I. INTRODUCTION

1.1. Problems and research rationale

Penicillium marneffei is a thermally dimorphic fungus, which can exist in either a unicellular yeast form or a multicellular mold form. This morphological change is determined by the temperature. *Penicillium marneffei* is the only known dimorphic fungus that has been found among the genus *Penicillium*. At 25-30 °C, the fungus grows as a mold with the typical morphology of the genus. The mold phase can be converted to a parasitic, yeast-like phase on the appropriate media *in vitro* at 37 °C or in infected animals and humans. *Penicillium marneffei* causes disease both in healthy individuals and immunocompromised patients, especially within the endemic area. This area includes Southeast Asian countries and southern China. The fungus may disseminate in immunocompromised patients and, if untreated, is fatal. Infection by *P. marneffei* has become an important problem among HIV-infected individuals and it has been considered to be an AIDS indicator disease since 1992 (Supparatpinyo *et al.*, 1992).

Thailand is one of the countries in the endemic area that has been plagued by infection from *P. marneffei*. Several cases of penicilliosis due to *P. marneffei* had been reported before the epidemic of AIDS. This infection occurred primarily in immunocompetent patients with predisposing illness (Jayanetra *et al.*, 1984). After the onset of AIDS in Thailand in 1984, *P. marneffei* infection has increased in frequency. The organism takes advantage of the HIV-induced compromised immune status of the host and has become one of the common causes of opportunistic infection in AIDS patients. The first case of *P. marneffei* infection in a patient infected with HIV was reported in 1989 from Bangkok, Thailand (Satapatayavongs *et al.*, 1989). Thereafter, many cases of this disease have been reported with the highest rate of infection in the north (Ruxrungtham & Phanuphak, 2001). Chiang Mai Province accounts for one-third of the country's reported cases of *P. marneffei*

infection. According to a report from Maharaj Nakorn Chiang Mai Hospital, the total number of *P. marneffei* infected patients reached 1,690 by the end of 2003 (data from central laboratory unit, Maharaj Nakorn Chiang Mai Hospital, personal communication), and a total of 6,112 patients for the entire country was reported by the Division of Epidemiology, Ministry of Public Health, Thailand on March 15th, 2004 (http://203.157.19.193/aids/Aidstab4.html). In all probability, the number of *P. marneffei* infected patients is higher than reported since the empirical treatment of patients may either pre-empt positive cultivation or in some instances prohibits re-isolation of clinical samples which may have failed to thrive. Figure 1 shows the correlation between the number of patients with culture identified *P. marneffei* from Maharaj Nakorn Chiang Mai Hospital (data from the Central Laboratory Unit, Maharaj Nakorn Chiang Mai Hospital, personal communication) and the number of HIV-infected patients in Chiang Mai. This figure clearly illustrates the conclusion that penicilliosis due to *P. marneffei* is an AIDS-defining disease. Although the overall number of HIV-infected patients has declined, the proportion of those infected with *P. marneffei* has slightly increased, especially in recent years. This may be due to the increasing number of patients reaching an advanced stage of HIV infection. Furthermore, such observation reflects the continuing threat of *P. marneffei* as a pathogen within the AIDS community.

Host responses to pathogenic fungi, either humoral or cellular mediated type, are induced by immunogenic proteins or antigens of the organisms. At present, knowledge of the composition and characteristics of *P. marneffei* antigenic proteins are still in its infancy. Mostly, the reports have focused on attempts to develop a diagnostic test for *P. marneffei* infection. Historically, initial attempts in antibody detection have relied on the use of crude, unfractionated mixtures of antigens, which inevitably resulted in problems with low specificity and sensitivity (Sekhon, Li & Garg, 1982; Yuen *et al.*, 1994; Kaufman *et al.*, 1995; Imwidthaya *et al.*, 1997). Strenuous efforts have been made to identify specific antigens by using western immunoblotting assay (Chongtrakool *et al.*, 1997; Vanittanakom *et al.*, 1997) and subsequently to purify those antigens from crude culture filtrate (Jeavons *et al.*, 1998; Poolsri, 1999). However, their attempts were only partially successful. Ultimately, one cell wall antigen could be identified by the application of recombinant

technology (Cao *et al.*, 1998b). This purified recombinant mannoprotein, Mp1p, was further used in the ELISA assays for both antigen and antibody detection in sera of HIV-infected patients with disseminated *P. marneffei* infection (Cao *et al.*, 1998a and 1999). However, those tests are not commercially available, and the antibody to Mp1p seems to be unreliable in the detection of the protein antigen of *P. marneffei* in Thai isolate (preliminary data).

