An experiment of using e. coli cells to help creating egfp genes

Health & Medicine

Introduction

The EGFP, enhanced green fluorescence protein, gene is a commonly used reporter gene because it allows for easy determination of gene expression due how it codes for proteins that emit visibly green light when exposed to ultraviolet light. EGFP gene is a mutation of the GFP, green fluorescence protein, that produces 35 times as much light when exposed to ultraviolet light. The emitted light from the protein peaks at 509 nm and becomes excited at 488 nm (Tsien, 1998). The original GFP gene was isolated from the jellyfish Aequorea victoria (Zhang, 1996), from which many different variations have been produced and have been expressed in many different organisms from bacteria to mammalians. E. coli cells were used in this experiment to amplify the number of plasmids containing EGFP genes for further analysis of the success of ligations used to make the recombinant plasmids containing this gene. In order for the EGFP gene to be visibly expressed in the E. coli cells it was necessary to use IPTG since this gene is controlled by a lac operator. The use of polymerase chain reaction allows for further amplification of the EGFP gene for analysis by splitting the plasmid DNA lengthwise, whereupon primers attach to complimentary nucleotide sequences on the DNA for the region that is intended to be copied. Taq polymerase is then used to finish the DNA sequence where primer strand ends, using nucleotides that are in solution during the PCR. After 30 cycles PCR billions of double stranded DNA containing the EGFP gene may be produced from a relatively small number in the initial sample.

This series of experiments was designed to create a recombinant plasmid by ligating the EGFP gene cut from the pEGFP-N1 plasmid using the restriction enzymes NcoI and NotI onto a pET-41a(+) vector, which was cut using the same enzymes, transforming this new plasmid into competent E. coli cells and expressing the gene using IPTG in a portion of the cells, and not in another portion that would serve as an unknown sample to test different analytical techniques for identifying the presence of the EGFP gene in the transformed plasmids. A portion of the ligations were saved and run through gel electrophoresis for the purpose of characterization and identification of the unknown sample. The E. coli cells that were successfully transformed would have been made identifiable by their kanamycin resistance that they would have acquired from the pET-41a(+) vector, since these cells were grown on a Luria broth medium containing kanamycin. The transformed cells were then run through a miniprep to isolate the plasmids they had taken up and multiplied; these plasmids were run through gel electrophoresis to characterize them. A portion of the plasmids from the miniprep were run through a restriction digest, using NotI and NcoI, just NcoI, or neither to serve as a control. These restriction digests were subsequently run through gel electrophoresis, this served to identify the presence of EGFP gene since the cut sites for both enzymes should isolate the gene from the vector. In order to characterize the plasmids from the miniprep they were run through a PCR using pAD1 sense primer and pAD1 anti primer to isolate a section of the DNA sequence of a predictable size that would contain most of the EGFP gene and a portion of the vector. Online bioinformatics tools were used to identify the size and sequence of the vector, EGFP gene, recombinant

plasmid, and subsequent PCR product in order to make the characterization of them through gel electrophoresis possible.

Methods

The initial experiment performed was a ligation of the EGFP gene into a pET-41a(+) vector. In order to do this a sample of pET-41a(+) vector that had been cut using the restriction enzymes NcoI and NotI and this was ligated to a sample of EGFP gene that had likewise been cut using the same enzymes. Five separate ligations were performed. The first was 1: 1 molar ratio of the vector with the EGFP gene with 5 ng of vector per µl and 0. 35 ng EGFP per µl. The second ligation had 5 ng of vector per µl and 1. 05 ng EGFP per µl for a 1: 3 molar concentration. The third ligation was a positive control containing a recombinant plasmid DNA of the vector ligated to the EGFP gene. The fourth ligation was a negative control only containing water, and fifth was a second negative control with the same concentrations of as the second ligation but without DNA ligase. The five samples of the ligations were run through gel electrophoresis with a DNA ladder for comparison, the gel had a concentration of 0. 8% agarose and was run at 100V. The gel was allowed to run until the track dye was visibly 2/3 of the way across, it was then stained with GelRed for 20 minutes on an orbital shaker and then viewed under a UV transilluminator and photographed.

A transformation of the previously completed ligations was performed in E. coli cells to express the recombinant plasmids in them. Four separate transformations were performed corresponding to ligations one through four above. During the transformation the E. coli cells were kept on ice, then heat shocked at 42°C for 2 minutes before being returned to ice for another 2

minutes. The cells were then added to room temperature Luria broth and placed in a shaking incubator for 45 minutes at 37°C. Six different petri dishes were prepared, the first four of which contained Luria broth, kanamycin, and IPTG, the fifth plate contained Luria broth and kanamycin, and the sixth plate contained only Luria broth. For the first four plates an 80 µl sample of their correspondingly numbered transformations were plates, while the fifth plate received a 20 µl sample of the second transformation and the sixth plate received a 20 µl sample of the fourth transformation. These plates were incubated overnight at 37°C. These plates were illuminated under UV light and the number of colonies was counted and the efficiency of the transformations was calculated.

