Sequence organization and gene expression of pGD1, a plasmid found in a wild isolate of Dictyostelium

Hidefumi Orii*, Yoshimasa Tanaka* and Kaichiro Yanagisawa

Institute of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

Received December 20, 1988; Accepted January 23, 1989 EMBL accession no. X13703

ABSTRACT
We have determined the complete nucleotide sequence of pGD1, a plasmid found in a wild isolate of Dictyostelium. The 4439-bp long pGD1 contains only one, 2718-bases-long, open reading frame (ORF) and nearly perfect inverted repeats of 551 bp and 552 bp. Northern-blot analysis showed that only one 2.7-kb poly (A)+ RNA transcript was expressed at a maximum level, 2 h (early aggregation stage) after the onset of development. The expression of this transcript was suppressed by the addition of cAMP. In the upstream region of the ORF, there are several putative consensus sequences, e.g. (1) TGACTTAGAA- AAATTT which is a putative site for cleavage by topoisomerase I, and (2) TGACGACA which may be a cAMP-responsive element, found in several genes that are regulated by cAMP at the level of transcription. A possible mechanism of the partitioning of pGD1 into daughter cells is discussed.

INTRODUCTION
The cellular slime molds have been used extensively in studies of the regulatory mechanisms of cell differentiation and cell movement (for review, see 1). After starvation, solitary amoebae of the cellular slime molds enter the aggregation stage and form a multicellular slug which differentiates into stalk cells and spores after morphogenetic movement.

Several plasmids have been found in this organism. In Dictyostelium discoideum, five plasmids ranging in size from 5.8 kb to 27 kb have been found (2, 3) and one of these, Ddpl (13.8 kb), has been used to construct shuttle vectors (4, 5). Furthermore, the transcription of the plasmid Ddpl has been analyzed in detail, and it has been shown that three kinds of RNA are transcribed from the plasmid in growing cells and five additional transcripts are found at distinct phases during development (6). However, the functions of the plasmid and its transcripts are still unclear.

In a wild isolate of Dictyostelium species strain GA11, we found a uniquely configured plasmid. The plasmid resembles a round fan; it is 4.5 kb in length and contains closely situated, long inverted repeats (about 0.5-
Cappello et al. (8) have reported the entire sequence of DIRS-1 (4.8 kb), an apparent retrotransposon in D. discoideum, which has long, inverted terminal repeats. These authors have also proposed a mechanism for the biosynthesis of a putative, circular, transposon Intermediate DNA (8). From the structural similarity between the plasmid and this intermediate, we previously suggested a possible relationship between the plasmid pDGl and the putative Intermediate DNA of DIRS-1, even though these DNAs are found in different species and have different restriction maps (7).

It is of interest to examine the above-mentioned possibility and to examine the possible function and the unique structure of pDGl. Furthermore, thus far, there are no reports of nucleotide sequences of plasmids isolated from cellular slime molds. In this report, in addition to providing the entire sequence of pDGl, we show that the plasmid encodes only one 2.7-kb-long RNA transcript which is expressed at the early aggregation stage and is suppressed by the addition of cAMP. Furthermore, we have found, in the 5' flanking region of the transcript, sequences similar to the putative consensus sequence for the site of cleavage by topoisomerase I in the ribosomal DNA of D. discoideum (9-11) and to a cAMP-responsive element found in higher eukaryotic cells (12).

MATERIALS AND METHODS

Growth and development of the Dictyostelium strain GA11

The wild isolate of Dictyostelium species GA11 which harbored the plasmid pDGl was grown at 22°C in association with Klebsiella aerogenes on SM agar plates, as described previously (7). For synchronous development on agar plates, cells growing exponentially (4-6 x 10⁶ cells/ml) were harvested and freed from bacteria by repeated low-speed centrifugations. The cells were resuspended in Bonner's salt solution (NaCl, 0.61 g; KCl, 0.75 g; CaCl₂, 0.3 g per liter), plated on non-nutrient agar at a density of 2 x 10⁷ cells/cm² and were allowed to develop at 22°C. For synchronous development in suspension, the washed cells were suspended in phosphate buffer (3.2 mM Na₂HPO₄, 12.8 mM KH₂PO₄, pH 6.4) at a density of 2 x 10⁷ cells/ml and incubated on a gyratory shaker (150 rpm) at 22°C. To examine the effect of cAMP and its analogues, the chemicals were added at a final concentration of 0.1 mM at the onset of development. The treated cells were harvested at various times, as indicated in legends to Figures, and used for extraction of RNA.