Since Thailand is facing significant problems with regard to *P. marneffei* infections, it is worthwhile looking for diagnostic markers to aid in the development of a reliable and rapid diagnostic test. The first reason that a rapid serological method should be developed for *P. marneffei* infection is recognition of the pathogen since there are no specific signs for penicilliosis due to *P. marneffei*. Penicilliosis due to *P. marneffei* infection results in a wide spectrum of clinical manifestations that often mimic other common opportunistic infections in Thailand, such as cryptococcosis, histoplasmosis, and tuberculosis, i.e. weight loss, fever, respiratory symptoms, hepatosplenomegaly (Deng & Conner, 1985; Yuen *et al.*, 1986). Such presumptive diagnoses are frequently based on clinical experience, knowledge of the disease prevalence, and detection of skin lesions. These lesions are quite specific, but not totally diagnostic since they mimic molluscum contagiosum lesions (Chiewchanwit *et al.*, 1991; Supparatpinyo *et al.*, 1994a). Skin lesions always appear when the infection is in an advanced disseminated stage, which is life threatening (Chiewchanwit *et al.*, 1991). When suspected, a presumptive diagnosis can be made on the basis of characteristic histopathology, but this can sometimes be confused with the histologic appearance of *Histoplasma capsulatum*. Definite diagnosis requires the isolation of the fungus from pathological specimens. However, this process is time consuming. Other identification methods that have been developed such as immunodiffusion assay (Sekhon, Li & Garg, 1982), immunohistochemistry (Kaufman *et al.*, 1995), and western immunoblot (Vanittanakom *et al.*, 1997; Jeavons *et al.*, 1998) are not practical for routine laboratory use. They require specialized equipment and well-trained laboratory staff that are not always available in small hospitals, especially in rural areas. In general, both presumptive and definitive identification methods are usually highly sensitive when the fungus is already

disseminated. Yet, there is no rapid method for diagnosis of the early stage of infection. To date, information on protein antigens of *P. marneffei* is not sufficient for development of such a test.

The second reason for the development of serological tests is the likelihood that patients may harbor latent, asymptomatic *P. marneffei* infections. This is supported by results obtained by Vanittanakom *et al.* (1997), where serum from one of the AIDS patients used as a negative control repeatedly gave positive reaction against the 54- and 50-kDa specific antigens of *P. marneffei*. This patient finally developed clinical symptoms of *P. marneffei* infection with positive culture 2 months after initial testing, indicating that antibody responses can be observed before the symptomatic appearance. The application of specific antigens to detect at-risk populations without obvious symptoms of penicilliosis due to *P. marneffei* merits further investigation. From the information described above, an effective serodiagnostic test should be developed since the test will be useful not only for ascertaining the real incidence of *P. marneffei* infection, but also for immunosurveillance in AIDS patients.

The basic knowledge gained from characterization of antigenic proteins will lead to further applications in future research. Such applications include epidemiological studies, pathogenesis of the fungus, host defense mechanism, and therapeutic interventions. Once the genes that encode proteins have been cloned, it is simple to induce the expression and purification for other purposes. For instance, the purified recombinant proteins, which are selected to serve as diagnostic markers, could also be used directly to detect specific antibodies. On the other hand, antigen detection assay could be developed by immunizing animals with purified antigen, and then preparing immunoglobulins from hyperimmune sera or hybridoma clones. Serological tests for epidemiological studies could also be developed using appropriate antigens. Some antigens may be involved in pathogenesis and could serve as targets for drug or vaccine development.

1.2. Construction of a lambda-based complementary DNA library

Each organism and tissue type has a unique population of messenger RNA (mRNA) molecules. This mRNA population is difficult to maintain, clone, and amplify. Therefore, it must be converted to a collection of more stable complementary DNA molecules (cDNA). A cDNA library is an array of DNA copies of an mRNA population that is propagated in a cloning vector. This library is composed of sequences representing a population of genes that are expressed in tissue or cell type at the time of RNA collection.