A miniprep of the recombinant plasmids transformed into the E. coli cells was performed using the Promega wizard miniprep kit. From the plated non-control E. coli sample a colony that fluoresced green under UV, hence forth called the recombinant sample, was collected from a Luria broth, kanamycin, IPTG plate. Another colony that did not fluoresce on the Luria broth, kanamycin, IPTG plate, hence forth called the non-recombinant sample, was collected. A third from the plate with only Luria broth and kanamycin was also collected; this will be called the unknown sample. These three samples were run through a miniprep and a portion of each was run through gel electrophoresis with a 0. 8% agarose concentration along with a DNA ladder. The sample was run at 60V until it entered the gel, then at 110V until it was half way through. The gel was stained with GelRed for 25 minutes and then viewed under a UV transilluminator and photographed.

The remaining plasmid DNA from the miniprep was run through a restriction digest. The recombinant, non-recombinant, and unknown sample plasmids were each given three separate restriction digestions: a digestion with both enzymes NotI and NcoI (which shall be called the double digest), a digestion with only NcoI (which shall be called the single digest), and a third without any restriction enzymes that served as the negative control, yielding 9 different samples. The restrictions were then incubated at 37°C for an hour. Gel electrophoresis was performed on these samples with a DNA ladder and a 0. 9% agarose gel at 60V until the samples were visibly 2/3 across the gel. The gel was stained using GelRed for 25 minutes and then viewed under a UV transilluminator and photographed.

Five polymerase chain reactions were performed on the plasmids isolated from the miniprep, including a positive and a negative control. For the recombinant, non-recombinant, unknown, and positive control pET41/EGFP plasmids a concentration of 0. 5 µM pAD1 sense primer, 0. 5 µM pAD1 anti primer, and 10 ng plasmid DNA per µl was used. The negative control contained the same concentrations, without the plasmids. The PCR was run for 10 minutes at 95°C, then for 30 cycles of 95°C for 1 minute, 56°C for 1 minute, and 72°C for 1. 5 minutes, followed afterwards by 5 minutes at 72°C and then held at 4°C. Gel electrophoresis was performed on the PCR products with a 0. 9% agarose gel at 60V until the DNA had visibly crossed ¾ of the gel. The gel was stained with GelRed and viewed under a UV transilluminator and photographed.

Bioinformatics was performed for the EGFP, pET41 plasmid, their recombinant plasmid and its PCR product made using the pAD1 sense primer and pAD1 anti primer. The NCBI nucleotide database at http://www. ncbi. nlm. nih. gov/ was used to find the pEGFP-N1 plasmid sequence, which was then cut using the online WatCut restriction analysis tool, which is available at http://watcut. uwaterloo. ca/watcut/watcut/template. php, using the NcoI and NotI restriction enzymes. This made it possible to isolate the EGFP gene sequence. Then using the AddGene Vector Database, found at http://www. addgene. org/vector-database/2592/, the pET-41a(+) reverse compliment sequence was found, then cut with NcoI and NotI using WatCut and the EGFP sequence was ligated in the cloning site. The PCR primers', pAD1 sense primer and pAD1 anti primer, annealing sites were found in the EGFP/pET41 sequence and the sequence for the PCR products were found. A combination of further restriction enzymes was considered to fragment the PCR product to verify its sequence by determining the sizes of the sequences.

Results

The positive control containing the recombinant plasmid showed two bands in the gel, the first at about 6 kilobases and the second at approximately 750 base pairs. The water negative control showed no bands. Two distinct bands were seen in the negative control that had no DNA ligase, the first at above 10 kilobases and the second at about 4. 5 kilobases, it also displayed a faint third band considerably above 10 kilobases that cannot be interpreted. The negative control contained EGFP gene and the pET41 vector. The 1: 3 pET41 to EGFP concentration ligation showed two distinct bands at 5 kilobases and

about 750 bases. The 1: 1 pET41 to EGFP concentration ligation displayed the same bands as the 1: 3 concentration ligation, but were fainter.

The plated transformed E. coli colonies from the 1: 1 vector to EGFP concentration ligation showed a single colony that fluoresced green on the plate with Luria broth, kanamycin, and IPTG. The plate 1: 3 vector to EGFP plate with the same media had one colony that fluoresced green and another that did not. The positive control plate had 13 different green colonies and none that were white. The negative control LB/kanamycin/IPTG plate that only had water did not have any colonies. The plate with only Luria broth and kanamycin that used the 1: 3 vector to EGFP concentration did not grow any colonies. This plate was to serve as the unknown for other experiments, so a separate unknown sample was acquired. The negative control on the Luria broth only plate was saturated in colonies.

Figure 2. Gel of transformed plasmids. Track 1 New England Biolabs 1 kb ladder, Track 2 recombinant sample, Track 3 non-recombinant sample, Track 4 unknown sample.

Figure 3 . Gel of restriction digest.

