen et al. (30). Pooled populations of transformants were grown in HL5 medium (43) in shaken suspension. Transformants were developed on membrane filters (Millipore Corp.), after washing by centrifugation in 20 mM KH_2PO_4-K_2HPO_4, pH 6.0 (40). To obtain migrating slugs, cells were developed as single streaks of approximately 10^9 cells per ml on 2% (wt/vol) water agar. The plates were kept moist and dark except for a single slit, which provided a unidirectional light source. Prestalk and prespore cells were separated by the method of Ratner and Borth (33).

Analysis of gene expression by Northern (RNA) transfer. Total cellular nucleic acid was extracted by alternating phenol and chloroform extractions. Five micrograms of each nucleic acid sample was electrophoresed through a formaldehyde denaturing gel and transferred onto nitrocellulose as described by Maniatis et al. (25), except that the filters were baked and prehybridized immediately after transfer. The hybridization was carried out in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at 32°C with the kinase-labeled, noncoding strand of the c-*myc* oligonucleotide as a probe. The filters were washed for 5 min (twice) at 37°C in 5× SSC–0.1% SDS.

Immunohistochemical analysis of transformants. Disaggregated cells in KH_2PO_4-K_2HPO_4 were prepared by passaging slugs through a series of syringes with needles of successively decreasing bore size. The disaggregated cells were washed and suspended at about 5×10^6/ml in KH_2PO_4-K_2HPO_4 before fixing in 60% methanol and dispensing as 5-μl drops onto poly-L-lysine subbed multiwell slides.

Slides were stained with the affinity-purified c-*myc* MAb, 9E10, prepared from culture fluid (10 μl per field) (10). This was diluted 1:50 in phosphate-buffered saline, and the slides were incubated in a humidified chamber for 30 min at room temperature. The specimens were washed extensively with phosphate-buffered saline for 15 min which was followed by the addition of 10 μl of rhodamine-conjugated rabbit anti-mouse immunoglobulin, diluted 1:30 in phosphate-buffered saline, again for 30 min at room temperature. Washing was carried out as before. The slides were mounted in Gelvatol and examined under a fluorescence phase-contrast microscope.

RESULTS

Purification of PsA by MUD1 affinity chromatography. PsA was purified from slug cell extracts of three strains of *D. discoideum* by MAb MUD1 affinity chromatography. Extracts prepared by lysing WS380B cells in NP-40 were first chromatographed over two guard columns. These columns removed nonspecific material, particularly a 37-kDa protein (this contaminant was later identified as the lectin discoidin). Figure 2 shows three stages of the purification procedure; the original NP-40 extract (lane 2); the material which did not bind to the MUD1 column (lane 3); PsA eluted off the column, seen here to migrate at 27 kDa (lane 4); and the same material as in lane 4, except that it was electroblotted onto nitrocellulose and probed with MAb MUD1 (lane 5).

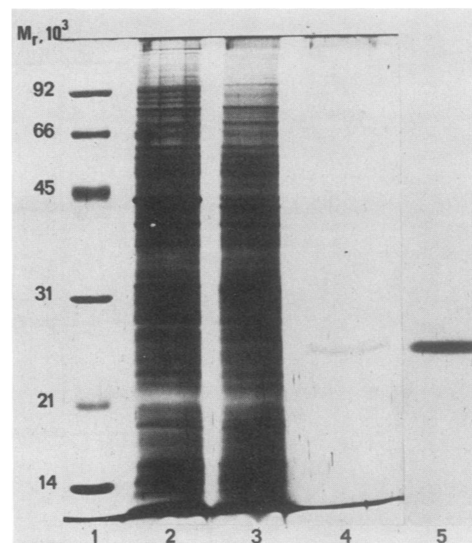

FIG. 2. Purification of PsA from strain WS380B. Protein samples were separated by SDS-PAGE and silver stained. Lanes: 1, molecular weight markers; 2, original 0.1% (vol/vol) NP-40 cell extract; 3, MUD1 affinity column flow through; 4, PsA eluted off the column at pH 11.5; 5, the same material as in lane 4, electroblotted onto nitrocellulose and stained with MAb MUD1 culture supernatant and peroxidase-conjugated second antibody.

PsA purified from this WS380B extract was subjected to automated Edman degradation, and 12 of the first 14 residues were assigned (Table 1). This procedure was repeated for a polymorphic wild strain, WS576 (PsA, 33 kDa), and sequence analysis identified 16 of the first 21 residues, which were identical to the N terminus of WS380B (Table 1). This established that two different-sized versions of PsA had the same N terminus. Typical yields of PsA by this method were 50 to 100 μg of protein per 10^{11} cells. Since insufficient material was recovered to allow further analysis, a purification protocol involving a combination of sodium deoxycholate and NP-40 was devised (see Materials and Methods).