Successful cDNA synthesis should yield full-length copies of the original population of mRNA molecules. Hence, the quality of the cDNA library can be only as good as the quality of mRNA. Pure, undegraded mRNA is essential for the construction of large, representative cDNA library (Chromczynski & Sacchi, 1987). After synthesis, the cDNA is inserted into an *E. coli*-based vector (plasmid or λ), and the library is screened for the clones of interest. Since 1980, lambda has been the vector system of choice for cDNA cloning (Han & Rutter, 1987; Meissner, Sisk & Berman, 1987; Murphy & Efstratiadis, 1987; Palazzolo & Meyerowitz, 1987; Scherer *et al.*, 1981; Young & Davis, 1983). The fundamental reasons are that in vitro packaging of λ generally has a higher efficiency than plasmid transformation, and λ libraries are easier to handle (amplify, plate, screen, and store) than plasmid libraries. However, most λ vectors have the disadvantage of being poorer templates for DNA sequencing, site-specific mutagenesis, and restriction shuffling. The development of excisable λ vectors, such as those based on restriction enzyme digestion (Swaroop & Weissman, 1988), site-specific recombination (Palazzo *et al.*, 1990), or filamentous phage replication (Short *et al.*, 1988), has increased the flexibility of DNA cloning. Now it is possible to clone and screen libraries with the efficiency and ease of λ systems, and be able to analyze positive clones with the ease and versatility of a plasmid. The vector that is compatible with the cDNA synthesis described in this study is based on the lambda ZipLox excision system (Gibco BRL). This vector uses a Cre-*loxP* site-specific recombination mechanism. The major steps in constructing a directional cDNA library from an mRNA population using the SuperScript™

Lambda System (Gibco BRL) are summarized in Figure 2. The map of the λZipLox vector is depicted in Figure 3.

A hybrid oligo(dT) primer adapter containing a *Not*I site is used to make directional cDNAs (Figure 4). This 44-base oligonucleotide was designed with a protective sequence to prevent the *Not*I restriction enzyme recognition site from being damaged in subsequent steps. The oligo also contains a 15-base poly(dT) sequence that binds to the 3' polyA region of the mRNA template. In first strand synthesis, the mRNA is primed with the *Not*I primer adapter, and it is transcribed by reverse transcriptase in the presence of dNTPs, radioactive nucleotide (^{32}P-dCTP) and buffer. SuperScriptII, an RNase H-deficient reverse transcriptase, enhanced yields of longer cDNA transcripts. The cDNA/mRNA hybrid is treated with RNase H in the second-strand synthesis reaction. The mRNA is nicked to produce fragments that serve as primers for DNA polymerase I, synthesizing the second strand cDNA. The uneven termini of the double-stranded cDNA must be polished with T4 DNA polymerase to allow efficient ligation of a *Sal*I adapter (Hu G, 1993). The *Sal*I adapter consists of complementary oligonucleotides which, when annealed, create a phosphorylated blunt end and a dephosphorylated cohesive end (Figure 5). This double-stranded adapter will ligate to other blunt termini on the cDNA fragments and to other adapters. Subsequently *Not*I digestion released the *Sal*I adapter and protective sequence on the primer adapter from the 3'-end of the cDNA. Size-exclusion chromatography is then performed to exclude the small cDNAs, fragments released from *Not*I digestion and excess *Sal*I adapters. The size-fractionated cDNA is then precipitated and ligated to the λZipLox vector, *Not*I/*Sal*I arms.

Every good cDNA library has key characteristics that distinguish it from its mediocre counterparts (Fulle, 2003). The library should be large enough to contain representatives of all sequences of interest, some of which may be derived from low abundant mRNAs. A cDNA library has to contain 170,000 clones to ensure finding a very low abundant mRNA molecule of 1 in 37,000 with 99% probability (Sambrook & Russell, 2001). In most cases, a library size of at least 10^5 primary clones is sufficient to obtain adequate representation of clones for typical eukaryotic cells which contain 10,000-30,000 different mRNA sequences. Furthermore, a good cDNA library should include a minimal number of clones that contain small (often

arbitrarily defined as less than 500 bp) cDNA inserts. Standard cDNA libraries contain an average insert size of at least 1.0 kb. Moreover, at least 90% of clones within the library have to contain cDNA inserts in order to consider it a good cDNA library.

Additionally, detection of some representative genes that are expressed at a specific time and condition of cell type can be used as a quality control for the cDNA library construction (Hagen, Gray & Kuijper, 1988). In this study, actin and heat shock protein 70 (hsp70)-expressing genes would be detected. Actin is a protein that is ubiquitously expressed in all eukaryotic cells. It plays a role in cell shape maintenance, cell division, and cell motility (Kabsch & Vandekerckhove, 1992), making it a good indicator of cell growth. Hsp70 is a protein that is expressed when cells are exposed to heat or other stress conditions (Zűgel & Kaufmann, 1999). It could be used as the representative marker for genes that are expressed during the temperature-shift response of the fungal cell.

