SEQUENCING OF dsDNA

8th-12th of November 2002

Assistants:

Pavel Pavlo

Shashi Bushan

Name: Evangelos Papadopoulos.

# INTRODUCTION

DNA sequencing is a fundamental method used in biology research to help us identify and record the base sequence and composition of the actual nucleic acids of living organisms or artificially synthesized DNA.

The method we used in this lab (dideoxy method) was developed by F. Sanger. It makes use of the mechanism of DNA synthesis by DNA polymerizers. The DNA to be sequenced must be first denatured to single strand DNA. After that an appropriate primer is used. This is a short DNA fragment that is known to be complementary to one position in the original DNA. The primer binds to its appropriate site and then the DNA polymerization starts in the presence of DNA polymerizer and the four deoxynucleosides, dATP, dCTP, dTTP, dGTP.

But a small amount of ddATP, ddCTP, ddTTP and ddGTP is added to four different samples. The composition of dideoxy ddNTP (N= A, C, T, G) does not allow further continuation of the polymerization reaction. Therefore the reaction stops. The exact position where the reaction will stop is a random event for every molecule but since in each tube we have only one type of ddNTP we know in what type of base the reaction will stop.

In this way the molecular weight of the synthesized DNA fragments is an indicator of the positions of the according bases. After that we can use any kind of technique, like electrophoresis to determine the order of the different molecular weights or sizes synthesized and therefore sequence the gene.

MATERIAL AND METHODS

We were provided with the DNA sample and the primers. I used the SK primer. Then we proceeded according to the experimental procedure that is described in the laboratory notes.

1. First step is the Denaturing of ds DNA.

This is accomplished by the addition of 6M NaOH.

2. Then annealing of the primer to the DNA.

Just by adding the primer and a buffer in the appropriate conditions.

3. After we have the denatured DNA with the primer bound on it, it is ready for beginning the polymerization reaction. We prepare 4 different tubes containing the corresponding mixtures for termination of the polymerization in A, C, G and T.

4. We add polymerase and radioactive dNTP for labeling to the DNA+primer mixture. And we leave it for sometime so as the polymerization reaction starts taking place.

5. Then we transfer a portion of the above polymerizing mixture into each of the 4 base terminating tubes. The reaction is therefore terminated but in different position in each tube. Particularly in the A tube it will terminate in an A base etc.

6. We prepared the electrophoresis polyacrilamide gel and we let it polymerize for one day.

7. We loaded the samples in the gel and run the electrophoresis for two hours.

8. We took the gel out in a paper washed and dry it.

9. We used a radiation scanner to view the results.

# RESULTS

The sample we used for some reason did not work out well. The quality of the scanned gel image was very poor and although we could see the base lines their position was not clear enough to allow sequencing. Therefor we were given another gel image to analyze. (See: Page I )

In page II is a printout of the gel image and the sequence that was derived from that. I used this sequence consisting of 55 base nucleotides to try and locate the gene from which they come.

We used the internet database of NCBI to identify the gene. The results of the query are in pages II-IV. As you can see it has an almost perfect match with the genome of Azotobacter vinelandii. There were only 2 gaps. Page I. The first gap is a G base which although it is not present we can see a blank row in the gel. So it was something missing there. And the second gap is in a position where there should be 4 C-bases but we detect only three. Probably because of some base loop that the strands form and/or from not perfect denaturing.

In particular this base sequence is located in a protein encoding gene for a dimmeric 2Fe-2S protein that is not essential for aerobic nitrogen fixation but confers oxygen protection to the nitrogenase in vitro and in vivo. An essential task as the time of half-life of nitrogenases in the presence of Oxygen vary from 30 sec to a few minutes.

The base sequence encoding this protein is in Pages V-VI. And you can see highlighted the sequenced portion. Which is reverse and complementary to the gel sequence.

# COMMENTS

Answering questions

1) Why is it necessary to pre-run the gel before loading the sample?

The electrophoresis gel was pre-run half an hour before loading the

DNA sequencing sample. This is done for heating up the gel before so as to be in a temperature equilibrium. Like preheating an oven. The diffusion process is temperature dependant and an uneven distribution of the temperature throughout the sample would lead to different electrophoretic results in various positions in the gel.

The same is valid for other chemical component gradients that have to be in an equilibrium state for the electrophoresis to work reliably.

Moreover, we had some release of extra urea from the saturated gel. This had to be washed out before loading the sample or otherwise might block the pathway for initializing electrophoresis.

2) Why do we read the sequence from the bottom of the gel and not from the top?

The sequencing is accomplished by differentiating the size or/and molecular weight of the different produced DNA fragments. Therefore, if for example an A base is close to the primer and another one is farther away from the primer we will have in the A terminating tube a short fragment of copied DNA ending in A and a long fragment of DNA ending in A too. In electrophoresis the smaller and lighter molecules propagate faster and so cross a bigger distance while for the bigger and heavier molecules propagation is more difficult and they remain closer to the original position.

We conclude that a small DNA fragment ending in A means an A base is closer to the primer and we will find it more away from the original position (Top) and closer to the bottom. So, we have to start reading from bottom to the top.

It is also important to note that the gel gives the sequence complementary to the template where primer binds.

3) You cloned your gene of interest, which is approximately 1000bp into a vector. Now you want to sequence that gene in order to find out the nucleotide sequence. After the first sequencing, you have successfully sequenced 512bp. How will you find the rest of the sequence? (two alternatives)

We can find the rest of the sequence depending on the position of the already sequenced fragment in our gene of interest and our knowledge of the vector. Generally, we can continue by using primers derived from our first sequencing or use primers appropriate for the cloning vector.

Let’s say for example that the first 512 bp we have sequenced are the first in the 3’-5’ direction then we have another 488 bp. We can chemically synthesize another primer complementary to a few of the last sequenced bp. By using this primer, we can follow the same experimental technique as in our practical and continue sequencing of the gene in the 3’-5’ direction. We can also continue sequencing in the opposite direction if we use a primer equal to the first few bp of our gene. Because this primer will bind to the complementary strand and continue sequencing in the reverse direction.

However, this technique has the disadvantage that we should chemically construct a primer after our first sequencing. But if we know the sequence of our vector a few base pairs around our gene insertion point we could have already prepared appropriate primers to proceed sequencing both directions downstream and upstream. So by 512bp sequencing we would have easily reconstructed the whole 1000bp long gene with some possible overlapping in the middle for confirmation and correct spacing of the two sequences.

4) Define cDNA. Which are the advantages of sequencing cDNA?

cDNA is the DNA synthesized from an mRNA template by the use of reverse transcriptases. These are enzymes originally located in retrovirus and they serve the inverse transcription required for them to synthesize DNA so as to exploit the invading cell translational system. cDNA is an important tool in genome research since it allows the sequencing of a particular mRNA coding for a particular protein. So, it helps discriminate between introns and exons (expressed and not expressed sequences) in the original DNA.