Screening of a Genomic Library in Search of the Biosynthetic Gene Cluster of Epothilones in the myxobacterium Sorangium cellulosum

Etienne NGUIMBI [1-2] , Yue Zhong LI[1] , Gabriel AHOMBO[2] , Zhifeng LI[1] , Raoul AMPA[2] .

ABSTRACT

The methods used to search for epothilone gene cluster modules in Sorangium cellulosum were surveyed. In this work, lambda genomic libraries of two closely related strains were constructed. The packaging efficiencies were 261860 pfu/ug and 987179.4pfu/ug respectively for So.9733-1 and So. 9881. Using the software Primer Premier, two probes corresponding to epophilone A and epothilone B were designed and used to detect signals for ksq and pcp during the screening by hybridization of the constructed lambda genomic library. Among 24000 plaques tested with the KS probe of epothilone A, we obtained 78 positive clones, then these positive clones were tested with the PCP probe of epoB. Only 4 plaques gave simultaneous signals with the two probes for epothilone A and epothilone B. DNA of the four recombinant phages was extracted, followed by SalI digestion. The restriction digested map of these clones was compared with the already reported information about the epothilone gene cluster. These results were in concordance with the reported partial upstream and downstream sequences of the epothilone gene cluster from strains SMP44 and So90. The results led to the isolation and identification of a fragment which may be sequenced and analyzed.

Keywords: Sorangium, lambda, epothilone, hybridization, probes

Criblage d'une Banque Génomique pour la Recherche du Gène Cluster responsable de la Biosynthèse des Epothilones chez la Myxobactérie Sorangium Cellulosum.

RÉSUMÉ

Un méthode de recherche des modules du gène Cluster responsable de la biosynthèse des epothilones chez la Myxobactérie sorangium Cellulosum a été éssayée. Dans ce travail une librairie génomique utilisant les vecteurs lambda a été construite chez deux souches voisines. L'éfficience de l'emballage a été 261860pfu/µg et 987179,4pfu/µg respectivement pour So.9733-1 et So 9881. En utilisant le logiciel Primer Premier, deux sondes nucléotidiques correspondant à epothilone a et epothlone B ont été désignées. Ces sondes ont été utilisées pendant l'hybridation pour detecter les signaux de KSQ et PCP..Sur 24000 plaques testées avec la sonde KS de epothilone A, 78 clones ont donné des signaux positifs. Ces 78 clones testées avec la sonde PCP de epothilone B, ont permis d'obtenir 4 clones avec des doubles signaux pour les deux sondes. L'ADN des quatre phages recombinés a été purifié et digéré par SalI. La carte de restriction des ces quatre clones a été comparé avec des informations déja connues sur le gène cluster responsable de la synthèse de l'epothilone. Ces résultats sont concordants avec la structure du gène cluster en amont comme en aval tel chez SMP44 et So.90. Ces résultats ont conduit à l'identification et l'isolation du clone à séquencer et à analyser.

Mots clés : Sorangium lambda, epothilone, hybridation,sondes

Introduction

The myxobacteria are ubiquitous micropredators or saprophytes that live in soil environments. They have an intriguing morphological development culminating in the production of fruiting bodies and myxospores, as reported by (Molnar et al.2000 , Yossi et al.1999) . (Silakowski et al.

(1). State Key Laboratory of Microbial technology, Life Science School, Shandong University, Jinan, Shandong province, People's Republic of China.

(2). Département de Biologie cellulaire et Moléculaire, Faculté des Sciences et Techniques, Université Marien NGOUABI, Brazzaville, Congo

2001) have shown that the myxobacteria belong to a group of bacteria that produce a large number of secondary metabolites, including polyketides.

