Review

Functional cDNA expression cloning: Pushing it to the limit

By Hiroto OKAYAMA*1,†

(Communicated by Shigetada NAKANISHI, M.J.A.)

Abstract: The 1970s and the following decade are the era of the birth and early development of recombinant DNA technologies, which have entirely revolutionized the modern life science by providing tools that enable us to know the structures of genes and genomes and to dissect their components and understand their functions at the molecular and submolecular levels. One major objective of the life sciences is to achieve molecular and chemical understandings of the functions of genes and their encoded proteins, which are responsible for the manifestation of all biological phenomena in organisms. In the early 1980s, I developed, together with Paul Berg, a new technique that enables the cloning of full-length complementary DNAs (cDNAs) on the basis of their functional expression in a given cell of interest. I review the development, application and future implications in the life sciences of this gene-cloning technique.

Keywords: cDNA cloning, method development

Introduction

In September 1978, after studying as a research associate in Professor Osamu Hayaishi’s laboratory following graduation from Kumamoto University Medical School and the Ph.D. course at Kyoto University, I went to the Department of Biochemistry at Stanford University Medical School and began a post-doctoral research study under Professor Paul Berg, the father of recombinant DNA technology. At that time, exciting progress on several research projects was taking place in Professor Berg’s laboratory. The expression of rabbit O-globin in African green monkey kidney cells with a recombinant SV40 virus vector had just been accomplished.1) An element in the SV40 early promoter that augments the transcription of the rabbit β-globin gene in non-erythrocytes when placed near the promoter region was being discovered; this was later recognized to be one of the transcriptional enhancer elements that were being concurrently and independently discovered in various genes by others.2) Meanwhile, the development of mammalian drug-selection markers was progressing.3,4) Professor Berg was interested in making rabbit hemoglobin in monkey cells. To do so, rabbit α-globin had to be co-expressed with the β-globin in the target cells. However, nobody had succeeded in cloning a full-length α-globin complementary DNA (cDNA) copy of the mRNA. Dr. Berg asked me to clone a full-length α-globin cDNA that coded for the entire protein; however, there was a rumor that unlike the β-globin counterpart, a full size α-globin cDNA was difficult to clone. This was the beginning of the development of the functional cDNA expression cloning strategy. I decided first to develop from scratch a new cDNA cloning method that would allow for the high efficiency cloning of full-length cDNA of any size. For me, this was a significant challenge, but I was determined because I knew that technological advances always precede great progress in any field of science.

The cDNA cloning methods under development

Although currently retroviral reverse transcriptases are routinely used to synthesize cDNA copies of various mRNAs, it was S. Spiegelman and his associates who first showed that the reverse transcriptase contained in avian myeloblastosis virus (AMV) particles has no apparent template specificity to particular RNAs and can be used to synthesize...
cDNAs of any RNAs. Subsequently, several groups demonstrated that synthetic oligo(deoxythymidy- late) [oligo(dT)] can efficiently prime for the reverse transcription of mRNAs because of their 3′ end extension with a poly(A) tail and used radiolabeled cDNA copies as probes to quantify the levels of particular mRNAs in cells. Meanwhile, the modification of naturally occurring E. coli plasmids and the use of those modified plasmids as vectors for the cloning of DNA fragments were being developed.

Although radiolabeled cDNA probes are useful for the quantitative analysis of their mRNA counterparts and for the detection of their genes in genomes, the radiolabeled cDNA probes are not useful for knowing the primary structures of the encoded proteins. For such purposes, the monoclonal expansion of particular cDNA sequences in E. coli is needed. Unlike higher eukaryotes’ chromosomal genes, which are much longer because of the presence of multiple introns and huge repetitive sequences, cloned cDNAs are simple in structure, easy to genetically manipulate and versatile in application. Cloned cDNAs are absolutely essential for knowing the structures of the proteins encoded by particular genes and for expressing the encoded proteins in various organisms and cells to understand their biological functions in vivo and to produce the proteins in large quantity for biochemical studies and clinical applications.

The importance of the monoclonal expansion of cDNA copies of any given mRNAs was recognized early, and several groups attempted to develop a way to insert the generated cDNA copies into a drug-selectable plasmid vector and expand them in E. coli. It was the research group led by B. Mach at the University of Geneva that first reported the isolation of E. coli colonies that harbored the plasmid carrying cDNA fragments and the successfully cloning of a cDNA copy of rabbit β-globin mRNA; the procedure they developed is illustrated in Fig. 1. First, a DNA strand complementary to the mRNA is synthesized by oligo(dT)-primed reverse transcription with the AMV enzyme. Following the removal of the RNA template by alkali density gradient centrifugation and subsequent poly(dT) tailing at the 3′ end of the cDNA with calf thymus terminal deoxynucleotidyl transferase (TdT), the second strand cDNA is synthesized with an oligo(dA) primer and E. coli DNA polymerase I. The unpaired 3′ portion of the poly(dT) tail is then removed with single strand-specific S1 nuclease. The resulting double-stranded cDNA is oligo(dG)-tailed with TdT and inserted into the kanamycin resistance gene-harboring pCR1 plasmid that has been linearized by EcoRI digestion, blunt ended by E. coli DNA polymerase I-mediated gap-filling DNA synthesis and oligo(dC)-tailed. The final step is the insertion of the cDNA into the pCR1 plasmid via oligo(dG):oligo(dC) pairing, subsequent transfection into E. coli and selection for kanamycin-resistant colonies.