This method resulted in typical yields of 0.5 to 1.5 mg of protein per 10^{11} cells, although the level of purification of PsA was somewhat lower than that achieved by the original method. However, the eluate was much enriched for PsA (Fig. 3, lane 1), seen here to migrate at 22 kDa (strain HU2440, a *modB* mutant lacking a glycoconjugate). The identity of PsA was confirmed by electroblotting the material onto nitrocellulose and probing with MAb (MUD1) (Fig. 3, lane 2). Attempts to further purify PsA resulted in irreversible loss on high-pressure liquid chromatography columns and low recovery from G-252 Sephadex. However, precipitation of the protein in prechilled acetone and reduction and carboxymethylation resulted in either loss of the contaminant or blocking of its N terminus (see below).

N-terminal amino acid sequence of PsA purified from HU2440. PsA purified from HU2440 by MUD1 affinity chromatography was subjected to N-terminal degradation, and 54 residues were identified. Only one sequence could be identified from the preparation (Table 1). Therefore, the contaminating protein was not present in sufficient quantities to interfere with the sequencing of the intact PsA protein.

TABLE 1. Amino acid sequences of PsA purified from three strains

Strain	*pspA* allele[a]	Sequence[b]
WS380B	352	YDYFT--LAN QNPV
WS576[c]	353	YDYFT--LAN QNPV-A-V-V I
HU2440[d]	351	YDYFTTTLAN QNPVCASVDV IQNVCTEVCG RFVRYIPDAT NTNQFTFAEY TTNQ
HU2440[e]	351	YDYFTTTLAN QNPVCASVDV IQNVCTEVCG RFVRYIPDAT NTNQFTFAEY TTNQCTVQVT PAVTNTFTCA DQTSSHALGS D-S

[a] The gene locus which encodes the PsA protein is called *pspA*. Numbers denote alleles for different isolates (12).
[b] Cleavages at the glutamate residues are shown in boldface. Amino acids are designated by the single-letter code. -, Unassigned residue.
[c] WS576 sequence from the 0.1% (vol/vol) NP-40 extract.
[d] Sequence from the material purified by using sodium deoxycholate and NP-40.
[e] Sequences from the V8 digest of PsA purified from HU2440.

V8 peptide map of PsA purified from HU2440 and sequencing of selected peptides. Three major peptides from the V8 digest of PsA purified from HU2440 were separated by reversed-phase high-pressure liquid chromatography (Fig. 4). Peptide I spans from residue 1 to 27, and peptide II spans from residue 28 to 49. The double peak, III, was analyzed to determine whether it represented two separate V8 fragments or whether part of the peak was undigested PsA. Sequence analysis revealed that both peaks had the N terminus Tyr-Thr-Thr-Asn, which is identical to the N-terminal sequence at positions 50 through 53. This suggests that the double peak resulted from a form of microheterogeneity due either to bound SDS, which was present in the V8 digestion buffer (38), or to some form of posttranslational modification (PsA is phosphorylated [Alexander et al., in press]). The V8 fragments IIIA and IIIB were pooled, and the sequence was extended from residue 50 to serine at position 83, with residue 82 unassigned.

A fourth peptide (IV; Fig. 4) from the V8 digest of HU2440 contained at least three sequences, one of which showed 100% homology with a V8 fragment of the lectin discoidin (3). The lack of discoidin in PsA purified by using the guard columns is explained by its removal by binding to these columns.

Isolation and characterization of a D19 genomic clone. Southern blotting of genomic AX2 DNA at high stringency indicated that D19 was a single-copy gene, but at lower stringency, cross-hybridizing restriction fragments were detected (data not shown). A restriction fragment derived from within the D19 cDNA insert was used, at high stringency, to probe a genomic library in the plasmid pAT153. Four independent positive clones were isolated and characterized by restriction mapping and Southern blotting. These four clones have an identical structure in the regions in which they overlap (Fig. 1). The position of the gene was deduced by restriction mapping. The clone with the greatest amount of sequence flanking the D19 coding region was designated D19G and used for all further analyses (Fig. 1).

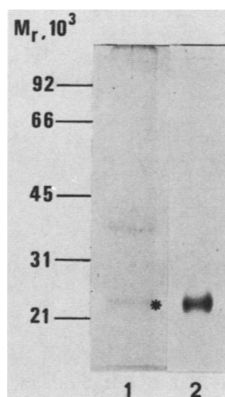

FIG. 3. MUD1 affinity column eluates of sodium deoxycholate–NP-40 HU2440 cell extract. Protein samples were separated by SDS-PAGE. Lane 1, Coomassie blue stain of PsA (*); lane 2, the same sample electroblotted onto nitrocellulose and stained with MAb MUD1 and peroxidase-conjugated second antibody.

The D19G clone contains an insert of 3.37 kilobases (kb). A total of 1.1 kb of nucleotide sequence, spanning the coding and 5′ and 3′ untranslated regions of the gene, was determined. A single start site of transcription was detected by primer extension by using RNA isolated from V12M2 prespore cells or from AX2 slugs (data not shown). The 5′ noncoding region is approximately 90 nucleotides long and has an A+T content of 95%. The length of the 3′ noncoding region was estimated from the cDNA clone to be 150 nucleotides. The coding region is 504 nucleotides long, contains no introns, and encodes a predicted polypeptide of 168 amino acids (Fig. 5). The overall size of the mRNA is estimated to be 736 nucleotides, which is in close agreement with the results of gel electrophoresis.

