The Dictyostelium Developmental cDNA Project: Generation and Analysis of Expressed Sequence Tags from the First-Finger Stage of Development

Takahiro Morio,1 Hideko Urushihara,1 Tamao Saito,2 Yoshihiro Ugawa,3 Hideaki Mizuno,1 Motonobu Yoshida,4 Ryuji Yoshino,5 Biswa Nath Mitra,1 Min Pi,1 Tomihiro Sato,1 Keiko Takemoto,5 Hiroo Yasukawa,6 Jeffrey Williams,7 Mineko Maeda,8 Ikuo Takeuchi,9 Hiroshi Ochiai,2 and Yoshimasa Tanaka1,*

Institute of Biological Sciences, Univ., Tsukuba, Tsukuba, Ibaraki 305-8572, Japan, 1 Division of Biological Science, Graduate School of Sciences, Hokkaido Univ., Sapporo, Hokkaido 060-0810, Japan, 2 DNA Information & Stock Center, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305-8602, Japan, 3 Research Institute of Food Science, Kinki Univ., Nara, Nara 631-8505, Japan, 4 Institute for Virus Research, Kyoto Univ., Kyoto, Kyoto 606-8507, Japan, 5 Faculty of Engineering, Toyama Univ., Toyama, Toyama 930-8555, Japan, 6 Department of Anatomy and Physiology, University of Dundee, Dundee DD1 5EH, UK, 7 Department of Biology, Graduate School of Science, Osaka Univ., Toyonaka, Osaka 560-0043, Japan, 8 and Novartis Foundation for the Promotion of Science, Takarazuka, Hyogo 665-0042, Japan 9

(Received 6 November 1998)

Abstract

In an effort to identify and characterize genes expressed during multicellular development in Dictyostelium, we have undertaken a cDNA sequencing project. Using size-fractionated subsets of cDNA from the first finger stage, two sets of gridded libraries were constructed for cDNA sequencing. One, library S, consisting of 9984 clones, carries relatively short inserts, and the other, library L, which consists of 8448 clones, has longer inserts. We sequenced all the selected clones in library S from their 3'-ends, and this generated 3093 non-redundant, expressed sequence tags (ESTs). Among them, 246 ESTs hit known Dictyostelium genes and 910 showed significant similarity to genes of Dictyostelium and other organisms. For library L, 1132 clones were randomly sequenced and 471 non-redundant ESTs were obtained. In combination, the ESTs from the two libraries represent approximately 40% of genes expressed in late development, assuming that the non-redundant ESTs correspond to independent genes. They will provide a useful resource for investigating the genetic networks that regulate multicellular development of this organism.

Key words: Dictyostelium discoideum; cDNA project; ESTs; multicellular development

1. Introduction

The protist Dictyostelium discoideum has been widely investigated as a useful model organism to study many basic problems in cell biology. One of the most remarkable aspects of this organism lies in its unique life cycle. Dictyostelium cells proliferate as unicellular amoeboid cells, but when their food source is exhausted, they aggregate and form a multicellular mass. It should be emphasized that the unicellular stage (growth phase) and multicellular stage (developmental stage) are clearly separated temporally and that the transition between these two stages is reversible until the onset of terminal differentiation. At the terminal differentiation stage, a fruiting body is formed with two distinct cell types, spores and stalk cells. This process includes most of the basic events of multicellular development such as intracellular signaling, cell differentiation, and pattern formation. Therefore, the elucidation of the molecular basis of morphogenesis of Dictyostelium may help us to reach a general understanding of the basic mechanisms involved in the establishment of multicellular organization in higher organisms.

In an effort to understand genetic networks involved in multicellular development, we initiated a cDNA sequencing project with the aim of identifying and characterizing genes expressed during development by analysis of expressed sequence tags (ESTs).1 For this purpose, we constructed cDNA libraries from the first finger stage...
2. Materials and Methods

2.1. Strain, culture and development of D. discoideum

D. discoideum Ax4 cells were grown axenically in HL-5 medium
3 to 3 x 10^6 cells/ml and harvested by centrifugation. After repeated washing with LPS buffer (20 mM KCl, 0.24 mM MgCl₂ and 40 mM K₂HPO₄/KH₂PO₄, pH 6.4), cells were developed on nitrocellulose filters (TOYO Advantec) saturated with LPS buffer at a density of 5 x 10^6 cells/cm² at 22°C.