Analysis of RNA transcripts

Total RNA was purified by the CsCl-cushion method (13). Cells were lysed
In 0.5% sarcosyl, 25 mM EDTA, 25 mM sodium citrate, 4.5 M guanidinium isothiocyanate, and centrifuged at 10,000 x g for 10 min. The supernatants were overlaid on 5.7 M CsCl, 50 mM EDTA, and centrifuged at 34,000 rpm for 12.5 h in a Beckman SW60 rotor. After centrifugation, the pellets were dissolved in 0.5% sodium dodecylsulfate (SDS), extracted successively with phenol, phenol/chloroform, and chloroform, and precipitated in ethanol. The precipitated RNA was dissolved in sterilized distilled water. Poly (A)^+ RNAs were obtained by further purification of the total RNA on messenger-activated paper according to the manufacturer's directions.

For Northern-blot hybridization, the total RNAs were separated on a 0.8% denaturing formaldehyde gel by electrophoresis (14) and blotted onto GeneScreen Plus filters according to the manufacturer's protocol. Then, each filter was incubated in a prehybridization solution of 50% formamide, 6 x SSC, 5 x Denhardt's solution, 0.5% SDS, 0.5 mg/ml of denatured salmon sperm DNA (1 x SSC: 0.15 M NaCl, 0.015 M sodium citrate; 100 x Denhardt's solution: 20 g Ficoll, 20 g polyvinylpyrrolidone, 20 g bovine serum albumin per liter) at 42° C for more than 3 h, and then hybridized to the ^32P-labelled DNA probe in hybridization solution (prehybridization solution containing 10% dextran sulfate) at 42° C overnight. Washing of the filters was performed in 1% SDS, 2 x SSC at 60° C for several hours with repeated changes of the washing solution. The filters were then exposed to FUJI X-ray film with an intensifying screen at -70° C. pDG1 DNA was labelled with [α-^32P]dATP (3000 Ci/mmol; 1 Ci=37 GBq), using a random-primed labelling kit, and used as hybridization probe. The specific activity of the probe was more than 10^8 cpm/μg. The concentration of the probe in the hybridization solution was less than 10 ng/ml.

**Sequencing of DNA**

Isolation of plasmid pDG1 was performed as previously described (7). Nucleotide sequences were determined by the dideoxy chain-termination method (15), using M13mp phages and pUC plasmids (16). However, the 0.2-kb ClaI-ClaI fragment, in the center of the long IRs which contain the unclonable 21-bp HaeIII fragment, was sequenced by the chemical method (17). To obtain the 0.2-kb ClaI-ClaI fragment, the 1.6-kb HindIII fragment of pDG1 was amplified in Escherichia coli strain H1604, as described previously (18). Strain H1604 has a temperature-sensitive gyraA43 (nalA43) mutation and was kindly provided by H. Ikeda (The Institute of Medical Science, The University of Tokyo). The determined sequences were analyzed using DBSYSTEM and ANALYSEQ programs on a VAX/780 computer (19, 20).
Materials

Adenosine 3',5'-cyclic monophosphate (cAMP), adenosine 5'-monophosphate (5'-AMP), 2'-deoxyadenosine 3',5'-cyclic monophosphate (2'-deoxy cAMP), guanosine 3',5'-cyclic monophosphate (cGMP), and calf-intestine alkaline phosphatase were purchased from Sigma. Messenger-activated paper, pUC18, Klenow fragment, T4 DNA ligase, M13 primer M1, four ddNTPs and four dNTPs were obtained from Takara Shuzo. M13mp18 and M13mp19 vectors, and Ficoll were obtained from Pharmacia. Restriction enzymes were from Toyobo, Boehringer Mannheim, Takara Shuzo, and Nippon Gene and used under the conditions recommended by the suppliers. GeneScreen Plus was from New England Nuclear and random primed DNA labelling kit was from Boehringer Mannheim. [α-32P]dATP (400 Ci/mmol and 3000 Ci/mmol) and M13mp8 cut at Smal site and dephosphorylated were from Amersham.