Soon after this method was reported, two research groups reported variants of the Mach procedure. One group used a method to avoid second strand DNA synthesis and inserted a cDNA into an E. coli plasmid without second strand DNA synthesis. To achieve this type of insertion, a pair of two EcoRI-linearized E. coli plasmid DNAs, one that had poly(dA) tails at both 3′ ends and the other that had poly(dT) tails at both 3′ ends, was used (Fig. 2). In this procedure, the poly(dT)-tailed mini-ColE1 plasmids serves as primers for cDNA synthesis with AMV reverse transcriptase. Poly(dT) tails are then added to the 3′ ends of the cDNAs with TdT. Finally, the product is annealed with the poly(dA)-tailed mini-ColE1 plasmid to obtain a circular mini-ColE1 head-to-head or head-to-tail dimer with the insertion of two cDNAs:mRNA hybrids through poly(dA):poly(dT) pairings. Although this strategy helps avoid the necessity of complicated second strand DNA synthesis, it risks the loss of clones because the hybrid region is highly prone to nuclease attacks during transfection of the product into E. coli and once the cDNA strand in the hybrid is nicked, there is no way to repair the nicks in E. coli. Moreover, the head-to-head dimers have a 50% probability of being produced and are generally unstable and are degraded in E. coli.

The other method (Fig. 3), which was independently developed by two groups led by T. Maniatis and W. Salser, is a simplification of the Mach method and uses the hairpin structure, which is spontaneously formed by single-stranded cDNA after strand separation by heat denaturation, as a primer for second strand cDNA synthesis with E. coli DNA polymerase I, which automatically removes any unpaired 3′ end by its 3′-to-5′ proof-reading exonuclease activity. Subsequently, the hairpin structure is opened with S1 nuclease, which was used to trim the double-stranded cDNA ends in the Mach method. Although this method highly simplifies second strand DNA synthesis, the 3′-end sequence of the original cDNA that corresponds to the 5′-end...
sequence of mRNA is ultimately lost. However, this method gained popularity because of its simplicity.

The development of a high-efficiency cDNA cloning method

There are three major shortcomings in these procedures. First, obtaining a full-length cDNA copy of a relatively long mRNA is difficult because of the quality of the mRNA that is used for reverse transcription. Moreover, if self-priming is used for second strand cDNA synthesis, a loss of a cDNA sequence corresponding to the N-terminal region of the protein is imminent as noted above. Such 5' truncated cDNAs are useless for expressing the entire encoded protein for the analysis of its structure and function. Second, the bidirectional insertion of the
cDNA into the plasmid vector can be problematic. This is not a problem if oligonucleotide hybridization\textsuperscript{21} or antibody detection\textsuperscript{22–24} is intended to be used to screen for a particular cDNA clone, but it is not a trivial issue if the use of selection or the screening of cDNA clones based on the functions of the proteins they encode is envisioned to isolate a particular cDNA clone. Third, a relatively large quantity of mRNA is required to obtain a cDNA library of even moderate size because of overall low cloning efficiencies.

1. Initial ideas. To develop a method to overcome these three shortcomings, I designed an outline of the method with new ideas by performing

---

**Rabbitts Method**

![Diagram](image)

Fig. 2. The cDNA cloning strategy invented by Rabbitts. This method is unique for inserting a synthesized cDNA into the plasmid without converting it to a double-stranded cDNA. The mini-ColE1 plasmid is linearized by EcoRI digestion. One half is poly(dT)-tailed with TdT and the other half is poly(dA)-tailed. The poly(dT)-tailed mini-ColE1 is then used as a primer for the reverse transcription of mRNA. The resulting linear mini-ColE1 that is linked to a mRNA-cDNA hybrid at both ends is then tailed with poly(dT) with TdT and annealed to the poly(dA)-tailed mini-ColE1 plasmid for circularization.
rigorous experiments to test the validity of the ideas one by one. The first experimental design of the method is illustrated in Fig. 4. To clone only full-size cDNAs, the idea of a HindIII linker attached to the 5’-end of mRNA was initially tested. In this protocol, the mRNA is digested with tobacco acid pyrophosphatase to cleave off the Cap structure and generate a 5’ monophosphate-linked mRNA. Second, a 3’-terminally cytosine ribonucleotide (rC)-added synthetic deoxyribonucleotide decamer with a HindIII recognition sequence is then ligated to the 5’-end of the mRNA with a T4 phage RNA

Fig. 3. The cDNA cloning method reported by the Salser group. This method eliminates the use of an exogenous primer for second cDNA strand synthesis, which simplifies double strand cDNA synthesis. The first strand cDNA synthesis with reverse transcriptase is primed with oligo(dT) and following heat denaturation to separate the mRNA-cDNA hybrid, the single-stranded cDNA frequently forms a hairpin structure with the 3’-end folding back and pairing with its body. This 3’-end serves as the primer for the second cDNA strand synthesis with E. coli Pol I. The hairpin is then digested with S1 nuclease to generate an open-ended double-stranded cDNA, which is subsequently trimmed at the 5’-ends with λ-exonuclease, poly(dA)-tailed with TdT and inserted into the poly(dT)-tailed pMB9 or pSC101 plasmid.
ligase because this enzyme ligates only the 3' and 5' ends of RNA.\(^{26,27}\) This decamer is actually a commercially available HindIII linker that was routinely used to construct a HindIII cohesive end at an end of double-stranded DNA, and the addition of one rC to the 3-end of the decamer was performed with TdT and CTP because this enzyme can use ribonucleotide triphosphates as a substrate and DNA molecules as a target for nucleotide transfer, but it cannot recognize DNA as a target any longer after one ribonucleotide is added to the 3'-end of the DNA.