In the systematic classification suggested by (Silakowski et al.2001); there are two suborders: Cystobacterineae and Sorangineae. Strains of both suborders have been shown to produce a wide variety of bioactive secondary metabolites, for example: -Cystobacterineae (e.g Stigmatella aurantica): auranchines, stigmatellins, myxalamids, myxothiazols. -Sorangineae (So.cellulosum): maracins, maracens, Chivozazoles, epothilones. Many of these structures are most likely synthesized by a type I polyketide synthase (PKS), a non ribosomal peptide synthase (NRPS) or a combination of both (PKS_NRPS).(Molnar et al. 2000) have determined that the epothilone polyketides produced by the Gram-negative myxobacterium Sorangium cellulosum have a narrow antifungal spectrum, and it was established that they have microtubule stabilizing activity with the same mechanism as that of the important anticancer agent taxol. (García-Martínez et al. 1999) have reported that Epothilones are effective against taxol-resistant tumors, making people to perceive epothilones as a potential successor to taxol. (Julien et al.2000), reported in two strains of Sorangium, SMP44 and So90 (GenBank accession numbers AF210843 and AF 217189, respectively) the sequences of the epothilone gene clusters are very similar. an entire 56kb fragment encodes a polyketide synthase (PKS), including a nonribosomal peptide synthase module, and a cytochrome P450 epoxidase. In the present work, two strains of Sorangium cellulosum, already morphogenetically characterized, and tested as epothilone producers, were assayed. We screened the lambda genomic library through hybridization. DNA was extracted from the four clones of double signals with respectively the probe of epothilone A and epothilone B. The recombinant DNA from the four double-signal plaques was Sal I digested. The results were in concordance with the already reported one.

Materials and methods

1-Bacterial strains and vectors

Sorangium cellulosum strains So9881 and So9733-1 screened from soil were available in our laboratory. The characteristics of these two strains are already given by (Yan Z-C et al.2003).The GenBank accession numbers are AY039304 and AF467672 respectively for So9733-1 and So9881. Escherichia coli LE392 used for the infection experiment and BamHI EMBL3,4 lambda Arms used as vectors for cloning were from Promega.

2-Media and culture conditions

The strains were cultured in M26 medium containing (w/v) potato starch 0.8%; glucose 0.2%; Tryptone 0.2%; yeast extract powder 0.2%; CaCl2·2H2O 0.1%; and trace elements solution 1 ml/L. The solid M26 contained 1.5% agar. Before autoclaving, the pH of the medium was adjusted to 7.0. The culture temperature was 30°C.

3-Genomic DNA Extraction

Using the stationary phase bacteria culture, the cells were pelleted by centrifugation and resuspended in STE at about 0.2 g of cells in 1 ml of STE. Then, 0.5 ml of the cell culture was put in a 5-ml centrifuge tube, 0.75 ml of STE was added and the mixture was vortexed. Next, 250 ul of 10% SDS and 25 ul of 20 mg/mg of proteinase K were added. The contents of the tube were mixed and then incubated at 37 .C for one hour (during this hour, the tube was gently agitated every 15 min). After adding 250 ul of 5 M NaCl, the tube was left at room temperature for 10 min, then 250 ul of CTAB/NaCl were added and the tube was incubated at 65 .C for 20 min. CTAB was put in first at 65. C to stabilize the temperature. An equal volume of phenol /choroform/isoamyl alcohol was added, the contents were mixed for one hour with the DNA mixer and then centrifuged at 8000 rpm for 10 min. Finally the supernatant was transferred into a new tube.An equal volume of chloroform/isoamyl was added, the solution mixed with the DNA mixer for 30 minutes, and then centrifuged at 8000 rpm for 10 minutes. The supernatant was transferred in a new tube and two volumes of ethanol were added. After mixing, the tube was kept at room temperature for ten minutes. Following centrifugation at 12000 rpm for 15 minutes,

the supernatant was discarded, 70%
ethanol was added to wash the DNA precipitate, the tube was kept inverted at room temperature for 15 min to dry the precipitate, and then 200 ul of TE was added. The DNA was electrophoretically resolved in 0.6% of Agarose Gel .

4-Partial Digestion of Genomic DNA , recovering of needed restriction DNA fragments (9-23Kb) and ligation to the Lambda EMBL 3 Arms.