1.3. Immunological screening of lambda phage cDNA expression library

Expression of the cloned gene from the λZipLox-based cDNA library could be driven by a controllable *lac* promoter that located in the vector. Unlike random libraries, in which only 50% of the clones contain cDNA inserts oriented properly for expression, one-third of members of a directional library are potentially able to express the antigen. Thus, specific cDNAs can be detected by screening only half as many clones. This two-fold reduction in time and resources needed for detection is particularly beneficial in the case of rare cDNAs, which may require screening as many as 10^6 clones in the case of a random library (Ian & Caroline, 1997).

Two general methods can be used to screen an expression library, DNA hybridization or immunological screening method. Deciding which method to use depends upon the probe available. In this study, the immunological screening method was chosen to screen the cDNA clones encoding antigenic proteins with a polyclonal antiserum. The λZipLox vector can produce the protein of a cloned insert fused to β-galactosidase within an appropriate *E. coli*. Theoretically, antibody to the protein of interest can recognize one-third of specific cDNA clones, in the case of directional cloning.

The essence of the technique is to infect an *E. coli* lawn as much as possible in primary screening. Initially, expression of the cloned inserts is suppressed in case the protein products are toxic. Once phage lysis is detected, the plates are overlaid with a nitrocellulose filter and switched to conditions that will allow fusion protein production onto the filter. Subsequently, the filter can be probed with the antibody as for a western blot to detect phage plaques producing the antigenic protein. Repeated rounds of screening will allow the purification of specific cDNA clones.

1.4. Analysis of the cloned sequences

Analysis of the data generated from cDNA sequences can provide insights into gene structure and function as well as help to direct experimental approaches to obtaining the sequence of a full-length clone even when only preliminary data are available. There are currently several software programs for sequence analysis. These resources may be databases and programs accessible to the users at individual sites or may be network-based program servers. This study used the latter one. There are numerous types of information that can be extracted from a DNA sequence. This study will extract valuable data using various types of analyzes. The types of analyzes performed in this study are as follows;

1.4.1) Homology to other proteins/genes

Probably the most important question to ask when obtaining a new sequence is "Is this sequence similar to another already deposited in the database?" One would like to know if this is a homologue of, or identical to, a previously identified sequence. BLAST (Basic Local Alignment Search Tools) is very fast searches types of analysis uses to find local segments of homology. No gaps in the aligned segments are allowed (Altschul *et al.*, 1990). This program can be used to query nucleotide sequences against nucleotide sequence databases (e.g., BLASTN), or the queried sequences can be translated into amino acid sequences to compared with protein sequence databases (e.g., BLASTX). Likewise, protein sequences can be used to search either the protein or nucleotide databases (e.g., BLASTP).

1.4.2) Multiple alignment of homologous sequences

Once a sequence has been identified as encoding a member of a potential gene family, important information concerning gene structure and function can be

extracted from the multiple alignment of related sequences. The alignment of two residues implies that those residues are performing similar roles in the two different proteins. Multiple sequence alignment can be performed by several programs through bioinformatics resource available online. One good example is the Clustal W program on The European Bioinformatics Institute (EBI) website (http://www.ebi.ac.uk/clustalw). Alternatively, using a user interface-based program such as Clustal X can do multiple sequence alignment as well.

1.4.3) Finding of an open reading frame (ORF)

The correct translation frame can be identified using the Translate program. The correct frame should appear as an open reading frame without an inappropriate early stop. However, sometimes more than one possible reading frames can be deduced from the six-frame translation, especially when only the fragment of a gene is obtained. In such cases, searching all reading frames of a cDNA sequence against protein sequences in databases by using program such as BLASTX can also identify the possible correct reading frame. The frame detected as having significant function could be reckoned as the correct reading frame.

1.4.4) Identification of functional domains

Conserved protein sequences or motifs in a protein sequence can be scanned via specialized databases such as PROSITE, protein families (Pfam), Protein Domain families (PRODOM), SBASE protein domain families, *etc.*, by using Internet resources. This study used Motif Scan (Falquet *et al.*, 2002). This program searches and reports motif matches in a query protein sequence against PROSITE and the protein families (Pfam) databases.