The resulting HindIII linker-attached mRNA is then used as a template for cDNA synthesis with a poly(dT)-tailed BamHI decamer linker as a primer for reverse transcription. Subsequent second strand DNA synthesis was performed with RNase H to cleave the RNA strand, E. coli DNA polymerase I to perform repair DNA synthesis and E. coli DNA ligase to close the nicks. Finally, double digestion with HindIII and BamHI would create a full-length double-stranded cDNA with HindIII and BamHI cohesive ends at the 5'- and 3'-termini, respectively, if reverse transcription reached the end of the added HindIII sequence. Consequently, only a full-length cDNA would be inserted into an appropriate E. coli plasmid with a HindIII and BamHI replacement insertion site. In this strategy, HindIII and BamHI cut specific 6-base sequences that occur once every 4–5 kb, which are infrequent but not rare and can be present in the mRNA aimed to be cloned. If this strategy works, this problem could be mitigated by using 8 base-cutting restriction enzyme sites, which are estimated to be present once every 65 Kb in mRNA.
As soon as I began to perform the pilot experiments with the rabbit reticulocyte mRNA prepared by routine phenol extraction, I quickly discovered some serious problems that were associated with this experimental protocol. Treatment of the mRNA with tobacco acid pyrophosphatase and subsequent RNA ligation to add a HindIII linker at the 5'-end greatly decreased the reverse-transcribability of the mRNA partly due to the aggregation of proteins with the mRNA. Moreover, the overall efficiency of the addition of the HindIII linker was very low. Despite an extensive search for optimal conditions and re-extraction of the RNA with phenol, no satisfactory results could be obtained. Furthermore, these commercially available enzyme preparations were never RNase-free. I therefore abandoned strategies that used any enzyme treatments of thoroughly deproteinized mRNA including the decapping and the ligation to the HindIII linker although this strategy was later used successfully to determine the 5'-end sequence of mRNA in combination with DNA amplification with the polymerase chain reaction.28)

2. The finalized procedure. After struggling for a couple of years, I developed the current protocol that fulfilled all of the requirements we identified for the cloning of full-size cDNAs that would allow the expression of the encoded proteins in any given organisms when an appropriate promoter and the sequences necessary for the proper processing of the transcribed RNA are placed in the vector plasmid (Fig. 5).29) In this protocol, two specially modified plasmid DNA fragments have to be prepared prior to

---

Fig. 5. Finalized method. (A) The preparation of special DNA reagents. The vector primer to be used as a primer for the reverse transcription of mRNA is prepared by KpnI digestion of the pBR322-SV40(0.71–0.86) plasmid DNA, poly(dT) tailing with TdT, subsequent HpaI digestion and purification of the poly(dT)-tailed large plasmid fragment by agarose gel electrophoresis and adsorption to oligo(dA)-cellulose. The oligo(dG)-tailed linker DNA that is to be used for the circularization of the mRNA:cDNA hybrid linked to the plasmid vector is prepared by PstI digestion of pBR322-SV40(0.19–0.32), oligo(dG) tailing, HindIII digestion and purification by agarose gel electrophoresis. (B) The steps of cDNA cloning with vector primer and oligo(dG)-tailed linker DNA. The cDNA is synthesized with vector primer for priming reverse transcription. The resulting cDNA:mRNA hybrid that is linked to the vector primer is oligo(dC)-tailed with TdT and digested with HindIII endonuclease. The resulting oligo(dG)-tailed cDNA:mRNA hybrid that is linked to the vector primer with a newly created HindIII cohesive end is then mixed with oligo(dG)-tailed linker DNA and incubated with *E. coli* DNA ligase for circularization. Finally, the RNA strand in the cDNA:mRNA hybrid segment is converted to DNA by repair DNA synthesis with *E. coli* RNase H, Pol I and DNA ligase.
cDNA cloning. One plasmid is the “vector primer,” a linear pBR322 containing a poly(dT) tail at one end (Fig. 5). This fragment is prepared from the pBR322-SV40 hybrid plasmid, in which the tetracycline-resistance gene between the HindIII and PvuII sites in the pBR322 is replaced with the SV40 PvuII-HindIII (map unit, 0.71–0.86) fragment that contains a single KpnI and Hpal site, the former of which provides the ends for poly(dT) tailing with TdT and the latter provides the digestion site for the removal of one poly(dT) tail from the plasmid DNA. After corresponding treatments, the large pBR322 DNA product is purified by agarose gel electrophoresis and oligo(dA)-cellulose column chromatography, the latter procedure is to remove untailed pBR322 DNA. The other product is “oligo(dG)-tailed linker DNA,” which has an oligo(dG) tail at one end and a HindIII cohesive end at the other end. This DNA is also prepared from another pBR322-SV40 hybrid plasmid containing a PstI site in the SV40 region by PstI restriction endonuclease digestion, oligo(dG) tailing with TdT, HindIII digestion and purification by agarose gel electrophoresis. Although we used the pBR322-SV40 hybrid DNAs to prepare these vector primer and oligo(dG)-tailed linker DNA, any plasmid DNAs containing a HindIII site, a unique restriction enzyme digestion site and a KpnI or PstI site in this order outside the E. coli plasmid DNA can be used.