Structure of the predicted polypeptide. The N terminus of the D19 gene encodes a polypeptide tract which has the characteristics of a signal sequence (6). A hydrophobic "core" of 15 amino acids, with a predicted α-helical structure, follows a short hydrophilic region adjacent to the methionine. Comparison of the sequence on the immediate C-terminal side of the predicted cleavage site (i.e., following Ala-Asn-Ala) shows it to be identical to the 81 N-terminal residues of the PsA protein. Since we have shown the D19 gene to be unique in the genome, it must encode the PsA protein. The protein encoded by the D19 gene has a predicted molecular weight of 17,879 and a predicted pI of 4.62. It is composed of a higher-than-average proportion of nonpolar residues and a lower-than-average proportion of potentially charged residues (Fig. 6). Remarkably, it contains over 21% threonine residues. These are potential sites for O-linked glycosylation. Although there is one potential N-linked glycosylation site (Asn-Pro-Thr; Fig. 5), this is not utilized (4). Bause (4) has shown that no glycosyl transfer was detected on those peptides which contain a proline residue in position X in the Asn-X-Ser/Thr triplet sequence.

The extreme C terminus contains a largely hydrophobic stretch of approximately 18 amino acids, but these are not followed by polar residues which could act as a membrane anchor sequence (Fig. 5 and 6). This region is interspersed with hydrophilic residues, most commonly serine. Immediately before the hydrophobic tail there are four copies of a

FIG. 4. Separation of peptides from a V8 digest of PsA purified from HU2440. The digestion mixture was separated on a Vydac C8 pH-stable column in an acetonitrile-propanol (solvent B) gradient at a flow rate of 1 ml/min. PsA peptides are designated I, II, and IIIA and B with respect to the time they eluted off the column. Peak IV is a peptide of the lectin discoidin. %B, Percentage of solvent B.

perfectly conserved, four-amino-acid repeat, Pro-Thr-Val-Thr. These repeats are followed by three similar, but imperfect, repeats. The repeats form a proline-rich region, containing 10 of the total 13 proline residues present in the protein. Within this 31-amino-acid stretch there are nine short regions where proline alternates with a hydroxyamino acid, normally a threonine residue. This pattern is interspersed with valine, asparagine, or alanine residues, present as single amino acids. Overall, the primary amino acid sequence predicts a polypeptide with an extended structure.

Expression of a marked copy of the gene after transformation into *D. discoideum* and localization of the protein product. The PsA gene was marked by insertion of an oligonucleotide encoding an eight-amino-acid epitope of the human c-*myc* protein (29), which is recognized by the 9E10 MAb (10). The epitope was inserted in frame within codon 81. The marked gene was subcloned into the *D. discoideum* transformation vector pB10TP2 (8). The final construct, pB10D19G, contained the entire coding region with approximately 1.0 kb of sequence upstream of the cap site and about 1.9 kb of downstream sequence (Fig. 7).

After transformation into AX2 cells and selection of stable transformants by using G418, expression of the gene was analyzed in pooled populations of cells. The vector copy number within the population was approximately 10. The mRNA transcribed from the marked D19 gene was analyzed by Northern transfer hybridization by using the noncoding strand of the c-*myc* oligonucleotide as a probe. This mRNA accumulated with the same kinetics as the transcript of the endogenous PsA gene (Fig. 8A). The mRNA was first

```
  1                                          ATTTTATTAATATAATTCTTAATAATTATAAAAAAAAAAAAAAAA

 46 AAACAACCAAAATATTATATTATTTTTTGTAAATATAAATTAAAAAA ATG AAA TTC CAA CAT ACA
                                                    MET Lys Phe Gln His Thr   6

111 TTT ATT GCA TTA TTA TCA CTA TTA ACA TAT GCA AAT GCA TAT GAT TAT TTC ACA
    Phe Ile Ala Leu Leu Ser Leu Leu Thr Tyr Ala Asn Ala Tyr Asp Tyr Phe Thr  24

165 ACA ACA TTG GCC AAT CAA AAT CCA GTT TGT GCT TCA GTA GAT GTC ATA CAA AAT
    Thr Thr Leu Ala Asn Gln Asn Pro Val Cys Ala Ser Val Asp Val Ile Gln Asn  42

219 GTT TGT ACT GAG GTT TGT GGT AGA TTT GTC CGT TAC ATT CCT GAT GCT ACC AAT
    Val Cys Thr Glu Val Cys Gly Arg Phe Val Arg Tyr Ile Pro Asp Ala Thr Asn  60

273 ACA AAT CAA TTC ACT TTT GCC GAA TAT ACT ACA AAC CAA TGT ACT GTT CAA GTA
    Thr Asn Gln Phe Thr Phe Ala Glu Tyr Thr Thr Asn Gln Cys Thr Val Gln Val  78

327 ACA CCA GCT GTA ACA AAT ACA TTT ACT TGC GCT GAT CAA ACT TCT TCA CAT GCA
    Thr Pro Ala Val Thr Asn Thr Phe Thr Cys Ala Asp Gln Thr Ser Ser His Ala  96

381 CTC GGT TCT GAT TGG AGT GGA GTT TGT AAA ATC ACT GCC ACA CCC GCA CCA ACA
    Leu Gly Ser Asp Trp Ser Gly Val Cys Lys Ile Thr Ala Thr Pro Ala Pro Thr 114
                                                                    -------
435 GTT ACA CCA ACA GTT ACA CCA ACA GTT ACA CCA ACA GTA ACA CCA ACA CCA ACA
    Val Thr Pro Thr Val Thr Pro Thr Val Thr Pro Thr Val Thr Pro Thr Pro Thr 132
    ------- --------------- --------------- ---------------
489 AAT ACA CCA AAT CCA ACC CCG TCC CAA ACA TCA ACA ACA ACT GGT TCT GCT TCA
    Asn Thr Pro Asn Pro Thr Pro Ser Gln Thr Ser Thr Thr Thr Gly Ser Ala Ser 150
                 •   •   •
543 ACT GTT GTT GCA AGT TTA TCA CTC ATT ATT TTC TCA ATG ATT CTT TCC CTT TGT
    Thr Val Val Ala Ser Leu Ser Leu Ile Ile Phe Ser MET Ile Leu Ser Leu Cys 168

597 TAATTTTAAAAATAACAAATGGTAAAAATTATTTGGCATTTTCAAGTACAATTTTAAATTTTTCAAAAAAA

668 AATAAAATTAAATAAATAATTAAATAAATTAATTAATAAAAAAAATAAAATTAAAACTTTTTTAATTTA
```