1.4.5) Restriction site mapping

The locations of restriction sites within a sequence are valuable landmarks that can be used to compare locations within and between DNA sequences. In addition, identification of restriction enzyme sites in a sequence would provide tools that can be used to further manipulate a cloned fragment. In this study, restriction map analysis was performed using programs available on various web sites (Table 2).

1.5. An pRSET system for cloning and purification of recombinant protein in *Escherichia coli*

Numerous expression systems are currently available (Hodgson, 1993). When choosing a system, consideration should be given to an intended application of the expressed protein. An *E. coli* system is the most widely used host to make a recombinant protein for an immunological assay. *Escherichia coli* is easy to transform, grows quickly in simple media, and requires inexpensive equipment for growth and storage. In this study, a pRSET system (Gibco BRL) for expression of a histidine (His) tag protein was chosen.

The vectors are composed of pRSET A, B, and C, which are differentiated by their coding frames (Gibco BRL). The map of the pRSET vectors is shown in Figure 6. The N-terminus of multiple cloning sites includes an ATG translation initiation codon, a polyhistidine tag that functions as a metal binding domain in the translated protein, a transcript stabilizing sequence from the gene 10 of phage T7, the Xpress™ epitope, and the enterokinase cleavage recognition sequence. The T7 promoter controls expression of the gene of interest from the pRSET. T7 RNA polymerase specifically recognizes this promoter. For expression of the cloned gene, it is necessary to deliver T7 RNA polymerase to the cells.

The BL21(DE3)pLysS strain (Appendix D) is used for expression of T7 RNA polymerase. This strain carries the DE3 bacteriophage lambda lysogen. The lambda lysogen contains the *lacI* gene, the T7 RNA polymerase gene under control of the *lac*UV5 promoter, and a small portion of the *lacZ* gene. This *lac* construct, which is inserted into the *int* gene, inactivates the *int* gene. Disruption of the *int* gene prevents excision of the phage (i.e. lysis) in the absence of helper phage. The *lac* repressor represses expression of T7 RNA polymerase. Addition of isopropyl-β-thiogalactopyranoside (IPTG) allows expression of T7 RNA polymerase. In addition, T7 lysozyme that is produced from pLysS has been shown to bind to T7 polymerase and inhibit transcription. This activity is exploited to reduce basal levels of T7 RNA polymerase and prevent the basal level of expression of the fusion protein. The T7 lysozyme is a bifunctional enzyme. In addition to its T7 RNA polymerase binding activity, it also cleaves a specific bond in the peptidoglycan layer of the *E. coli* cell wall. This activity increases cell lysis by repeated freeze-thaw prior to purification.

The metal binding domain of the fusion peptide allows purification of recombinant proteins by charged nickel (Ni^{2+}) affinity chromatography (Probond™ resin, Gibco BRL). The affinity of the poly-His to Ni^{2+} is very high, allowing contaminating proteins to be washed away. Elution is accomplished by adding buffer with pH 4 that will cause dissociation of His_6 tagged protein from the nickel ions. Alternatively, imidazole gradient elution under native condition method can be used to displace the polyhistidine-containing fusion proteins from the matrix if the expressed protein is unstable or denatures at low pH.

1.6. The catalase-peroxidase

Catalase-peroxidase is classified as a class I peroxidase (peroxidase of prokaryotic origin). It is a unique bifunctional enzyme, capable of either reducing H_2O_2 with an external reductant (peroxidase activity) or disproportionating it to water and oxygen molecules (catalase activity) (Long & Salin, 2000). The catalase reaction is the overwhelming activity of the bifunctional enzyme similar to monofunctional catalases, but it also contains an identical active site to that of peroxidases (Mutsuda *et al.*, 1996). This growing class of enzyme has been isolated from several archeobacteria, pathogenic eubacteria, fungi, and plants. There are no reports of this enzyme in mammalian cells. The known and studied catalase-peroxidases usually contain 726-753 amino acids per subunit and consist of two highly homologous halves, each of which shares significant amino acid identity with yeast cytochrome C peroxidase (CCP), but not with typical catalases (Welinder, 1991).