Optimization of Sau3AI Partial Digestion of Genomic DNA for So 9881 and So9733 was done. Using the optimized conditions determined above, we undertook a large –scale preparation of high molecular weight genomic DNA for the two strains of Sorangium. V-type tank electrophoresis elution was used to recover the fragments, as developed by (Sambrook J. et al.1989).The ligation conditions were as described in Promega's Protocols and Applications Guide.

5-Packaging reactions and Titering of phages on LB plates.

The packaging and titering were conducted as done by (Hohn, E.G 1979). Using the phage buffer, serial dilutions were made following the standard conditions. The plaque number, titer (pfu/ml) and packaging efficiency were then calculated.

6-Probes Design

(Berg et al.1989) have revealed modular fashion of the organization of the epothilone gene cluster. Each module contains domains; each domain is responsible for one reaction. Our approach consisted in loading the sequence into a software program (Primer Premier), a program used for preparing primers and probes, and according to the parameters choosing the probes the correspond to KS of epoA and PCP of epoB.

7-Plaque blotting

All steps (plaque lifts, lysis and fixation, rinsing, hybridization and detection) were performed with an ECL kit from Promega and the protocols are described in the ECL direct system for plaque screening. Recombinant Phage DNA Extraction and SalI digestion The recombinant DNA in the four clones was extracted using a Wizard lambda Prep DNA Purification System (Promega) Kit, and SalI digested.

Results

Genomic DNA of Sorangium cellulosum strain So9733-1 or So9881 is high molecular weight. Fragments were about 300 to 350 kb as determined by PFGE (Figure 1). Figure 2 is showing the large scale preparation of Sau3AI partial digestion of genomic DNA of So9881, the results was the same for the So9733-1.

The DNA fragments of range between 9 -23 kb were recovered . Assisted computer analysis has given the size of the recovered fragments. We used the following relation to compute the amount of DNA insert to be ligated to the BamHI EMBL3 :left and right arms: ng of (I)= ng (vector) x Kb (size of I)/Kb size vector x molar ratio of I/vector. I = insert. The ligated product was packaged.

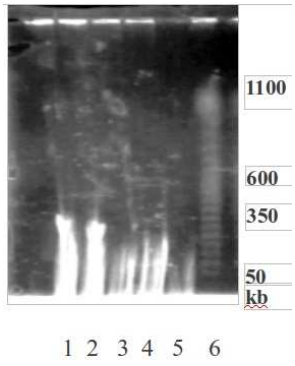

Figure 1 : 0.6% Agarose Gel PFGE of Genomic DNA.

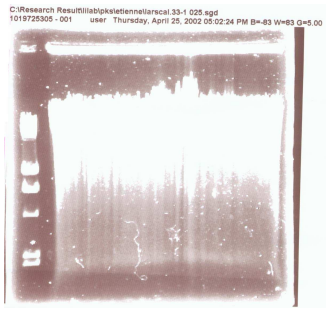

Figure 2 : 0.4% Agarose Gel Electrophoresis of Sau3AI digested Genomic DNA (Large scale preparation) of So.ce 9733-1

Figure 3 shows plaque-forming units of So9733-1 or So9881 on agar plates. The bacterial cells grew to form a lawn in which regions of phage infection appears as cleared areas or plaques. Each plaque contains many phages and represents a clone, meaning that there is one kind of insert in the phages of the same clone. Because many plaques are seen, we can understand that there are many inserts with different sizes. These inserts are different genes or parts of genes, and different sequences of DNA. This result illustrates the polymorphism of inserts (fragments from genomic DNA) in a genomic library.

Figure 4 (a) and (b) are showing the signals during the hybridization of KS of epoA probe and double positive signals. Figure 4: (a) Signals of the in situ hybridization using KSQ of epo A as a probe, (b) double positive signals.