FIG. 5. Sequence of the D19 mRNA. The sequence was obtained from the genomic clone D19G by the chemical cleavage method (26). The open reading frame shown is confirmed by comparison with the protein sequence (Table 1). The length of the 5′ noncoding region was predicted by primer extension analysis (data not shown), while the length of the 3′ noncoding region was deduced by comparison with sequence from the cDNA clone. Symbols: ——, position of the four tetra-amino acid repeats; ●●●, Asn-Pro-Thr triplet sequence which, although it conforms to the Asn-X-Thr sequence, is not a predicted site for N-linked glycosylation (4). The numbers on the left represent nucleotides, while the numbers on the right represent amino acids.

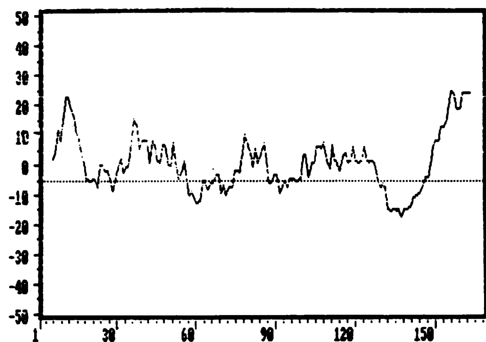

FIG. 6. Plot of the hydropathic index of the predicted primary translation product of the D19 gene. The plot was computed according to the method of Kyte and Doolittle (21) with a window size of nine amino acids. The 168 amino acids of the predicted D19 coding region are represented along the x axis. On the y axis, positive values indicate hydrophobicity, while negative values indicate hydrophilicity.

detectable at the tipped aggregate stage, reached a peak of abundance at the first finger and standing slug stages, and fell in concentration at culmination. The mRNA also showed correct spatial localization. Slugs of the transformed population were disaggregated, and the cells were separated by two rounds of Percoll density gradient centrifugation. Comparison of the light (prestalk) and heavy (prespore) mRNA fractions showed strong enrichment of the mRNA in prespore cells (Fig. 8B). The D19 cDNA insert detected mRNA from both the endogenous and transformed genes as a single band when it was used as a probe in Northern transfer. In a transformed population with approximately 5 to 10 copies of the transforming gene, there was approximately five times as much mRNA at this position in the gel as was found in an equivalent number of AX2 cells harvested at the same stages of development (data not shown). Thus, 990 base pairs of sequence upstream of the cap site contains all the information required for correctly regulated expression at approximately the level of the endogenous gene.

The transformed population was screened for expression of the marked gene by indirect immunofluorescence (Fig. 9). Methanol-fixed, disaggregated slug cells were stained with the anti-c-*myc* MAb and with a rhodamine-conjugated second antibody. Bright, cell surface staining of a high proportion of the transformed cells was seen, but there was no staining of (control) cells harvested from nontransformed slugs (data not shown).