RESULTS

Nucleotide sequence of pDG1

Almost the entire sequence of pDG1 was determined by the method of di-deoxy chain-termination, using M13mp phages and pUC plasmids (15, 16). However, pDG1 contains closely situated, long inverted repeats (IRs) (7) and the instability of such IRs in Escherichia coli has been reported (21, 22). Thus, as we expected, we were unable to clone a 21-bp HaeIII fragment located at the center of the IRs into M13mp phages or pUC plasmids. Accordingly, this region was sequenced by the standard chemical method (17) using the 218-bp Clal-Clal fragment (see MATERIALS AND METHODS). Figure 1 shows the strategy for sequencing pDG1.

Figure 2 shows the complete nucleotide sequence of pDG1. The sequence has the following composition: A, 1720; T, 1458; G, 672; and C, 589 (4439 bp in all). Thus, the GC content is 28%, a value close to the overall GC content of the chromosomal DNA of cellular slime molds (23, 24).

The plasmid pDG1 contains closely situated, long IRs of 551 bp and 552 bp (Figure 2, IR1-left and IR1-right, respectively), which are identical except for one mismatch (positions 563 and 3877) and the deletion of a single A residue (at any position from 581 to 589). A central 80-bp region separates the left and right IR1 and this region includes a pair of short IR2 (25 bp) which are identical except for 3 mismatches (at positions 1 and 4439, 6 and 4434, 11 and 4429). The complete nucleotide sequence is compatible with our previous results obtained by electron-microscopic analysis and restriction-enzyme mapping (7). Computer analysis did not reveal any significant homology.
between the IRs in pDG1 and the IRs (330 bp) of DIRS-1 (8) and that of the 2-um DNA (599 bp) from yeast (25). Another set of unique sequences in the IRs consist of six nearly homologous GC-rich clusters each of which can be cleaved by HaeIII and is flanked by stretches of A or T with a consensus sequence AAAATGGCGGCCGTTTTT (Figure 2, closed triangle). Furthermore, this latter region itself consists of local short IRs.

**Open reading frame (ORF)**

By computer analysis, using the universal genetic code, a 2718-base-long ORF was found outside the IRs (Figure 2). If ATG (of which A is at position 743) is an initiation codon, the corresponding protein has a deduced sequence of 906 amino acids (103,041 daltons). There is a sequence resembling a TATA box in the upstream region of the ORF (TATTAATA at position 604). Thirteen base-pairs and sixty base-pairs downstream from the TATA box, there are stretches of A (box b') and T (box b), respectively. Two overlapping polyadenylation signals (AATAAA) occur 3-bp downstream from the termination codon TAA. As these features appear to satisfy the minimal requirements for transcription (26 and 27 for the stretch of A's), the ORF appears to have the potential for being transcribed in cells of Dictyostelium strain GA11. Moreover, use of several programs for searching for genes by content, in Staden's ANALYSEQ (20), showed that the ORF should code for a protein (data not shown). Other ORFs are not expected to be transcribed because they are short and show patterns that differ unfavorably from that of protein-coding ORFs according to the ANALYSEQ program. The pattern of codon usage in the ORF is similar to those described by Warrick and Spudich (28). A computer search for homology to other known proteins, using the National Biomedical Research Foundation (NBRF)-PDB (Release 16.0) and SWISS-PROT (Release 7.0) data banks.
has not identified any protein significantly homologous to that encoded in the ORF (data not shown).

Because of the structural similarity between the pDGl and the putative intermediate DNA of retrotransposon DIRS-1, as discussed previously (7), we searched for consensus sequences for reverse transcriptases in the deduced sequence of amino acids. In fact, the sequence Tyr-X-Asp-Asp, which is a highly conserved sequence in reverse transcriptases (29), was found near the N-terminal region (box in Figure 2). However, the region lacks the three hydrophobic amino acids that are found adjacent to the aspartic acid residues in all reverse transcriptases. Furthermore, of seven strictly conserved amino acids, only one residue outside the box matched appropriately (vertical arrowhead at R, at base 968, in Figure 2). Statistical analysis using randomized sequences of polypeptides of the same amino-acid composition did not show any positive signals with respect to homology between the protein encoded by the plasmid and reverse transcriptases (data not shown).