The recombinant DNA in these four clones was extracted and the inserts were SalI digested and resolved electrophoretically. Figures 5 and 6 and are showing the extracted DNA of the recombinant phage and the SalI digestion o

Figure 3: Plate showing the phage plaques (So9733-1).

the related double positive clones, respectively. According to Figure 6, it is obvious that for all recombinant phages, the SalI digested DNA displays two common bands at two specific positions, 23 kb and 9 kb, testifying to the pre-

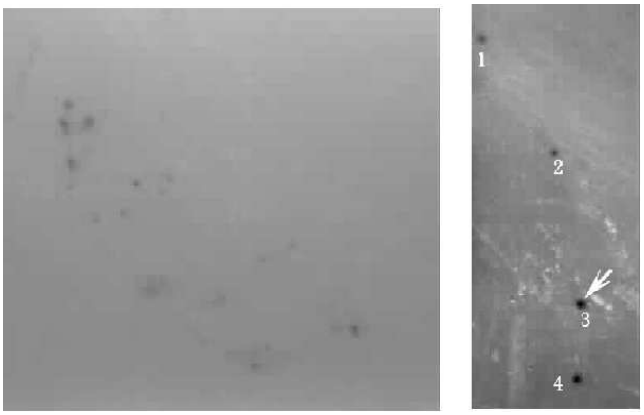

Figure 4 : Signals of the in situ hybridization using KSQ of epo A as a probe (a), double positive signals(b)

sence of the two arms (lambda vectors) released. However each recombinant phage DNA has a characteristic banding pattern suggesting heterogeneity of the inserts. At this point, only sequencing and sequence analysis can give further

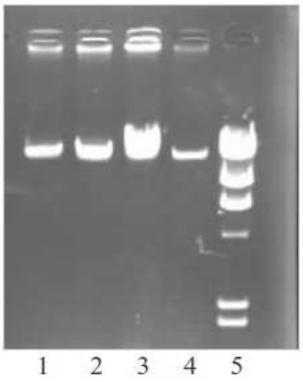

Figure 5 : Agarose Gel Electrophoresis of Extracted and purified lambda recombinant Phage DNA .

information.

Discussion

A genomic library is very useful. Once it is established, we can search for any gene we want. The only problem is that there is no on-line catalog for such a library, so we need probes to tell us which clones contain the gene of interest.

Fragments smaller than 9 kb and bigger than 23 kb cannot be packaged with the The EMBL3,4 vectors . Robert F. 2001 claimed that one clear advantage of the lambda phages over plasmid vectors is that they accommodate much foreign DNA. Tang L et al. 2000 has reported cloning the

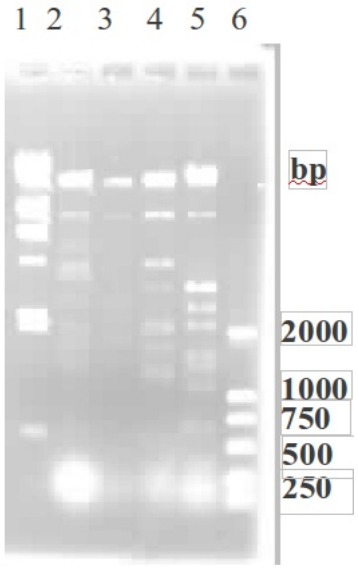

Figure. 6 : 0.8% Agarose Gel Electrophoresis of SalI digested recombinant Phage lambda DNA of four double positive clones.(lanes 2-5)

complete sequence of the gene cluster responsible for epothilone biosynthesis in the myxobacterium Sorangium cellulosum. So whe use an alternative method to search for the epothilone gene cluster and compare the results, to test the efficiency of our method.

Whenconstructing a library , the PFGE is used to determine the exact size of the extracted DNA fragments, for Long fragments are advantageous for constructing a library.