Step 1 is cDNA synthesis with the vector primer as a primer for reverse transcription with AMV reverse transcriptase. Step 2 is oligo(dC) tailing at the 3′-ends of the cDNA and the vector with TdT. Step 3 is HindIII digestion to remove the oligo(dC) tail end and to generate a HindIII cohesive end at the vector terminus. We chose this restriction enzyme digestion because unlike EcoRI and SalI, HindIII restriction endonuclease does not cleave DNA-RNA hybrids so that cDNAs containing a HindIII site can be cloned with this protocol. Step 4 is annealing and circularization with oligo(dG)-tailed linker DNA and E. coli DNA ligase. Step 5 is second strand DNA synthesis with RNase H to nick the mRNA strand, E. coli DNA polymerase I to replace the RNA strand with DNA by repair-type DNA synthesis and E. coli DNA ligase to seal nicks, yielding closed circular plasmid DNA with a double-stranded cDNA inserts. The last step is transfection of E. coli with the product followed by selection for ampicillin resistance. Using this method for cloning rabbit globin cDNAs, we obtained roughly 100,000 colonies containing recombinant α- or β-globin cDNAs per 1µg mRNA, of which 30–50% were full-length or nearly full-length. Thus, this method enables not only high-efficiency cloning of full-length cDNAs with a small amount of mRNA but also unidirectional insertion of cDNA in the vector plasmid as intended so that the placement of an appropriate transcriptional promoter, a splice junction and a transcription termination signal upstream and downstream of the insert, respectively, should allow the cDNA-encoded full-size protein to be expressed after delivery of the recombinant plasmid DNA into cells.

The next step was the placement of the elements that are necessary for transcription of the cDNA inserts and appropriate processing of the transcripts in mammalian cells. Basically, the plasmid for the preparation of the vector primer is required only to contain a fragment composed of one HindIII site, a unique restriction site and a 3′-protruding rare restriction site in this order, whereas the plasmid for the preparation of the linker DNA is required to contain a fragment composed of a 3′-protruding unique restriction site and one HindIII site. I decided to use a mosaicked SV40 segment containing the T antigen gene polyadenylation signal for the construction of the vector primer plasmid (pcDV1) and a similarly mosaicked SV40 segment containing the T-antigen promoter and the truncated late splice junction for the construction of the linker DNA plasmid (pL1) (Fig. 6). When the vector primer and the linker DNA are prepared from these plasmids and used for cDNA cloning as described in Fig. 5, a full length cDNA inserted in the pcD expression vector should be automatically constructed (Fig. 6 left lower figure). The cDNA is expected to be transcribed from the strong SV40 early promoter, processed with proper splicing and polyadenylation and finally translated into a functional protein upon delivery into mammalian cells. In addition to these modifications to construct the cDNA expression libraries, we adopted the guanidine thiocyanate solubilization-CsTFA high speed centrifugation method, a modification of the guanidine-HCl solubilization method, instead of regularly used phenol extraction to prepare protein-free high quality mRNA from cultured cells or animal organs. This modified method was later proved to be highly effective for preparing readily reverse-transcribable RNA even from human body specimen.

3. First validation of the procedure. The first report that verified the merit of the procedure was the isolation of a full-length functional cDNA for human hypoxanthine-guanine phosphoribosyl transferase (HGPRT) defective in Lesh-Nyhan syndrome.
patients, which was performed by Ted Friedmann and his group.\(^{35}\) The group initially identified a human Alu-containing genomic DNA fragment that complemented HGPRT-defective mouse fibroblasts,\(^{36}\) and then used the genomic fragment to isolate the full-length human HGPRT cDNA by colony hybridization from the human fibroblast cDNA library constructed with the pcD vector. They demonstrated that the cDNA in the vector expresses functional HGPRT and directly complements the HGPRT-defective mouse fibroblasts. Subsequently, full-length or nearly full-length cDNAs for several hundred genes were isolated from the human or mouse pcD libraries by using the same strategy as above or simple nucleic acid hybridization with probes obtained by conventional methods.

Furthermore, the success of this cloning method inspired the development of a simple and efficient method for conventional cDNA cloning by combining classical first strand cDNA synthesis with the RNase/PoI/DNA ligase-mediated second strand DNA synthesis used in our method\(^{37}\) and much later a strategy for cloning full-length cDNAs covering the entire mRNA sequences, in which prior to first strand cDNA synthesis the cap structure at their 5'-end is replaced with a RNA oligomer by using tobacco acid pyrophosphatase and RNA ligase.\(^{38}\) Since then, both methods were extensively used in the Human Genome Project for cDNA sequencing and database construction.

Although the full-length expressible cDNA libraries are extremely useful even when conventional

---

**Fig. 6.** Structure of pcD-X and pcD2-X. The pcD and pcD2 vectors are designed to clone full-length cDNAs based on the phenotypes they express upon transfection into appropriate eukaryotic cells. pcDV1 is the precursor plasmid for the preparation of common vector primer. This modified vector primer contains the SV40 early gene polyadenylation signal (a dark green segment). pL1 and pL2 are for the preparation of oligo(dG)-tailed Linker DNAs according to the same procedure as for its original. The pL1-derived linker DNA contains the SV40 early promoter and the late 16S splice junction to drive the expression of the inserted cDNA, whereas the pL2-derived linker DNA additionally contains a drug-selectable neomycin transcriptional unit driven by the same SV40 promoter and polyadenylation signal upstream of the SV40 promoter and splice junction for cDNA expression. pcD-X and pcD2-X depict the cDNA-harboring plasmids produced with the pcDV1-derived vector primer and pL1- and pL2-derived linker DNAs, respectively, according to the procedure shown in Fig. 5.
hybridization probes are available, my original intention for the construction of such a method and libraries was to clone a particular cDNA that can be identified only by the biological function of its encoded protein. A typical case for such use is isolation of cDNA clones for genes defective in particular mutant cells by the phenotypic complementation of the mutant cells with the expression of cDNA libraries. In some instances, wild-type cells could be used as a potential screening host for the isolation of genes involved in certain cell phenotypes, such as malignant transformation.

**An improved pcD vector and the development of a high-efficiency method for delivering plasmid DNA into cultured mammalian cells**

To achieve this goal, a highly efficient method for delivering the pcD expression libraries into cultured mammalian cells had to be developed. After attempting to use *E. coli* lambda phage as a vector, we found that a slight modification of the original calcium phosphate method highly enhances the efficiencies for delivering *E. coli* plasmid DNA into various mammalian cell lines under culture and allows for the establishment of stable transformants of these cells at frequencies up to 50%.