Figure 2. Complete nucleotide sequence of plasmid pDGl and the predicted amino acid sequence of the ORF. The numbering begins at C at the center of the long IRS. IR1-left (position 41-591), IR1-right (position 3848-4399) and IR2 (position 1-25 and 4415-4439) are indicated by arrows. Letters shown above the IR1-left and the 5'-flanking region indicate nucleotide differences between sequences of the left and right IRS. Several elements are indicated above the nucleotide sequence as follows: arrows 1, 2 and 3 indicate short, direct repeats of ACACCA, GAAGAT and AAAAAACCTA, respectively; arrow 4 indicates perfect short IRS; closed triangles indicate GC-rich clusters having a Haelll site flanked by stretches of A or T; open circles indicate putative sites of cleavage by topoisomerase I; CRE indicates the sequence homologous to the cAMP-responsive element (TGACGTCA). Boxes a, b, b' and c indicate TATA box-like sequences, stretches of T, stretches of A, and polyadenylation signals, respectively. The predicted amino acid sequence of the ORF is shown under each codon using single-letter abbreviations. Letters shown under the amino acid sequence of the ORF indicate a part of the amino acid sequences of the pol gene of Moloney murine leukemia virus (MoMLV). The YYXD box is found in all reverse transcriptases (29). Bars under the amino acid sequence of MoMLV indicate seven amino acids that are highly conserved among reverse transcriptases. The vertical arrowheads indicate four conserved amino acids in the ORF corresponding to four out of seven amino acids that are highly conserved in reverse transcriptases. Asterisks and dots show homology in the amino acid sequences of the ORF and MoMLV: asterisks indicate identical amino acids; dots show that the amino acids belong to the same group. The cleavage sites of restriction enzymes (ClaI, EcoRI, HindIII and PstI) are indicated. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.
Figure 3 Expression of pDGl genes during development of Dictyostelium GA11. Total RNA (10 μg), isolated from cells at the indicated stages of development, were separated on a formaldehyde agarose gel, transferred to a filter and hybridized with 32P-labelled pDGl DNA. 28S and 18S size markers are calf rRNAs. V, vegetative stage; EA, early aggregation stage; LA, late aggregation stage; M, mound stage; C, culmination stage.

In the carboxy-terminal regions, 28 (73%) out of 38 amino acid residues were either aspartic acid or glutamic acid.

Unique nucleotide sequences found in pDGl

Several interesting sequences were found in pDGl. One is the putative site of cleavage by topoisomerase I (open circles from positions 578 and 3846 in Figure 2) which is present at the termini of the IRs and results in two

Figure 4. Effect of cAMP on the expression of pDGl genes during development of cells of Dictyostelium strain GA11 in suspension. GA11 cells were developed in liquid shaking culture without cAMP (lanes 1-5) and with cAMP (lanes 6-9). cAMP was added at a concentration of 0.1 mM at the onset of development. Northern hybridization was performed as described in the legend to Figure 3.
sites. The sequences are highly homologous to the putative cleavage site of topoisomerase I found in the spacer region of *D. discoideum* rDNA (9, 11). The second is a cAMP-responsive element (CRE in Figure 2), similar to that reported (12). As indicated by arrows 1, 2 and 3, three short, direct repeats were found. The sequence marked by 1 (positions 1526-1588) appears to code for ten alternating proline residue. The sequence GAAGAT (arrows preceded by 2, positions 3347-3457) is present in the terminal region of the ORF and the sequence encodes a deduced amino acid sequence with high negative charge. There are also perfect inverted repeats with nine AT dinucleotides (arrows preceded by 4, positions 3737-3764).