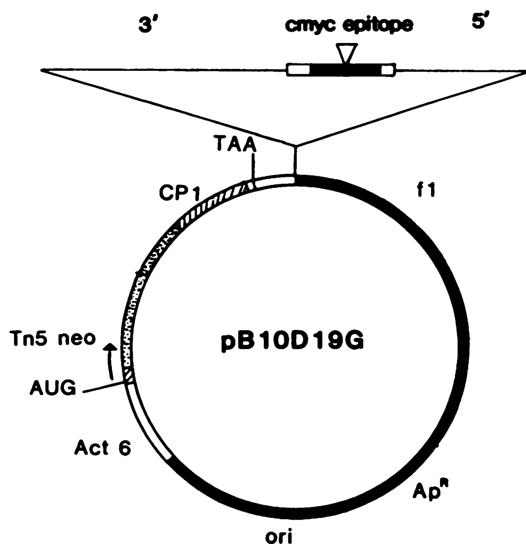

FIG. 7. Structure of pB10D19G. The plasmid was constructed by insertion of the entire D19G genomic fragment into the transformation vector, pB10TP2 (8). The position and orientation of the D19 gene within the insert are shown. The relative position of the c-*myc* epitope within the coding region is also indicated. The solid black line within the vector represents sequences derived from pEMBL. The actin 6 promoter (Act 6; ▭) drives expression of the Tn5 neomycin resistance gene (Tn5 neo; ▤). This transcript, which enables the selection of *Dictyostelium* cells resistant to G418, is terminated with sequences from the cysteine proteinase 1 gene (CP1; ▨). ori, Origin of replication; f1, phage f1.

DISCUSSION

The predicted amino acid sequence of the coding region of the D19 gene is identical to the first 81 residues of PsA purified by MAb MUD1 affinity chromatography. Other aspects of the DNA sequence correlate well with the physical properties of the PsA protein. The pI of PsA is between 4 and 5 (12), and the pI predicted from the primary sequence of the D19 polypeptide is 4.62. The sequence of the PsA gene indicates that there is a signal peptide and the predicted cleavage site would generate an N terminus identical to that obtained for the protein. However, the molecular weight of the protein is predicted to be 17,879, while that of PsA has been judged to be 30,000 (19). This apparent discrepancy can be accounted for by posttranslational modifications. PsA and *D. discoideum* contact sites A (CsA) are glycoproteins which show major decreases in size in *modB* mutants (11, 44; Alexander et al., in press). The primary defect caused by the *modB* mutation is thought to be in the mechanism responsible for the retardation of the protein until type 2 glycosylation occurs (15).

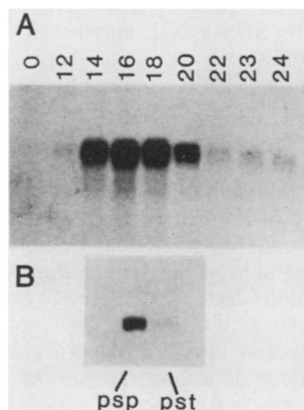

FIG. 8. Expression of the transformed gene during development. Pooled populations of AX2 cells transformed with pB10D19G were analyzed. Total cellular RNA was extracted and used for Northern transfer analysis. The noncoding strand of the c-*myc* oligonucleotide was end-labeled with ^{32}P by using T4 polynucleotide kinase. This probe specifically detects mRNA from the transformed gene. (A) Times shown (h) indicate the interval after initiation of development. The corresponding morphological stages are as follows: 12 h, loose mounds; 14 h, tipped mounds; 16 h, first finger; 18 h, standing slug; 20 h, late preculminates; 24 h, culminates. (B) Disaggregated slug cells were separated by two rounds of Percoll density gradient centrifugation (33), and RNA was extracted from isolated prespore (psp) and prestalk (pst) cells.

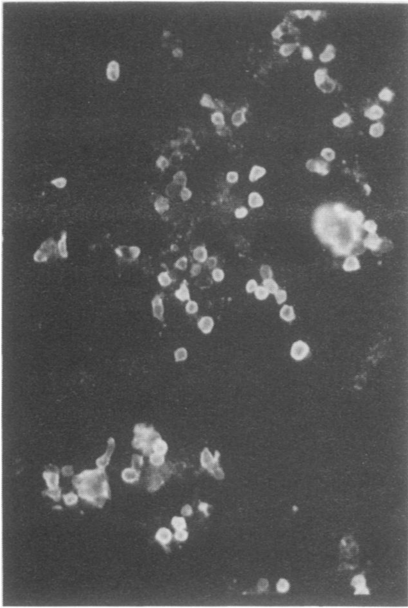

FIG. 9. Immunostaining of transformants. Slug cells, disaggregated and fixed with methanol, were stained with the anti-c-*myc* MAb 9E10 and a rhodamine-conjugated second antibody before microscopy. Magnification, ×180.

Type 2 carbohydrate accounts for 12 kDa of the molecular mass of CsA (11) and for 7 kDa of the PsA apparent molecular mass as determined by SDS-PAGE (30 kDa in NC4 and 23 kDa in the *modB* mutant of NC4, HU2428; data not shown). Therefore, in PsA the *modB*-sensitive carbohydrate accounts for 7 kDa of the 12-kDa difference between the molecular mass predicted from the translated D19 gene and the apparent molecular mass determined by SDS-PAGE. The difference between the molecular mass predicted from the DNA sequence and the apparent molecular mass determined by SDS-PAGE may result from the unusually high content of hydroxyamino acids (21% threonine and 8% serine). A similar discrepancy, observed between the predicted size of VP9 (22 kDa), a minor structural component of a rotavirus capsid, and the apparent size of the nonglycosylated protein on SDS-PAGE (28 kDa), has been attributed to its unusually high content (20%) of serine residues (5). None of the Thr or Ser residues in the first 81 residues of PsA can be glycosylated, since glycosylated residues are not soluble in the sequencing reagents. Therefore, all of the carbohydrate is predicted to be on the Thr and Ser residues downstream of residue 81.