This delivery efficiency should be sufficient for screening expression libraries comprised of 1–2 million cDNA clones for an appropriate cDNA that induces a phenotypic change in the stably transformed library recipients. In addition, for the convenient selection of stable transformants in which transfected pcD-X DNA is integrated into their genome, we modified the pL1 linker DNA by inserting a SV40-early promoter-
driven neo marker gene for the construction of neomycin-selectable pcD2 libraries (Fig. 6 lower right figure). One successful application of this new transfection method and the pcD2-X vector was the isolation of the eukaryotic elongation 1 gene as a determinant of susceptibility to malignant transformation.40) Although the use of cultured mammalian cells as the prime host for expression cDNA cloning with the pcD vector series was initially intended, I gradually realized that the lack of proper mutant cells and the phenotypic instability of mammalian cells greatly hamper the use of this cloning strategy particularly when the targeted genes are expressed much less than their cDNA copies are represented as a minor species in the expression libraries.

**Trans-complementation cloning of cDNAs for human cell cycle genes**

One way to overcome this limitation would be use of unicellular eukaryotes such as yeasts if the targeted genes are expected to be evolutionarily well conserved because more mutants are available and phenotypic instability is likely to be less severe.

The first success in cloning a mammalian gene with yeast as a complementation host was made by Paul Nurse and his group, who isolated a human CDC2 cDNA by trans-complementation of a fission yeast temperature-sensitive mutant of the cdc2 gene with one of the pcD/pcD2 expression cDNA libraries by trans-complementation of the corresponding fission yeast mutants,41),46)–48) whereas the human CDK2 cDNA was cloned by trans-complementation of a budding yeast cdc28 mutant.52),53)

![Fig. 8. Mammalian cell cycle factors identified by the trans-complementation cloning of cDNAs.](image-url)

In mammalian cells, D-type cyclin-bound Cdk4/6 promotes the G0–G1 transition by activating the E2F-DP transcriptional factor complexes that govern the expression of some of the genes that are essential for the onset and progression of S phase. Cyclin E/A-bound Cdk2 is required for S phase onset, whereas cyclin B-bound Cdc2 triggers the onset of M phase. These cyclin-dependent kinases are negatively regulated by phosphorylation at tyrosine 15 or at a tyrosine residue equivalent to this site. When cells suffer DNA damage, these kinases are forced to be inactivated with this site remaining phosphorylated. Wee1, the mammalian ortholog of the fission yeast enzyme, phosphorylates Cdk2 and Cdc2 at this site. Cdc25A, a paralog of fission yeast Cdc25, dephosphorylates both Cdk4/6 and Cdk2, whereas Cdc25B and Cdc25C dephosphorylate Cdc2. The cDNAs for human CDC2, WEE1, CDC25A and CDC25B were cloned from one of the pcD/pcD2 expression cDNA libraries by trans-complementation of the corresponding fission yeast mutants,41),46)–48) whereas the human CDK2 cDNA was cloned by trans-complementation of a budding yeast cdc28 mutant.52),53)
mutants. A set of high-efficiency pcD library transducing vectors and a much improved method for delivering plasmid DNA into fission yeast were thus constructed (Fig. 7). Immediate outcomes of this effort were the successful cloning of full-length cDNAs for human WEE1 and CDC25A and CDC25B, orthologs and paralogs of fission yeast cell cycle control factors. In addition to the human CDC2 cDNA, trans-complementation cloning of these human counterparts provided a foundation for understanding how the mammalian cell cycle is controlled (Fig. 8). Trans-complementation of fission yeast mutants was also successfully used to isolate full-length cDNAs for mouse orotidine-5'-monophosphate decarboxylase and human liver fructose-1,6-bisphosphatase.

**Pushing it to the limit**

One potential advantage of using unicellular eukaryotes for an expression cloning host is the availability of temperature-sensitive mutants for proliferation and differentiation with strict non-permissive temperatures and high population homogeneities. One such mutant of the fission yeast *Schizosaccharomyces pombe* is the ran1/pat1 mutant, which undergoes lethal haploid meiosis upon inactivation of temperature-sensitive Ran1/Pat1 kinase (Fig. 9). Nutritional starvation is a key signal for mating and subsequent meiosis, a presumed prototype of cell differentiation in multicellular eukaryotes. Nutritional starvation activates the Ste11 transcriptional factor, which induces the expression of Mei2, the RNA-binding protein that is necessary and sufficient for the start and progression of meiosis. However, Mei2 is rendered inactive by phosphorylation by Ran1/Pat1 kinase until Mei3 protein, which is expressed upon mating, inactivates Ran1 kinase. The temperature-sensitive ran1/pat1 strains grown on poor nutrition medium are forced to commit lethal haploid meiosis when they are shifted to the non-permissive temperature, and any signals that down-regulate Ste11 would rescue the cells from lethal haploid meiosis.