**Gene expression of pDGl**

To examine the expression of the genes encoded by pDGl, Northern-blot hybridization analysis was performed using the total RNA fraction extracted from cells at different developmental stages. As shown in Figure 3, a single, developmentally regulated transcript was observed to reach a maximum level after 2 h of development, although there was no detectable band of this RNA in extracts of vegetative cells. In developing cells in suspension, the amount of this transcript also reached a maximum at 1.5 - 2.0 h (Figure 4, -cAMP). The poly (A)\(^+\) RNA fraction purified from crude RNA using messenger-activated paper also provided similar results (data not shown). The length of the transcript was calculated to be about 2.7 kb by comparison with calf 28S and 17S rRNAs. Since the calculated length is very close to that of the ORF (2718 bp), the transcribed RNA is probably derived from the ORF.

**Effects of cAMP and Its analogues on the expression of pDGl**

There is a CRE-like sequence in the upstream region of the ORF and the amount of the transcript reaches maximum at 2 h of development (formation of fruiting bodies is completed within 10 h). These results suggest that cAMP should regulate the expression of the transcript. In addition, the aggregating amoebae of strain GA11 respond chemotactically to cAMP (7). Figure 4 (+cAMP) shows that exogenously added cAMP inhibits the expression of the transcript. Moreover, 2'‐deoxy cAMP can substitute for cAMP in producing the inhibitory effect while 5'‐AMP and cGMP are ineffective (Figure 5).

**DISCUSSION**

In the cellular slime molds, six distinct plasmids have already been described (2, 3, 7). However, no nucleotide sequences of the plasmids or even of particular genes from these plasmids have been reported. Accordingly, this
Figure 5. Effects of cAMP and related nucleotides on the expression of pDG1 genes in Dictyostelium strain GAI1. Cells were developed in shaking culture in the absence or presence of cAMP or related nucleotides at a concentration of 0.1 mM. After 1.5 h cells were harvested and 5 µg of total RNA were analysed by Northern hybridization. Lane 1, vegetative cells; lanes 2-6, developed for 1.5 h; lanes 1 and 2, no cAMP; lane 3, cAMP; lane 4, 2'-deoxy cAMP; lane 5, 5'-AMP; lane 6, cGMP.

communication is the first to provide the complete nucleotide sequence of a plasmid from the cellular slime molds.

We discussed previously the possible relationship between pDG1 and the putative intermediate DNA of DIRS-1, an apparent retrotransposon in D. discoideum (7). However, the overall sequence of pDG1 does not resemble that of DIRS-1. It cannot be deduced from the hypothetical amino acid sequence encoded in the ORF whether or not the gene product of pDG1 may have some reverse transcriptase activity.

One of the most notable features of the plasmid is the existence of closely situated, long IRs. The sequence analysis revealed that nearly perfect IRs (IR1, 551 bp and 552 bp) were separated by shorter IRs (IR2). It is of interest to ask whether the other five plasmids from Dictyostelium also contain such highly conserved, long IRs, because several plasmids in yeast have long IRs (30). At present, the function of the long IRs of pDG1 is unknown. In the 2-µm plasmid DNA from yeast, it has been already established from experiments in vivo and in vitro that a part of the IRs and the adjacent 100-bp region function as the origin of DNA replication (31, 32). Moreover, the IRs of the 2-µm plasmid are involved in site-specific and reciprocal recombination (33). In the ciliated protozoan Tetrahymena, rDNA exists as linear and palindromic extrachromosomal molecules, resembling those of Dictyostelium (34, 35). This rDNA is replicated bidirectionally within a 600-bp region on either side of the center of the palindrome, as shown by electron-
microscopic analysis (36). Thus, the IRs region of pDGI may participate in autonomous replication.

It is noteworthy that there are putative cleavage sites by topoisomerase I in pDGI. In the rDNA of *Tetrahymena*, sites hypersensitive to DNase I, which are located at both ends of the rRNA transcription unit, correspond to the hexadecamer cleavage site by topoisomerase I that has the sequence AGACTTAG-A(A/G)AAA(A/T)(A/T)(A/T) (10). Recently, in the extrachromosomal rRNA genes of *D. discoideum*, Ness et al. (11) identified several cleavage sites for site-specific topoisomerase I which have similar hexademeric sequences, at -200 bp and -2200 bp, upstream from the site of initiation of transcription.