2.2. Isolation of total and polyadenylated RNA

Cells were harvested at the first-finger stage (14-16 hr development) on nitrocellulose filters, washed with 20 mM K₂HPO₄/KH₂PO₄ buffer (pH 6.4) containing 20 mM EDTA and total RNA was extracted by the acid guanidium thiocyanate-phenol-chloroform method using ISOGEN reagent (WAKO Chemicals Inc.) with the standard protocol of the manufacturer. Poly(A)^+ RNA was purified using oligotex-dT30 Super (Nihon-Roche), on ISOGEN reagent (WAKO Chemicals Inc.) with the standard protocol of the manufacturer. Poly(A)⁺ RNA was recovered. Using the kit, 1-3 μg of plasmid DNA was obtained from 1 ml of culture. To measure the insert size of cDNA, the plasmid DNAs from library S were digested with SacI and ApaI, and those from library L with PstI and BamHI, and subjected to 1% agarose electrophoresis.

Sequencing reactions were performed on plasmid templates using a Perkin-Elmer 9600 or 9700 Thermal Cycler and Applied Biosystems PRISM Ready Reaction Dye Terminator Cycle Sequencing Kit or Amersham DYEnamic Direct cycle sequencing kit. The reaction was primed by the M13 forward primer. Reaction products were precipitated with 99.5% ethanol and dried under a vacuum, and were stored at -20°C in the dark until electrophoresis. For sequencing, the products were suspended in sample loading buffer (83.3% formamide, 4.2 mM EDTA and 8.3 μg/ml blue-dextran) and run on a 373A, 377, or 377XL automated DNA sequencer (Perkin-Elmer, Division of Applied Biosystems).

2.5. Sequence data analysis

The nucleotide sequence data was deposited in the database (Dicty.cDB) after correction of ambiguous basecallings and removal of vector and primer/adapter sequences, and also in the GenBank/EMBL/DDBJ databases.

2.6. Construction of cDNA libraries

DNA synthesis was performed using the SuperScript Plasmid System (GIBCO BRL, Life Technologies, Inc.) following the standard protocol of the manufacturer. Using 5 μg of poly(A)^+ RNA, the first strand cDNA was primed by an oligo (dT)₁₅ primer carrying a NotI adapter. After the second strand synthesis of cDNA, the termini of the cDNAs were made blunt by T4 DNA polymerase. SalI adapters were ligated to both ends, and the NotI site at the 3'-end was generated by NotI digestion. The cDNAs were size-fractionated on a cDNA Size Fractionation Column (BRL) to recover the fraction over about 0.4 kb (fraction S). The size fraction over about 1.0 kb (fraction L) was obtained by 1% Sea-Plaque GTG agarose (FMC) gel electrophoresis. The recovery of cDNAs from agarose was performed using GELase (Epicentre Technologies Corp.) according to the standard protocol of the manufacturer.

The cDNAs included in fraction S were ligated to the SalI-NotI sites of pBluescript II KS- (Stratagene) and transformed into Escherichia coli DH5α by electroporation with E. coli Pulser (Bio-Rad) set at 18 kV/cm to generate library S which contained more than 6.8 x 10⁵ recombinant clones. Complementary DNAs of fraction L were ligated into the corresponding sites of pSPORT1 (GIBCO BRL, Life Technologies, Inc.) and transformed into E. coli DL795 cells [K12 SH28 Dhsd, mcrBC, mrr; e14 (mcrA)0; slc201; recA::Cm⁺; supE44], kindly provided by Dr. David Leach in University of Edinburgh and Dr. Masashi Fukuzawa in University of Dundee. In the same way, library L containing more than 1.8 x 10⁶ recombinant clones was generated.
Table 1. Summary of insert data.

|                | Library S | Library L |
|----------------|-----------|-----------|
| Total Clones   | 9984      | 3448      |
| With Insert    | 9378      | 6450      |
| Uncut          | 64        | 712       |
| No insert      | 330       | 1209      |
| No plasmid     | 212       | 77        |
| Mean Length Inserts (kb) | 0.96 | 1.31 |

The insert sizes of the plasmid DNAs were determined as in Materials and Methods. With Insert: clones with cDNA insert longer than 400 bp, Uncut: clones produced no linear fragment by restriction enzyme digestion, No insert: clones with no cDNA insert, but only containing linearized vector plasmid, No Plasmid: clones with no DNA fragment found on electrophoresis.

Figure 1. Distribution of inserted cDNA sizes in library S (panel A) and library L (panel B). The insert sizes were determined as described in Materials and Methods.