---

**Fig. 9. Control of the early phase of fission yeast sexual development.** Nutrient (carbon and nitrogen sources) starvation triggers the onset of the sexual development of fission yeast by promoting the expression of the Ste11 transcriptional factor, which activates genes for mating pheromone and Mei2, the factor that is required and sufficient for the start and progression of meiosis. However, Mei2 is rendered inactive by phosphorylation by Ran1/Pat1 kinase until Mei3 protein, which is expressed upon mating, inactivates Ran1 kinase. The temperature-sensitive ran1/pat1 strains grown on poor nutrition medium are forced to commit lethal haploid meiosis when they are shifted to the non-permissive temperature, and any signals that down-regulate Ste11 would rescue the cells from lethal haploid meiosis.
suppressors of ran1 and other cell cycle mutants from fission yeast pcD2 cDNA libraries and genomic libraries, we succeeded in identifying several new genes that influence sexual development (conjugation). Epistatic analyses of these genes with known factors involved in the sexual development led us to draw a signal cascade controlling the sexual development (Fig. 10). Interestingly, all of these factors appear to be classified into four functional groups: differentiation signals mediating starvation for carbon and nitrogen, stress signals mediated by stress MAP kinases, an RNA binding protein that potentially controls a threshold for differentiation commitment, and cell cycle start and arrest factors that ensure mutual exclusiveness between proliferation and differentiation. There are mammalian orthologs of these fission yeast differentiation factors, some of which were isolated by trans-complementation of the ran1 mutant (see Fig. 10 figure legend). These results suggest the general principle that the differentiation process is composed of two steps: commitment of differentiation and expression of differentiated phenotypes, the former of which lacks strict cell-type specificity and is evolutionarily conserved to a certain extent, whereas the latter holds strict cell-type specificity (Fig. 11).
DIFFERENTIATION CONTROL

1. Differentiation signal
2. Stress signal
3. Threshold control
4. Mutual exclusiveness with cell cycle start

Commitment → Expression of differentiated phenotypes

Cell type-nonspecific → Cell type-specific

Fig. 11. Aspects of differentiation control. Differentiation is divided into two steps: commitment and execution (expression of differentiated phenotypes). The process of commitment is less specific to cell type and appears to be regulated by at least four distinct signals or cellular mechanisms: 1) the differentiation signal, 2) MAP kinase-mediated stress signals, 3) a type of threshold control and 4) the mechanism ensuring mutual exclusiveness between differentiation and cell cycle initiation.

Fig. 12. The power of functional expression cloning. In this hypothetical signal cascade, or chain of interactions, Factor B is a temperature-sensitive mutant. Genes or their cDNA encoding A, C, F, D that dimerizes with B, or in the inflow signal 2 pathway, E or F if nonfunctional F/F homodimer molecules interfere with the G/F heterodimer, might be isolated as weak multi-copy suppressors of the temperature sensitivity of the mutated B.
Future prospects

In 1990, the Human Genome Project was launched as an international collaborative work to determine the nucleotide sequences of the entire human genome; nearly a decade ago, the project was completed with the identification of approximately 25,000 genes and the primary structure of the encoded proteins aided by cDNA sequence projects. However, we are still far from understanding how each biological phenomenon is executed and by which proteins. This understanding is made more difficult by the multi-functionality of many mammalian proteins. Although cDNA expression cloning is no longer needed for isolating new genes, we may have to rely on this technique to find additional functions of already known proteins by identifying their roles in unexpected cellular phenomena. In this regard, as already described in the previous section, I would like to stress the power of the strategy to search for weak multi-copy suppressors of the mutant phenotype, which have traditionally been ignored because of complication of analysis, to identify genes encoding the proteins functionally interacting with the mutated protein in the signal cascade. If the expression of the mutant phenotype is conditional, such as being temperature-sensitive or nutrition-dependent, genes encoding factors upstream and downstream of the mutant protein might be identified as multi-copy suppressors of this mutant, as illustrated in Fig. 12. Alternatively, if the mutated protein is a subunit of a heterodimer protein, the gene for the association partner could be isolated as a strong multi-copy suppressor. Moreover, if different signals activate the cascade downstream of the mutated factor, genes for their transducing factors could also be isolated as multi-copy suppressors. Once factors constituting the signal cascades of interest are identified, their mammalian counterparts could easily be retrieved from the appropriate databases that already exist if the cascades are evolutionarily conserved. If the cascades are evolutionarily diverted, the host unicellular eukaryote could be partially humanized by replacing one or a few of downstream or upstream factors with their human homologs and used as a host for direct expressing cloning of mammalian cDNAs.

References

1) Mulligan, R.C., Howard, B.H. and Berg, P. (1979) Synthesis of rabbit beta-globin in cultured monkey kidney cells following infection with a SV40 beta-globin recombinant genome. Nature 277, 108–114.
2) Fromm, M. and Berg, P. (1983) Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. Mol. Cell. Biol. 3, 991–999.
3) Mulligan, R.C. and Berg, P. (1980) Expression of a bacterial gene in mammalian cells. Science 209, 1422–1427.
4) Southern, P.J. and Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. 1, 327–341.
5) Spiegelman, S., Watson, K.F. and Kacian, D.L. (1971) Synthesis of DNA complements of natural RNAs: a general approach. Proc. Natl. Acad. Sci. U.S.A. 68, 2843–2845.
6) Ross, J., Aviv, H., Scolnick, E. and Leder, P. (1972) In vitro synthesis of DNA complementary to purified Rabbit globin mRNA. Proc. Natl. Acad. Sci. U.S.A. 69, 264–268.
7) Diggelmann, H., Faust, C.H. Jr. and Mach, B. (1973) Enzymatic synthesis of DNA complementary to purified 14S messenger RNA of immunoglobulin light chain. Proc. Natl. Acad. Sci. U.S.A. 70, 693–696.
8) Chen, J.H., Lavers, G.C. and Spector, A. (1973) Synthesis of complementary DNA from lens mRNA with RNA-dependent DNA polymerase. Biochem. Biophys. Res. Commun. 52, 767–773.
9) Berns, A.J.M., Bloemendal, H., Kaufman, S. and Verma, I.M. (1973) Synthesis of DNA complementary to 14S calf lens crystallin messenger RNA by reverse transcriptase. Biochem. Biophys. Res. Commun. 52, 1013–1019.
10) Aviv, H., Packman, S., Swan, D., Ross, J. and Leder, P. (1973) In vitro synthesis of DNA complementary to mRNA derived from a light chain-producing myeloma tumour. Nat. New Biol. 241, 174–176.
11) Faust, C.H. Jr., Diggelmann, H. and Mach, B. (1973) Isolation of poly(adenyl acid)-rich ribonucleic acid from mouse myeloma and synthesis of complementary deoxyribonucleic acid. Biochemistry 12, 925–931.
12) Verma, I.M., Temple, G.F., Fan, H. and Baltimore, D. (1972) In vitro synthesis of DNA complementary to rabbit reticulocyte 10S RNA. Nat. New Biol. 235, 163–167.
13) Keedes, L.H., Chang, A.C., Houseman, D. and Cohen, S.N. (1975) Isolation of histone genes from unfractionated sea urchin DNA by subculture cloning in E. coli. Nature 255, 533–538.
14) Hershfield, V., Boyer, H.W., Yanofsky, C., Lovett, M.L. and Helinski, D.R. (1974) Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. U.S.A. 71, 3455–3459.
15) Crick, F. (1979) Split genes and RNA splicing. Science 204, 264–271.
16) Schmidtke, J. and Epplen, J.T. (1980) Sequence organization of animal nuclear DNA. Hum. Genet. 55, 1–18.
25) Efstratiadis, A., Vournakis, J.N., Donis-Keller, H., Brock, K.V., Deng, R. and Riblet, S.M. (1992) 3-Terminal cloning and properties of bacteriophage T4-induced RNA ligase. Nature 260, 221–225.