Despite the fact that PsA is a known plasma membrane protein, the sequence data does not predict a hydrophobic, transmembrane segment followed by a hydrophilic cytoplasmic "anchor" domain. Rather, the C terminus closely resembles the equivalent regions of a set of cell surface glycoproteins known to be attached to the membrane via a glycosyl-phosphatidylinositol link (23). At the extreme C terminus of these proteins, largely hydrophobic residues are interspersed with hydrophilic residues, often serine or threonine. In the mature polypeptide, the hydrophobic tails are removed by cleavage and the lipid is attached via an amide linkage to the new C-terminal amino acid. The PsA protein shows direct homology in the C-terminal region to two proteins involved in cell adhesion, *D. discoideum* CsA (31) and a form of chicken neural-cell adhesion molecule known to be linked to the membrane via a glycosyl-phosphatidylinositol link (Fig. 10) (14). It seems almost certain that both PsA and CsA are also attached to the cell membrane via a glycolipid. Direct analysis of CsA shows that it is attached in such a manner (J. Stadler, G. Bauer, and G. Gerisch, personal communication). The predicted glycosyl-phosphatidylinositol linkage would allow the rapid removal of PsA protein from the cell surface by the action of a specific phospholipase, releasing cleaved protein into the extracellular matrix. This potential rapid loss could be important during differentiation of *D. discoideum*, which involves rapid changes in the presentation of surface molecules (34).

There is homology between PsA and CsA upstream of the extreme C terminus in the proline-rich repeat region (31) (Fig. 10). In PsA from strain AX2, this sequence is 31 amino acids long and contains nine short, interspersed groups of alternating threonines and prolines. In the CsA protein, the comparable region is 32 amino acids long and contains three longer clusters of proline residues which alternate with threonine or serine. Such a primary sequence is capable of forming an extended chain in proline-proline conformation, which could be important in allowing the projection of these proteins into the extracellular space. It is possible therefore that the homology merely indicates a requirement for a similar stalk-like structure, but in combination with the extreme C-terminal homology to CsA and neural-cell adhesion molecules, this internal homology suggests that these three genes may belong to the same family. Direct evidence is, however, clearly required to prove a role of PsA in cell adhesion.

The repeat region in the C terminus could also be responsible for the polymorphism in size of the PsA protein. In different *D. discoideum* strains there is a difference in apparent molecular mass of ±2 kDa (12). The repeats in PsA are four amino acids long and there are four perfectly conserved copies. Unequal crossover would result in the loss of up to three copies of the repeat from one chromosome and the insertion of up to three extra copies into the other; at least one copy of the repeat would always be retained. In this way, exact insertion or deletion of up to 12 amino acids

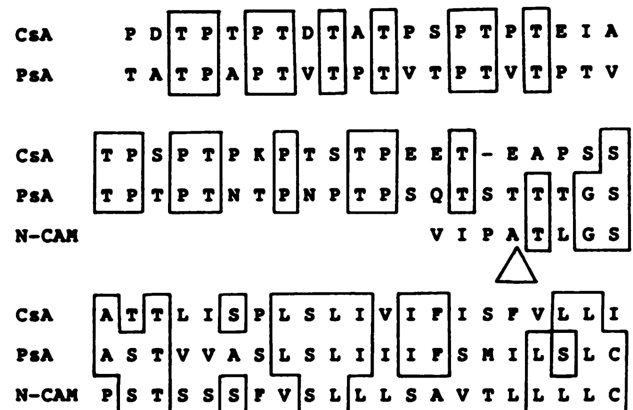

FIG. 10. Comparison of PsA amino acid sequence with CsA and neural-cell adhesion molecules. The C-terminal 63 amino acids of PsA are shown, aligned with the C-terminal 62 amino acids of CsA (31) and with the 25 amino acids unique to the form of chicken neural-cell adhesion molecule known to be attached to the membrane via phosphatidylinositol (14). The arrowhead indicates the boundary between this unique sequence and sequence from the previous exon. The boxed regions indicate the occurrence of identical amino acids.

might occur. The repeat contains threonines which are potential sites for O-linked glycosylation, and so these 12-amino-acid changes could account for the 2-kDa size polymorphisms.

The results of the transformation show that 990 base pairs upstream of the cap site of D19 are sufficient for correct temporal and spatial mRNA accumulation. It is not clear whether the signals for prespore-specific expression reside within the promoter or whether a generalized promoter element combined with posttranscriptional control leads to a specifically localized transcript. The upstream and 5′ noncoding sequences of one other prespore-specific gene, EB4, have been reported (2). The EB4 and PsA genes are expressed at similar times in development. The transcripts from both genes are rapidly lost because of selective destabilization when aggregates are disrupted, and exogenous cAMP acts to prevent this destabilization (24). One suggestion was that the extremely long, relatively GC-rich, 5′ noncoding region of EB4 might be involved in its posttranscriptional regulation. However, there is no obvious similarity to the 5′ noncoding region of PsA, which is much shorter and has a much lower GC content.