The catalase-peroxidases have been implicated as a virulence factor among the intracellular pathogenic species of *Mycobacteria,* including *M. tuberculosis*, *M. bovis*, *M. leprae*, and *M. avium* (Jackett, Aber & Lowrie, 1978; Manca *et al.*, 1999; Wilson, Lisle & Collins, 1995). The enzymes could protect them from the deleterious effects of macrophage-generated hydrogen peroxide. The virulence of *M. bovis* in guinea pigs is dependent on the *KatG*-encoded catalase-peroxidase. The *KatG*-lacking strains were significantly less virulent than the parental strains, and the integration of a functional *KatG* gene into most mutants restored full virulence (Wilson Lisle & Collins, 1995). In addition, *Kat*-encoded proteins have been implicated as a virulence factor in a wide range of bacterial pathogens such as *E. coli*

O157 enterohemorrhagic strain (Brunder, Schmidt & Karch, 1996), *Legionella pneumophila* (Bandyopadhyay & Steinman, 1998), *Yersinia pestis* (Garcia *et al.*, 1999), *Burkholderia cepacia* (Lefebre & Valvano, 2001), *Burkholderia pseudomallei* (Loprasert *et al.*, 2003), and *Pseudomonas aeruginosa* (Brown *et al.*, 1995).

Among the fungi, catalase-peroxidase encoding genes have been reported in *Penicillium simplicissimum* (Fraaije *et al.*, 1996), *Neurospora crassa* (Peraza & Hansberg, 2002), *Aspergillus nidulans* (Scherer *et al.*, 2002), and *Aspergillus fumigatus* (Paris *et al.*, 2003). However, only the *A. fumigatus* Cat2p has been shown to be a putative virulence factor by involving in the H_2O_2 degradation *in vitro* and it could transiently protect the fungus against the oxidative burst in the rat model (Paris *et al.*, 2003).