26) Last, J.A. and Anderson, W.F. (1976) Puriﬁcation and properties of bacteriophage T4-induced RNA ligase. Arch. Biochem. Biophys. 174, 167–176.

27) England, T.E. and Uhenbeck, O.C. (1978) 3-Terminal joining of RNA with a synthetic oligonucleotide as a hybridization probe. Proc. Natl. Acad. Sci. U.S.A. 75, 5066–5070.

28) Brock, K.V., Deng, R. and Riblet, S.M. (1992) Nucleotide sequencing of 5′ and 3′ termini of bovine viral diarrhea virus RNA by RNA ligation and PCR. J. Virol. Methods 38, 39–46.

29) Okayama, H. and Berg, P. (1982) High efﬁciency cloning of full length cDNA. Mol. Cell. Biol. 2, 161–170.

30) Molloy, P.L. and Symmons, R.H. (1980) Cleavage of DNA:RNA hybrids by Type II restriction enzymes. Nucleic Acids Res. 8, 2939–2946.

31) Okayama, H. and Berg, P. (1983) A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. 3, 280–289.

32) Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294–5299.

33) Okayama, H., Kawauchi, M., Brownstein, M., Lee, F., Yokota, T. and Ariai, K. (1987) High-efficiency cloning of full-length cDNA: construction and screening of cDNA expression libraries for mammalian cells. Methods Enzymol. 154, 3–28.

34) Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I. and Okayama, H. (1992) Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol. 65, 1105–1113.

35) Jolly, D.J., Okayama, H., Berg, P., Esty, A.C., Filipula, D., Kohlent, P., Johnson, G.G., Shively, J.E., Hunkapiller, T. and Friedmann, T. (1983) Isolation and characterization of a full-length cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. U.S.A. 80, 477–481.

36) Jolly, D.J., Esty, A.C., Bernard, H.U. and Friedmann, T. (1982) Isolation of a genomic clone partially encoding human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. U.S.A. 79, 5038–5041.

37) Gubler, U. and Hoffman, B.J. (1983) A simple and very efﬁcient method for generating cDNA libraries. Gene 25, 263–269.

38) Maruyama, K. and Sugano, S. (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene 138, 171–174.

39) Chen, C. and Okayama, H. (1987) High-eﬃciency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. 7, 2745–2752.

40) Tatsuka, M., Mitsui, H., Wada, M., Nagata, A., Nojima, H. and Okayama, H. (1992) Elongation factor-1 α gene determines susceptibility to transformation. Nature 359, 333–336.

41) Lee, M.G. and Nurse, P. (1987) Complementation and characterization of an AP-dominant mutation in fission yeast Schizosaccharomyces pombe. Cell 5299.

42) Toyama, R. and Okayama, H. (1999) Human chorionic gonadotropin α and human cytomegalovirus promoters are extremely active in the fission yeast Schizosaccharomyces pombe. FEBS Lett. 425, 217–221.

43) Käufer, N.F., Simanis, V. and Nurse, P. (1985) Fission yeast Schizosaccharomyces pombe correctly excises a mammalian RNA transcript intervening sequence. Nature 318, 78–80.

44) Jones, R.H., Moreno, S., Nurse, P. and Jones, N.C. (1988) Expression of the SV40 promoter in fission yeast: identiﬁcation and characterization of an AP-1-like factor. Cell 53, 659–667.

45) Okazaki, K., Okazaki, N., Kume, K., Jinno, S., Tanaka, K. and Okayama, H. (1990) High-frequency transformation method and library transducing vectors for cloning mammalian cDNAs by trans-complementation of Schizosaccharomyces pombe. Nucleic Acids Res. 18, 6485–6489.

46) Igarashi, M., Nagata, A., Jinno, S., Suto, K. and Okayama, H. (1991) Weel−like gene in human No. 3] Development of cDNA expression cloning 117
47) Nagata, A., Igarashi, M., Jinno, S., Suto, K. and Okayama, H. (1991) An additional homolog of the fission yeast cdc25 gene occurs in humans and is highly expressed in some cancer cells. New Biol. 3, 959–968.