Since we have obtained correctly regulated expression of the transformed D19 gene and an appropriate localization of the marked protein, it will now be possible to use transformation both to localize the signals required for prespore-specific gene expression and to investigate the possible functions of the protein in the migratory slug.

ACKNOWLEDGMENTS

This research was supported by ARGS grant A08415732, Macquarie University Research grants to K.L.W. and E.S., and NH&MRC grant 860245. S.B.P. was supported by a Macquarie University Research Fellowship. The visit of A.A.G. to the laboratory of H.E.M. was supported by the Deutsche Forschungsgemeinschaft, SFB 168.

We thank S. McRobbie for his suggestion that the D19 gene might encode the PsA protein.

LITERATURE CITED

1. **Barklis, E., and H. F. Lodish.** 1983. Regulation of *Dictyostelium discoideum* mRNAs specific for prespore or prestalk cells. Cell **32**:1139–1148.
2. **Barklis, E., B. Pontius, K. Barfield, and H. F. Lodish.** 1985. Structure of the promoter of the *Dictyostelium discoideum* prespore EB4 gene. Mol. Cell. Biol. **5**:1465–1472.
3. **Bartles, J. R., W. A. Frazier, and S. D. Rosen.** 1982. Slime mold lectins. Int. Rev. Cytol. **75**:61–99.
4. **Bause, E.** 1983. Structural requirements of N-glycosylation of proteins. Biochem. J. **209**:331–336.
5. **Both, G. R.** 1987. Replication of the Reoviridae: information derived from gene cloning and expression. *In* P. Ahlquist, E. Domingo, and J. Holland (ed.), RNA Genetics 1987. CRC Press, Inc., Boca Raton, Fla.
6. **Briggs, M. S., and L. M. Gierasch.** 1986. Molecular mechanism of protein secretion: the role of the signal sequence. Adv. Protein Chem. **38**:109–180.
7. **Chisholm, R. L., E. Barklis, and H. F. Lodish.** 1984. Mechanism of sequential induction of cell-type specific mRNAs in *Dictyostelium discoideum*. Nature (London) **310**:67–69.
8. **Early, A. E., and J. G. Williams.** 1987. Two vectors which facilitate gene manipulation and a simplified transformation procedure for *Dictyostelium discoideum*. Gene **59**:99–106.
9. **Erdos, G. W., K. B. Raper, and L. K. Vogen.** 1973. Mating types and macrocyst formation in *Dictyostelium discoideum*. Proc. Natl. Acad. Sci. USA **70**:1828–1830.
10. **Evan, G. I., G. K. Lewis, G. Ramsay, and J. M. Bishop.** 1985. Isolation of monoclonal antibodies specific for human c-*myc* proto-oncogene product. Mol. Cell. Biol. **5**:3610–3616.
11. **Gerisch, G., U. Weinhart, G. Bertholdt, M. Claviez, and J. Stadler.** 1985. Incomplete contact site A glycoprotein in HL220, a *modB* mutant of *Dictyostelium discoideum*. J. Cell Sci. **73**:49–68.
12. **Grant, W. N., D. L. Welker, and K. L. Williams.** 1985. A polymorphic, prespore-specific cell surface glycoprotein is present in the extracellular matrix of *Dictyostelium discoideum*. Mol. Cell. Biol. **5**:2559–2566.
13. **Grant, W. N., and K. L. Williams.** 1983. Monoclonal antibody characterisation of slime sheath: the extracellular matrix of *Dictyostelium discoideum*. EMBO J. **2**:935–940.
14. **Hemperley, J. J., G. M. Edelman, and B. A. Cunningham.** 1986. cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a membrane-spanning region consistent with evidence for membrane attachment via a phosphatidylinositol intermediate. Proc. Natl. Acad. Sci. USA **83**:9822–9826.
15. **Hohmann, H. P., S. Bozzaro, R. Merkl, E. Wallraff, M. Yoshida, U. Weinhart, and G. Gerisch.** 1987. Posttranslational glycosylation of the contact site A protein of *Dictyostelium discoideum* is important for stability but not for its function in cell adhesion. EMBO J. **6**:3663–3671.
16. **Hunkapillar, M. W., R. M. Hewick, W. J. Dreyer, and L. E. Hood.** 1983. High-sensitivity sequencing with a gas-phase sequenator. Methods Enzymol. **91**:399–413.
17. **Jermyn, K. A., M. Berks, R. R. Kay, and J. G. Williams.** 1981. Two distinct classes of prestalk-enriched mRNA sequences in *Dictyostelium discoideum*. Development **100**:745–755.
18. **Krefft, M., L. Voet, J. H. Gregg, H. Mairhofer, and K. L. Williams.** 1984. Evidence that positional information is used to establish the prestalk-prespore pattern in *Dictyostelium discoideum* aggregates. EMBO J. **3**:201–206.
19. **Krefft, M., L. Voet, H. Mairhofer, and K. L. Williams.** 1983. Analysis of proportion regulation in slugs of *Dictyostelium discoideum* using a monoclonal antibody and a FACS-IV. Exp. Cell Res. **147**:235–239.
20. **Kyhse-Anderson, J.** 1984. Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of protein from polyacrylamide to nitrocellulose. J. Biochem. Biophys. Methods **10**:203–209.
21. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
22. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
23. **Low, M. G., M. A. J. Ferguson, A. H. Futerman, and I. Silman.** 1986. Covalently attached phosphatidylinositol as a hydrophobic anchor membrane proteins. Trends Biochem. Sci. **11**:212–215.
24. **Mangiarotti, G., A. Ceccarelli, and H. F. Lodish.** 1983. Cyclic AMP stabilises a class of developmentally regulated *Dictyostelium discoideum* mRNAs. Nature (London) **301**:616–618.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
27. **Mehdy, M. C., D. Ratner, and R. A. Firtel.** 1983. Induction of cell-type-specific expression in *Dictyostelium*. Cell **32**:763–771.
28. **Meyer, H. E., E. Hoffmann-Posorske, C. C. Kuhn, and L. M. G. Heilmeyer.** 1988. Microcharacterisation of phosphoserine-containing proteins. Localisation of the autophosphorylation of skeletal muscle phosphorylase kinase, p. 185–212. *In* H. Tschesche (ed.), Modern methods in protein chemistry, vol. 3. Walter de Gruyter, Verlag, Holland.
29. **Munro, S., and H. R. B. Pelham.** 1986. An Hsp70-like protein in the ER: identity with the 78kd glucose-regulated protein and immunoglobulin heavy chain binding protein. Cell **46**:291–300.
30. **Nellen, W., C. Silan, and R. A. Firtel.** 1984. DNA-mediated transformation in *Dictyostelium discoideum*: regulated expression of an actin gene fusion. Mol. Cell. Biol. **4**:2890–2898.
31. **Noegel, A., G. Gerisch, J. Stadler, and M. Westphal.** 1986.