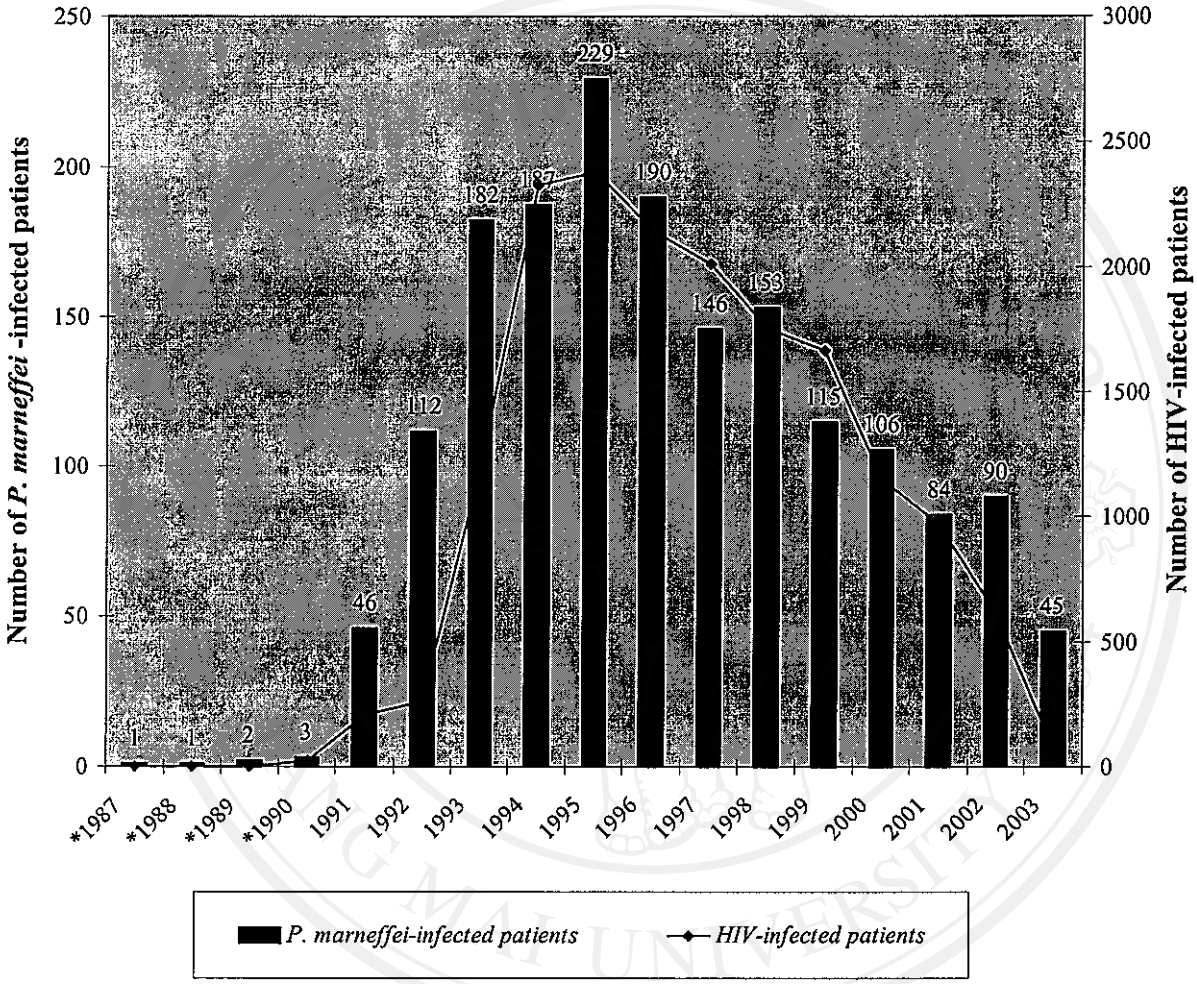


Figure 1. Positive correlation between the incidence rate of *P. marneffei* infection and the number of AIDS patients in Chiang Mai. A significant number of *P. marneffei*-infected patients correlates with the number of AIDS patients in Maharaj Nakorn Chiang Mai Hospital during 1987-2003. (Note: the asterisk * indicates the *Penicillium marneffei* infected patients from the year 1987-1989, and 2 from the year 1990 were non-AIDS)