C93776–C93849, and C93851–C94532. The data were checked for overlap with the existing cDNA sequences and examined for similarities in the GenBank, EMBL, and DDBJ nucleic acid databases. They were also translated in all six reading frames, and each translation was compared with the protein sequence database Protein Information Resource (PIR) and SWISS-PROT. All similarity searches were performed using the BLAST algorithm.5

3. Results and Discussion

3.1. Construction and characterization of cDNA libraries

Poly(A)+ RNA was isolated from the first finger stage of development of D. discoideum strain Ax4. Complementary DNA was synthesized using an oligo (dT) primer. The sizes of the obtained double-strand cDNA ranged from 0.2 kb to 10 kb (data not shown). Using the size fractionated cDNAs, two libraries, termed S and L, were constructed from the fraction greater than 0.4 kb and that greater than 1 kb, respectively. Both the cDNA fractions were unidirectionally ligated into vector plasmid, pBluescript II KS- or pSPORT1, and transformed into E. coli. For library S, 9984 clones were gridded out in 96-well microtiter plates, and 8448 clones were gridded for library L.

For preparation of sequencing templates, plasmid DNAs were isolated by the 96-well alkaline lysis method. After digestion with appropriate restriction enzymes to excise cDNA inserts, the distribution of insert sizes was determined by electrophoresis on 1% agarose (Table 1). The size distribution of the inserts showed a peak between 0.5 kb and 1 kb (Fig. 1A), and the mean length of the inserts was estimated to be 0.96 kb. For library L, the inserts could be determined for 6450 clones. The mean insert size was estimated to be 1.31 kb. As shown in Fig. 1B, the major peak of the inserts was present between 1.2 and 1.5 kb, although an additional small peak is also seen between 0.5 and 0.8. Insert size analysis of the two libraries apparently indicates that libraries S and L, respectively, represent the classes of genes coding for relatively short and long mRNAs.

Sixty-four clones in library S and 712 clones in library L, which are represented as ‘Uncut’ in Table 1, apparently showed the plasmid sizes containing inserts, but no inserts were generated by restriction enzyme cleavage. This might be due to partial deletion or rearrangement of cDNA inserts in E. coli cells. As some of these clones are expected to retain part of the cDNA inserts, they were also subjected to sequencing.
Table 2. Summary of sequenced clones.

|                      | Library S | Library L |
|----------------------|-----------|-----------|
| Total ESTs           | 7663      | 1132      |
| Non-redundant ESTs   | 3093 (40.4%) | 471 (41.6%) |
| Known Dictyostelium genes | 246 (8.0%) | 78 (16.5%) |
| Homologs             | 910 (29.4%) | 159 (33.8%) |
| Novel genes          | 1937 (62.6%) | 234 (49.7%) |

Non-redundant ESTs: EST groups considered to be derived from independent genes, Known Dictyostelium genes: genes showing significant similarity to Dictyostelium genes in the GenBank, EMBL and DDBJ nucleic acid databases with P-value < e-70, Homologs: genes whose longest open reading frame showed similarity to proteins of Dictyostelium or other organisms in the PIR and SWISS-PROT protein sequence databases with P-value < e-10.

3.2. Sequence analysis of library S

Clones expected to have cDNA inserts (‘With Insert’ and ‘Uncut’ in Table 1) were sequenced from their 3’-ends. Sequences less than 150 bp or those with more than 2% ambiguity were rejected. Among the 9444 clones sequenced, significant sequence data as ESTs that might permit identification of genes as ESTs were obtained from 7663 clones (81.2%, see Table 2). The obtained ESTs were compared with each other using the BLAST algorithm, and those showing similarity with a P-value < e-60 were clustered into non-redundant EST groups. Out of 7663 ESTs, 3093 non-redundant groups were identified. Then the ESTs were examined for homology in the GenBank, EMBL and DDBJ nucleic acid databases. In addition, the sequences were translated into three possible reading frames, and the longest open reading frames were compared with the Protein Information Resource (PIR) and SWISS-PROT protein sequence databases. Out of 3093 independent groups, 246 groups were assigned to previously sequenced Dictyostelium genes and 910 groups to those with significant similarity (P-value < e-10) to genes of Dictyostelium or other organisms (Table 2). Assuming that the non-redundant EST groups correspond to independent genes, the frequency of gene appearance in library S was indicated in Fig. 2A. Among the independent genes, 2559 genes were represented by either one or two ESTs. As listed in Table 3, the cDNA sizes of the genes with the highest frequency of appearance were relatively short and ranged from 0.6 kb to 1.2 kb. This is consistent with the view that library S mainly represents genes encoding relatively short mRNAs, and thus the frequency of gene appearance in this library may reflect the expression profile of shorter genes.