48) Jinno, S., Suto, K., Nagata, A., Igarashi, M., Kanaoka, Y., Nojima, H. and Okayama, H. (1994) Cdc25A is a novel phosphatase functioning early in the cell cycle. EMBO J. 13, 1549–1556.

49) Okayama, H., Nagata, A., Jinno, S., Murakami, H., Tanaka, K. and Nakashima, N. (1996) Cell cycle control in fission yeast and mammals: identification of new regulatory mechanisms. Adv. Cancer Res. 69, 17–62.

50) Bertolotti, R., Armbruster-Hilbert, L. and Okayama, H. (1995) Liver fructose-1,6-bisphosphatase cDNA: trans-complementation of fission yeast and characterization of two human transcripts. Differentiation 59, 51–60.

51) Yamamoto, M. (1996) Regulation of meiosis in fission yeast. Cell Struct. Funct. 21, 431–436.

52) Elledge, S.J. and Spottswood, M.R. (1991) A new human p34 protein kinase, CDK2, identified by complementation of a cdc28 mutation in Saccharomyces cerevisiae, is a homolog of Xenopus Eg1. EMBO J. 10, 2653–2659.

53) Ninomiya-Tsuji, J., Nomoto, S., Yasuda, H., Reed, S.J. and Matsumoto, K. (1994) Cloning of a human cDNA encoding a CDC2-related kinase by complementation of a budding yeast cdc28 mutation. Proc. Natl. Acad. Sci. U.S.A. 88, 9006–9010.

54) Takeda, T., Toda, T., Kominami, K., Kohnosu, A., Yanagida, M. and Jones, N. (1995) Schizosaccharomyces pombe aff1 encodes a transcription factor required for sexual development and entry into stationary phase. EMBO J. 14, 6193–6208.

55) Okazaki, N., Okazaki, K., Watanabe, Y., Kato-Hayashi, M., Yamamoto, M. and Okayama, H. (1998) Novel factor highly conserved among eukaryotes controls development in fission yeast. Mol. Cell. Biol. 18, 887–895.

56) Hiroi, N., Ito, T., Yamamoto, H., Ochiya, T., Jinno, S. and Okayama, H. (2002) Mammalian Red1 is a novel transcriptional cofactor that mediates retinoic acid-induced cell differentiation. EMBO J. 19, 5235–5244.

57) Kawasaki, H., Song, J., Eckner, R., Ugai, H., Chiu, R., Taira, K., Shi, Y., Jones, N. and Yokoyama, K.K. (1998) P300 and ATF-2 are components of the DRF complex, which regulates retinoic acid and E1A-mediated transcription of the c-jun gene in F9 cells. Genes Dev. 12, 233–245.

58) Kato, T. Jr., Okazaki, K., Murakami, H., Stettler, S., Fantes, P.A. and Okayama, H. (1996) Stress signal, mediated by a Hog1-like MAP kinase, controls sexual development in fission yeast. FEBS Lett. 378, 207–212.

59) Tsukahara, K., Yamamoto, H. and Okayama, H. (1998) An RNA binding protein negatively controlling differentiation in fission yeast. Mol. Cell. Biol. 18, 4488–4498.

60) Yamamoto, H., Tsukahara, K., Kanaoka, Y., Jinno, S. and Okayama, H. (1999) Isolation of a mammalian homologue of a fission yeast differentiation regulator. Mol. Cell. Biol. 19, 3829–3841.

61) Obara-Ishihara, T. and Okayama, H. (1994) A B-type cyclin negatively regulates conjugation via interacting with cell cycle `start' genes in fission yeast. EMBO J. 13, 1863–1872.

62) Yamaguchi, S., Murakami, H. and Okayama, H. (1997) A WD repeat protein controls the cell cycle and differentiation by negatively regulating Cdc2/B-type cyclin complexes. Mol. Biol. Cell 8, 2475–2486.

63) Tanaka, K. and Okayama, H. (2000) A pcl-like cyclin activates the Res2p-Cdc10p cell cycle “start” transcriptional factor complex in fission yeast. Mol. Biol. Cell 11, 2845–2862.

64) Ogasawara, T., Kawaguchi, H., Jinno, S., Hoshi, K., Itaka, K., Takato, T., Nakamura, K. and Okayama, H. (2004) Bone morphogenetic protein 2-induced osteoblast differentiation requires Smad-mediated down-regulation of Cdk6. Mol. Cell. Biol. 24, 6560–6568.
Profile

Hiroto Okayama was born in 1948 and graduated from Kumamoto University School of Medicine in 1973 and the Ph.D. course of Kyoto University School of Medicine in 1977. He started his research carrier with studies on poly ADP-ribosylation of nuclear proteins, which led to the discovery of a new enzyme that cleaves the ADP-ribose–histone linkage. In 1978, he went to Stanford University School of Medicine as postdoctoral fellow and developed a new technique that enables the cloning of full-length complementary DNAs on the basis of their functional expression in a given cell of interest. After moving to National Institutes of Health as Visiting Scientist in 1983, he developed a high-efficiency transfection method for delivering cDNA libraries into mammalian cells and cDNA cloning vectors that enable selection on stable cDNA expression. In 1988, he came back to Japan as Molecular Genetics Professor of Institute of Microbial Diseases at Osaka University. He continued development of various methods that maximize the utilization of the cDNA cloning technique, including those that expand its expression hosts from mammalian cells to fission yeast. In 1992, he moved to the University of Tokyo as Molecular Biology Professor of Faculty of Medicine. Meanwhile from 1991 to 1996, he served as Director of the Okayama Cell Switching Project, ERATO, JRDC. During these periods, he isolated various critical cell cycle- and cell differentiation-controlling genes.