Complete sequence and transcript regulation of a cell adhesion protein from aggregating *Dictyostelium* cells. EMBO J. **5**:1473–1476.
32. **Oyama, M., and D. D. Blumberg.** 1986. Changes during differentiation in requirements for cAMP for expression of cell-type-specific mRNAs in the cellular slime mould, *Dictyostelium discoideum*. Dev. Biol. **117**:550–556.
33. **Ratner, D., and W. Borth.** 1983. Comparison of differentiating *Dictyostelium discoideum* cell types separated by an improved method of density gradient centrifugation. Exp. Cell Res. **143**:1–73.
34. **Sadeghi, H., K. L. Williams, and C. Klein.** 1987. Characterization of 117 antigen: a developmentally regulated cell surface glycoprotein of *Dictyostelium discoideum*. J. Biol. Chem. **262**:16294–16299.
35. **Sandring, R., and A. F. Williams.** 1978. Glycoproteins and antigens of membranes prepared from rat thymocytes after lysis by shearing or with the detergent Tween-40. Biochim. Biophys. Acta **508**:85–96.
36. **Schaap, P., M. M. Van Looheren Campagne, R. Van Driel, W. Spek, P. J. M. Van Haastert, and J. Pinas.** 1986. Postaggregative differentiation induction by cAMP in *Dictyostelium*: intracellular transduction pathway and requirement for additional stimuli. Dev. Biol. **118**:52–63.
37. **Schaap, P., and M. Wang.** 1986. Interactions between adenosine and oscillatory cAMP signaling regulate size and pattern in *Dictyostelium*. Cell **45**:137–144.
38. **Shively, J. E.** 1986. Reverse phase HPLC, p. 41–87. *In* J. E. Shively (ed.), Methods of protein microcharacterisation: a practical handbook. Humana Press, Clifton, N.J.
39. **Smith, E., K. Roberts, G. W. Butcher, and G. Galfre.** 1984. Monoclonal antibody screening: two methods using antigens immobilized on nitrocellulose. Anal. Biochem. **138**:119–124.
40. **Sussman, M.** 1966. Biochemical and genetic methods in the study of cellular slime mold development. Methods Cell Physiol. **2**:397–410.
41. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
42. **Tsang, A., and J. M. Bradbury.** 1981. Separation and properties of prestalk and prespore cells of *Dictyostelium discoideum*. Exp. Cell Res. **132**:433–441.
43. **Watts, D. J., and J. M. Ashworth.** 1970. Growth of myxamoebae of the cellular slime mould *Dictyostelium discoideum*. Biochem. J. **119**:171–174.
44. **West, C. M., and W. F. Loomis.** 1985. Absence of a carbohydrate modification does not affect the level or subcellular localisation of three membrane glycoproteins in *modB* mutants of *Dictyostelium discoideum*. J. Biol. Chem. **260**:13803–13809.