3.3. Sequence analysis of library L

A total of 1262 clones from library L were randomly sequenced from their 3’-ends, and 1132 ESTs (89.7% of the sequenced clones, Table 2) derived from 471 non-redundant EST groups were obtained. Searching against nucleotide and amino acid sequence databases revealed that the EST groups hit 78 known Dictyostelium genes, homologs of 159 known genes and 234 unknown genes (Table 2). The profile of frequencies of gene appearance was similar to that of library S: 88.3% (416) of the groups were represented by just one or two EST(s) (Fig. 2B). However, as shown in Table 4, the actin 15 gene, the elongation factor 1α gene (efaA), several prespore-specific genes (Dp87, cotA, cotB, cotC, psvA and pspD) and the cAMP-inducible cysteine protease gene (cprB), which are known to be highly expressed during development, were more abundant. Among them, efaA, Actin 15, Dp87 and cotA comprise...
Table 3. Genes frequently appeared in library S.

| Representative clone | Gene    | Accession No. | mRNA size (kb) | Abundance (%) | References |
|----------------------|---------|---------------|----------------|---------------|------------|
| SSA101               | dutA    | D16417        | 1.2            | 3.8           | 15         |
| SSA360               | BP74*   | A34095        | —              | 1.4           | 16         |
| SSA238               | cdpA    | X82784        | 0.6            | 1.3           | 17         |
| SSA357               |         | —             | —              | 0.9           | —          |
| SSA420               | PEPT2*  | Q63424        | —              | 0.8           | 18         |
| SSA419               | rasB    | M96622        | 0.7            | 0.7           | 19         |
| SSA362               | ubqA    | M23752        | 1.1            | 0.6           | 20         |
| SSA153               | ORAB-1* | P22125        | —              | 0.6           | 21         |
| SSA225               |         | —             | —              | 0.6           | —          |
| SSA321               |         | —             | —              | 0.6           | —          |
| SSA592               | vatP    | X90516        | 0.7            | 0.6           | 22         |
| SSA202               |         | —             | —              | 0.6           | —          |
| SSA213               | cprB    | X03344        | 1.3            | 0.6           | 13         |
| SSA116               | pkIA    | U61986        | 0.7            | 0.6           | —          |
| SSB562               |         | —             | —              | 0.6           | —          |
| SSA132               |         | —             | —              | 0.6           | —          |

Among 7666 ESTs of library S genes whose abundance is more than 0.5% were listed. Asterisks show putative homologs and their accession number in the PIR or SWISS-PROT protein sequence databases is shown.

Table 4. Genes frequently appeared in library L.

| Representative clone | Gene            | Accession No. | mRNA size (kb) | Abundance (%) | References |
|----------------------|-----------------|---------------|----------------|---------------|------------|
| SLA107               | efaA            | X55972        | 1.8            | 10.8          | '6         |
| SLA105               | Actin 15        | M14146        | 1.3            | 6.7           | 7          |
| SLA106               | Dp87            | D13973        | 2.0            | 5.7           | 8          |
| SLA112               | cotA            | X16491        | 2.2            | 4.5           | 9          |
| SLA111               | 60S ribosomal protein L3 | L08391 | 1.3 | 2.3 | 23 |
| SLA246               | 60S ribosomal protein L1* | P49691 | — | 1.6 | 24 |
| SLA121               | sahA            | M19903        | 1.6            | 1.6           | 25         |
| SLA116               | psA             | X37149        | 2.2            | 1.3           | 10         |
| SLA195               | cotC            | M26239        | 1.8            | 1.3           | 11         |
| SLA123               | cotB            | M26238        | 2.2            | 1.0           | 11         |
| SLA203               | pspD            | U25144        | 2.1            | 1.0           | 12         |
| SLA267               | cprB            | X03344        | 1.3            | 1.0           | 13         |
| SLA193               | kspB            | X75263        | 2.4, 2.6       | 1.0           | 26         |

Among 1132 ESTs of library L genes whose abundance is more than 1.0% are listed. Asterisks show putative homolog and their accession number in the PIR or SWISS-PROT protein sequence databases is shown.

For further characterization of the genes, we are carrying out the following investigations: (1) determination of spatial and temporal expression profile of the identified genes by probing the high-density gridded filters on which the cDNA clones were blotted and in situ hybridization experiments,14 and (2) analysis of the function of individual genes by antisense mutagenesis, gene disruption and ectopic expression.