In the IRs of pDGI, there are six, nearly homologous, GC-rich clusters flanked by stretches of A or T. Recently, three groups obtained similar results which showed that two G-rich sequences in the 5' non-coding region function as cis-acting elements in *D. discoideum* for the transcription of the actin genes (37, 38) and the cysteine proteinase 2 gene (39, 40). In addition, the fact that the clusters have the potential ability to generate dyad symmetry, which is a common feature of control elements (41), leads us to assume that the GC-rich clusters may play some role in the regulation of gene expression. With respect to the GC-rich clusters, it is noteworthy that a deletion of 1.1 Kb occurred between two GC-rich clusters when the plasmid, or the 2.1-kb *HindIII* fragment which contains the IRs, was ligated in vectors and introduced into *E. coli* HB101 (18).

The cAMP-responsive element (CRE) has been found in the upstream region of cAMP-inducible genes in mammals and in long, terminal repeats of several leukemia viruses (12, 42, 43). However, there are no reports of CRE in the cellular slime molds. Although the effect of cAMP or 2'-deoxy cAMP on the expression of pDGI gene is inhibitory, the fact that exogenously added cAMP affects transcription of pDGI is suggestive of an involvement of the CRE-homologous sequence in the cAMP-regulated expression of genes in the cellular slime molds. 2'-deoxy cAMP, which can substitute for exogenous cAMP, has a high affinity for the cell-surface receptor but a very low affinity for cAMP-dependent protein kinase (44). Thus, exogenously added cAMP may inhibit the transcription of pDGI through interactions with cell-surface receptors for cAMP at an initial step, as discussed by others (45, 46).

The experimental results described in this report do not provide any clues as to the function of the plasmid and the protein encoded by the ORF. Recently, we introduced the long *Clal-Clal* fragment (4.2 kb), which contains
almost all the pDGl into the axenic strain AX3K of D. discoideum, using a
integration vector B10SX (47). The three AX3K transformants that contained
high numbers of copies of the hybrid plasmid contained several RNA bands
which hybridized with pDGl and whose appearance was developmentally regulated
although, in contrast to the events in the GA11 strain, the addition of cAMP
increased the accumulation of the transcripts (48). However, no differences
in the morphogenetic patterns of the three transformants, as compared to
those of the original AX3K strain, could be detected under a binocular micro-
scope. These results imply that pDGl in Dictyostelium species GA11 is like
the "molecular parasites" or "selfish DNA" described by Doolittle and
Sapienza (49).

As shown in this report, pDGl encodes only one gene which is expressed
during development. This result raises one question of major potential Inter-
est. How can the plasmid pDGl be stably partitioned into daughter cells? In
the "selfish" 2-μm DNA of yeast, the plasmid has adopted very sophisticated
strategies for its' stable maintenance (50). Because pDGl has long IRs and
putative cleavage sites by topoisomerase I, similar to those of the rDNA of
D. discoideum (35), it is conceivable that the partitioning of the pDGl is
modulated by the same maintenance mechanisms as those that are involved in
maintenance of the palindromic extrachromosomal rDNA molecules at constant
copy numbers during cell division. In fact, a 3-μm circular DNA in Saccharo-
myces cerevisiae is a rDNA plasmid possessing an ori sequence within the
single unit of the tandemly repeated rDNA gene (51, 52).

ACKNOWLEDGMENTS

We would like to thank R. Staden (MRC, Laboratory of Molecular Biology, 
Cambridge, England) for his programs, and H. Kimura (this University) for
installing the programs on VAX/780 at the Accelerator Center of this Univer-
sity. We also thank T. Miyata (Kyushu University) for the analysis of amino
acid homology between the ORF and reverse transcriptase and K. Saigo (The
University of Tokyo) for helpful discussions on retroviral genes. This work
was partly supported by grants from the Ministry of Education, Science and
Culture (Project Nos. 60105004 and 61304003), and was performed in the facili-
ties of the Gene Experiment Center of this University.

*To whom correspondence should be addressed
+Yoshima Tanaka, Institute of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki 305,
Japan
REFERENCES