The transfection process requires some specific conditions, in particular, the temperature is very important for phage particles to infect E.coli. In our case, the packaging efficiency was acceptable for further screening of the libraries by any method. The assisted computer analysis gave the average size of the fragments in the library at about 13 kb. If a microbial genome is cloned as a library of small uniform random fragments, the frequency of the gene X sequences in the library reflects that in the genome (Farrelly, V et al.1995) . In our case, the microbial genome is cloned in a library of non-uniform random fragments, and there is a polymorphism between the restriction fragments. Considering the Poisson distribution, the possibility that gene X is in the library is between 95% to 98%. If multiple copies of gene X are present in the genome, the same statistics can be applied. Thus, for example, if n copies of 10 kb gene X were present in the genome, then on average, nX fragments would be found in the library. With the knowledge of the genome size and assumption of the average size of a gene X, the identified gene could be used to estimate the total number of X genes in the genome. In our case it is important to understand that there are multiple copies of PKS, and sequence homology can lead to the screening of a PKS which is not the goal. Nowadays it is possible to calculate the theoretical number of plaques in the library, as particular library which it is necessary to screen to have a probability of finding any given sequence. This number is calculated by the following formula :

$N = A/B \quad A = \ln(1-P) \quad B = \ln(1-I/G)$

N: number of plaques that should be screened,
P: Probability of having a given sequence represented in this number of clones

I: is the insert size of the vector G: is the genome size of the organism .

The myxobacterial genome size is about 9000 kb, so it may be necessary to screen a great number of plaques.

The epoA and epoB genes are 8.5 kb in length: epoA is a loading module while epoB is an NRPS. (Farrelly,V et al. 1995) reported that, because the amino acid sequence of individual domains of modular PKSs is usually quite similar, an approach of generating probes from the results of the PCR products with degenerated primers can be successful. However primers and probes are often imperfect, and adding to the fact that organisms often contain multiple PKS gene clusters, probes can reveal a PKS which is not sought . Sorangium cellulosum possesses multiple PKSs, and there is a high sequence similarity among domains of different modules,which means that synthesis of probes can be missed. According to the already published structure of the epothilone gene cluster , (Jung-Hoon Yoon et al.1997) showed that the most important differences in the PKS gene cluster are located in the loading module , where the ksq gene is different from other ks genes of other modules of the cluster. Another feature is the presence of the NRPS located downstream of epoA. Using the software Primer Premier 5.0, , we matched all the parameters that can allowed us to choose the probes from the domain KS of epoA and the domain PCP of epoB.

According to analysis of the results in Table 1, it is remarkable that both probes have acceptable records, no hairpin, cross dimer are not found. The two probes were used with the BLAST search against the databases of sorangial sequences and we confirmed the efficiency of the probes.

From the 24000 plaques tested with the KS probe of epo A, we obtained 78 positive clones, then these positive clones were tested with the PCP probe of epoB. Only 4 plaques gave simultaneous signals with the two probes for epoA and epoB. On the other hand we found about 18 positives clones with only epoB probes. (figure not shown) SalI is an enzyme which can provide a helpful digestion map to analyze the heterogeneity of the fragments epoA and epoB (Table 2).We analyzed the enzymatic sites and the fragments present after Sal I digestion of the modules of epoA and

epoB, and compared them with the partial upstream and dowstream sequences reported for the epothilone gene cluster from strains SMP44 and So90.

The results of Figure 6 connected to the analysis of the results of the Table 2 will help when the inserts of any of the four clones will be sequenced. Regardless, sequencing is needed to acquire more information about the genes.

Conclusion

The method used in the search of epothilone gene cluster in Sorangium using lambda genomic library is an innovation, for all people has used cosmid. The screening of the library by hybridization

Table I : Design and characteristics of the probes for hybridization of epothilone genes

	Probe 1	Probe 2
Sites in sequences	2954bp	10243bp
Modules	*epoA*	*epoB*
The genes	*ksq*	*pcp*
Sequences	5'-ACA CGG CAC GGG CAC GAC GCT-3'	5'-GCT CGG GCT GGA GGT GGT TGG GC-3'
Melting temperatures	74.3	76.2
Degeneracy	1	1
Hairpin	None	None
Dimer	None	None
False primer	Found	Found
Cross dimer	None	None
Activity (µg/OD)	32.8	34.1
Rating	68	71

has allowed us through the designed probes to detect double signals. The restriction map with SalI upstream and down stream of the epothilone gene cluster is consistent with the structure of the already published one. The method can be used to screen any gene cluster.