Acknowledgments: We thank Hideki Nagasaki and Dr. Kimiko Yamamoto of STAFF-Institute and Dr. Takuji Sasaki at National Institute for Agrobiological Resources for technical advice and the use of facilities. We also thank the following individuals for their excellent technical assistance: Keiko Hashimoto, Antony Lizamma, Masako Matsuda, Hitomi Matsuzaki, Miyuki Nakajima, Nobuko Nakamura, Satoko Saito, Naoya Sakuragi, Nobuko Sekiguchi, Keiko Shinhoara, Reiko Shintani, and Masako Yokoyama. Finally, we deeply thank Dr. Naotake Ogasawara in Nara Institute of Science and Technology for his continuous support and encouragement. This study was supported by research grants from Research For the Future (JSPS-RFTF96L00105) of Japan Society for the Promotion of Science and also supported by a Grant-in-Aid for Scientific Research on Priority Areas from Ministry of Education, Science, Sports and Culture of Japan (08283105).
References

1. Adams, M. D., Kelly, J. M., Gocayne, J. D., Dubnick, M. et al. 1991, Complementary DNA sequencing: expressed sequence tags and human genome project, Science, 252, 1651–1656.
2. Blumberg, D. D. and Lodish, H. F. 1980, Changes in the messenger RNA population during differentiation of Dictyostelium discoideum, Dev. Biol., 78, 285–300.
3. Watts, D. J. and Ashworth, J. M. 1970, Growth of myxamoebae of the cellular slime mold Dictyostelium discoideum in axenic culture, Biochem. J., 119, 171–174.
4. Kuribayashi, K., Hikata, M., Hiraoka, O., Miyamoto, C., and Furuchi, Y. 1988, A rapid and efficient purification of poly-(A)-mRNA by oligo (dT)30-Latex, Nucleic Acids Symp. Ser., 19, 61–64.
5. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990, Basic local alignment search tool, J. Mol. Biol., 215, 403–410.
6. Blumberg, D. D. 1994, Identification of a new spore coat genes SP60 and SP70 of Dictyostelium discoideum, Mol. Cell. Biol., 14, 8853–8866.
7. Escalante, R. and Loomis, W. F. 1995, Whole-mount in situ hybridization of cell-type-specific mRNAs in Dictyostelium, Dev. Biol., 171, 262–266.
8. Yoder, B. K., Mao, J., Erdos, G. W., West, C. M., and Chisholm, R. L. 1989, Expression and organization of BPT74, a cyclic AMP-regulated gene expressed during Dictyostelium discoideum development, Mol. Cell. Biol., 9, 4170–4178.
9. Tasaka, M., Hasegawa, M., Ozaki, T., Iwabuchi, M., and Inui, K. 1996, Molecular cloning and tissue distribution of rat peptide transporter PEPT2, Biochim. Biophys. Acta, 1280, 173–177.
10. Daniel, J. M., Spiegelman, G. B., and Weeks, G. 1993, Characterization of a third ras gene, rasB, that is expressed throughout the growth and development of Dictyostelium discoideum, Oncogene, 8, 1041–1047.
11. Coukell, B., Moniakis, J., and Grinberg, A. 1995, Cloning and expression in Escherichia coli of a cDNA encoding a developmentally regulated Ca(2+) binding protein from Dictyostelium discoideum, FEBS Lett., 10, 362, 342–346.
12. Ohmachi, T., Giorda, R., Shaw, D. R., and Ennis, H. L. 1989, Molecular organization of developmentally regulated Dictyostelium discoideum ubiquitin cDNAs, Biochemistry, 28, 5226–5231.
13. Ngsee, J. K., Elferink, L. A., and Scheller, R. H. 1991, A family of ras-like GTP-binding proteins expressed in electromotor neurons, J. Biol. Chem., 266, 2675–2680.
14. Xie, Y., Coukell, B., and Gombos, Z. 1996, Antisense RNA inhibition of the putative vacuolar H+ ATPase proteolipid of Dictyostelium reduces intracellular Ca2+ transport and cell viability, J. Cell Sci., 109, 489–497.
15. Steel, L. F., Farnum, P. D., and Kunapoli, P. 1995, Sequence and developmental regulation of the gene that encodes the Dictyostelium discoideum L3 ribosomal protein, Gene, 162, 123–128.
16. Rafti, F., Gargiulo, G., Manzi, A., Malva, C., and Graziani, F. 1989, Sequence of the ribosomal protein cDNA of D. melanogaster homologous to the L1 ribosomal protein gene of X. laevis, Nucleic Acids Res., 17, 456.
17. Guittion, M. C., Part, D., and Veron, M. 1988, Cloning of a cDNA for the S-adenosyl-L-homocysteine hydrolase from Dictyostelium discoideum, Biochimie, 70, 835–840.
18. Haus, U., Trommler, P., Fisher, G. R., et al. 1993, The heat shock cognate protein from Dictyostelium affects actin polymerization through interaction with the actin-binding protein cap32/34, EMBO J., 12, 3769–3771.