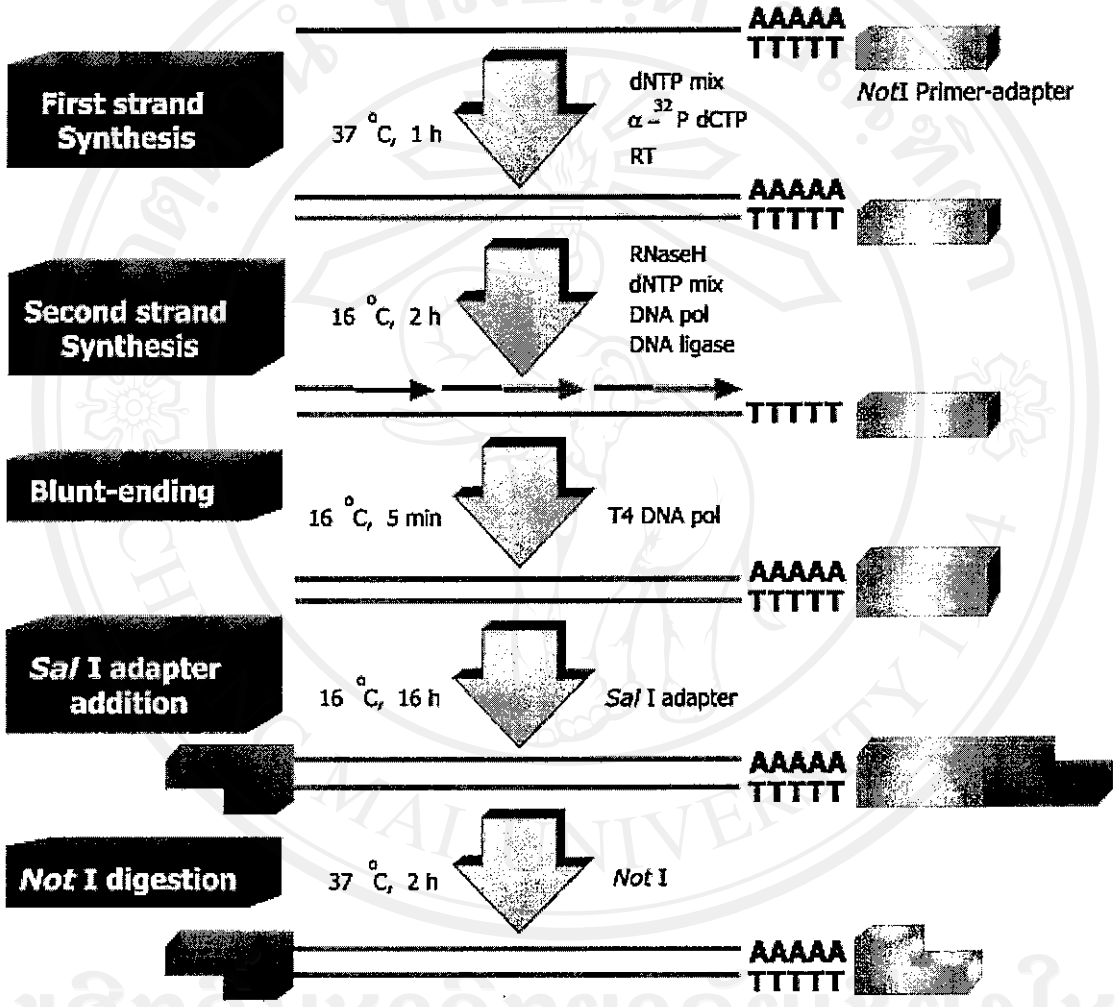
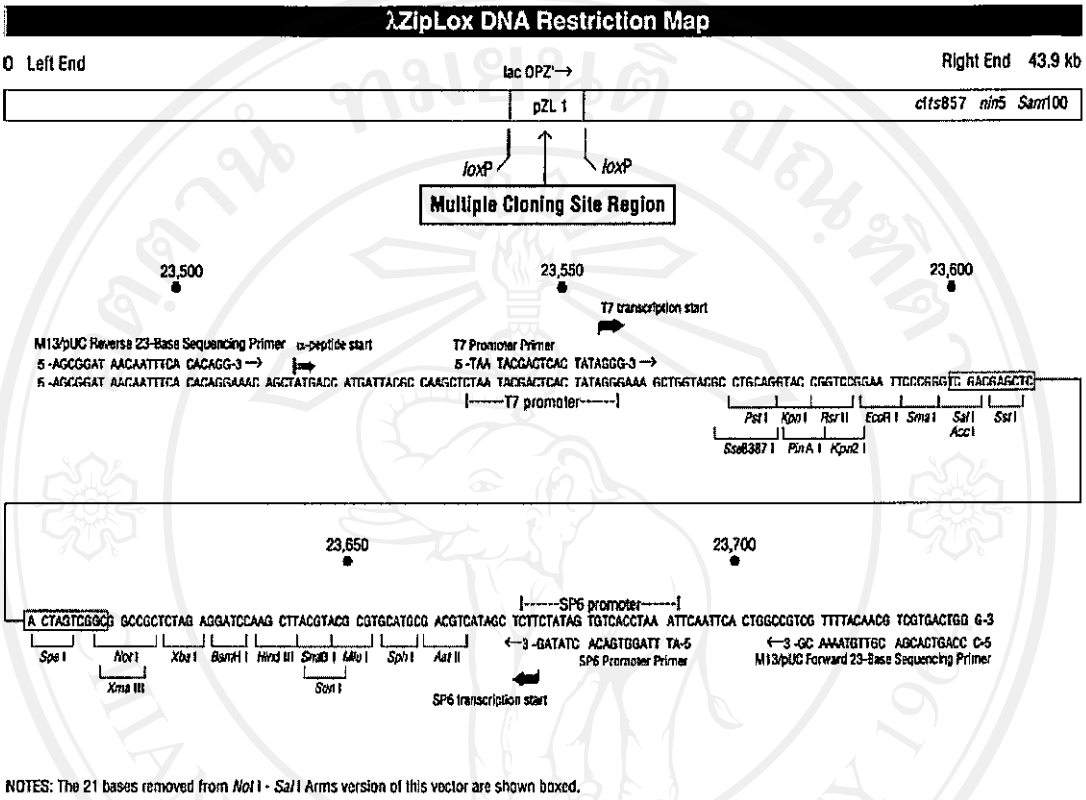


Figure 2. Construction of directional cDNAs by using the SuperScript™ Lambda System (adapted from the protocol manual from Gibco BRL).



NOTES: The 21 bases removed from *Not I - Sal I* Arms version of this vector are shown boxed.

Figure 3. Map of λZipLox vector (picture from the protocol manual of Gibco BRL)..

5'– pGACTAGT TCTAGA TCGCGA GCGGCCGC$(T)_{15}$ – 3'

Spe I *Xba* I *Nru* I *Not* I

Figure 4. Fourty-four bases of *Not*I primer adapter.

5'– TCGACCCACGCGTCCG–3'

3'– GGGTGCGCAGGCp–5'

Sal I *Mlu* I

Figure 5. Sequence of the *Sal*I adapter.

Figure 6. Map of pRSET vectors (picture from a manual protocol of Gibco BRL)