.

AcKnowledgements:

Table II : The enzymatic sites and the fragment sizes of *SalI* on the modules *epoA* and *epoB* and partial upstream and downstream sequences of reported epothilone gene cluster.

Strain SMP44			Strain So ce90		
Modules	Enzymatic sites (bp)	Fragment sizes (bp)	Modules	Enzymatic sites (bp)	Fragment sizes (bp)
upstream	1229		upstream	6235	
1-1997bp	1382	153	1-7609bp	6385	150
	1663	281		7274	889
EpoA	3783	2120	EpoA	9395	2121
1998-6263bp	4935	1152	7610-10875bp	10547	1152
	6086	1151		11698	1151
EpoB	6680	594	EpoB	12292	594
6260-10492bp	7655	975	11872-16104bp	13267	975
	9230	1575		14371	1104
EpoC	10789	1559	EpoC	16401	2030
10639-16137bp	10849	60	16251-21749bp	16461	60
	12844	1995		18461	2000

We thank Professor Li Yue Zhong for providing material and the financial support. We thank Dr. Pamela Holt for editing the manuscript.

REFERENCES

Berg, K. L., Squires, C., and Squires, C. L (1989).: Ribosomal RNA operon antitermination. Function of leader and spacer region box B-box A sequences and their conservation in diverse microorganisms. J. Mol. Biol. 209 . 345-358

Bryan Julien, Sanjay Shah, Rainer Ziermann, Robert Goldman, Leonard Katz, Chaitan Khosla. (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. Gene 249. 153-160

Farrelly, V., Rainey, F. A. & Stackebrandt, E. (1995). Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl Environ Microbiol 61, 2798–2801.

. García-Martínez, J., Acinas, S.G., Anton, A. I., and Rodríguez-Valera, F. (1999) Use of the 16S–23S ribosomal genes spacer region in studies of prokaryotic diversity. J. Microbiol. Methods. 36. 55–64

.Hohn, E. G. (1979) In vitro packaging of . and cosmid DNA. Methods in Enzymol. 68. 299–309
Jung-Hoon Yoon, Sung Taik Lee, Sam-Bong Kim, Michael Goodfellow and Yong –
Ha Park.(1977) Inter-and Intraspecific Genetic Analysis of the Genus Saccharomonospora with 16S to 23S Ribosomal DNA (rDNA) and 23S to 5S rDNA Internally Transcribed Spacer Sequences. International Journal of Systematic Bacteriology. p661-669

Molnar I, T Schpp, M Ono, RE Zirke, M Milnamow, b Nowak-Thompson,…..The Biosynthetic gene cluster for microtubule-Stabilizing agents epothilones A and from Sorangium cellulosum So ce90..Chemistry & Biology 7 (2000) 97-109.

Robert F. Weaver .(2001) Molecular Biology, . New York, NY. ..7 . 773-785

Sambrook, J. Fritsch, E.F. and Maniatis, T. (1989)

A laboratory manual Molecular
Cloning: 79-81

Silakowski B, Nordsiek G, Kunze B, Blöcker H, Müller R. (2001) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium Stigmatella aurantiaca Sga15 Chem Biol. 8. 59-69.

Tang L, Shah S, Chung L, Carney J, Katz L, Khosla C, Julien B. (2000) Cloning and
heterologous expression of the epothilone gene cluster. Science. 28. 287(5453):640-2

Yan Z C, Wang B, Li Y Z, Gong X, Zhang H Q, and Gao P J. (2003)
Morphologies and phylogenetic classification of cellulolytic myxobacteria. Syst. Appl.
Microbiol., 26(1) 104-108

Yossi Paitan, Elisha Orr, Eliora Z. Ron, Eugene Rosenberg. (1999) Cloning and
Characterization of Myxococcus xanthus cytochrome P-450 hydrolase required for the biosynthesis of the polyketide antibiotic TA. Gene 228 147-153