The recombinant sample displays two distinct bands, the first above 10 kilobases and a second at about 5. 5 kilobases, with some faint bands at above 10 kilobases. The recombinant sample contains EGFP gene in the pET41 vector. The non-recombinant sample, which contains the pET41 plasmid without the EGFP gene inserted, features two distinct bands like the recombinant sample, but shifted slightly lower, with the first at about 10

kilobases and a the second at about 4. 5 kilobases. Some faint banding can be seen above the 10 kilobase range in the non-recombinant sample. The unknown sample features banding identical to the recombinant sample with two bands in the same locations and faint banding above 10 kilobases. The unknown sample contains the pET41 plasmid and may contain the EGFP gene.

The DNA ladder in the restriction digest gel did not appear, which made interpreting data from the gel difficult and inaccurate. The double digested showed one significant band and contained the pET41 plasmid DNA strand with a small section removed that does not appear in the gel. The singly digested non-recombinant sample features one band that may be slightly higher than the previous sample. The undigested non-recombinant sample shows four distinct bands, the lowest two being most visible and the others showing more faintly above what is most likely 10 kilobases. The double digested recombinant sample shows a distinct band at about the same level as the non-recombinant double digest, but also shows a very faint band at what may be the 750 bases range. The singly digested recombinant sample shows one band higher than the distinct band for the doubly digested recombinant sample, likely in the 5 kilobases range. There are several bands in the undigested recombinant sample, much like the undigested non-recombinant sample, however all the bands are shifted upwards into higher base number ranges. The doubly digested unknown sample shows two bands in the same locations as the doubly digested recombinant, but much fainter.

Likewise the singly digested unknown shows a single band in the same location as the singly digested recombinant but fainter. The undigested unknown shows several very faint bands in the same locations as the undigested recombinant.

The gel of the PCR products shows heavy signs of contamination, but some of the bands remain interpretable. A significant band is present at about 1. 2 kilobases for the unknown sample. This same band at 1. 2 kilobases is also present in the recombinant sample. A heavy band is present at what might be 200 bases for the non-recombinant sample, along with three faint bands at about 350 bases, 1000 bases, and at approximately 6 kilobases. The positive control shows similar banding to the unknown and the recombinant samples with a significant band at 1000 kilobases. The negative control has a single band very low in the range far lower than 500 bases; this sample only contained the primers.

The bioinformatics research performed on the EGFP gene and the pET-41a(+) determined the sequence and size of the resulting recombinant plasmid and PCR products. The use of NcoI and NotI on the pEGFP-N1 yielded the EGFP gene with 724 base pairs. The pET-41a(+) plasmid cut with NcoI and NotI produced a DNA strand 5859 base pairs long. The EGFP gene ligated onto the pET-41a(+) vector produced a recombinant plasmid with 6583 base pairs. A PCR product for this plasmid was found when using the primers pAD1sense primer and pAD1 anti primer, these primers copied the sequence between a section of the EGFP and pET-41 vector respectively to produce a DNA strand 1182 base pairs long. A method that could be used to

determine if the PCR sequence is the intended product would be to use restriction enzymes to cut the sequence into smaller segments of predictable sizes, knowing where the cut sites would be. Using the enzyme BseRI the PCR would produce would be cut into 2 segments at base pair number 496, these two segments would be of length 496 base pairs and 686 base pairs. The restriction enzyme EaeI could also be used; it cuts the PCR sequence at the 537th base pair and 926th base pair, yielding three segments of lengths 537, 256, and 389 base pairs. If both of these restriction enzymes were used together four segments of DNA would be formed or lengths 496, 389, 256, and 41. The segment of length 41 would likely not be easily noticed in gel electrophoresis, but the double digest of EaeI and BseRI could be compared against an EaeI digest, comparing the location of the band of base pair length 537 to the band of base pair length 496.

Discussion

The ligation initially performed showed positive confirmation in the gel electrophoresis that it went successfully, the attempted ligations at concentrations 1: 1 and 1: 3 vector to EGFP ratios exhibited similar banding to the positive control suggesting they ligated successfully. The lower band in the 1: 1, 1: 3 concentration samples and positive control are likely excess EGFP segments that did not ligate. These can be compared against the negative control, the lowest reaching band for which is lower than the similarly placed bands for the positive control and the 1: 1 and 1: 3 concentration sample ligations, which suggests the negative control DNA strand has fewer base pairs and therefore was able to more through the gel quicker. This is to be expected since it did not contain the EGFP sequence in

the vector plasmid. The banding in the negative control may be dimers or trimmers that formed between the vector plasmids.

The plated transformed E. coli cells, other than the positive control, produced few colonies. The 1: 1 and 1: 3 vector to EGFP concentration transformations each produced a single colony that fluoresced green and expressed the EGFP gene. The 1: 3 vector to EGFP concentration transformation also yielded a single colony that did not express the EGFP gene and appeared white. The positive control sample produced 13 separate colonies, each of which expressed the EGFP gene and fluoresced green under UV light. The plate that only used a sample containing water did not produce any colonies, which was to be expected and confirmed there was no outside contamination. However, no colonies formed on the 1: 3 vector to EGFP gene transformation plate, this was to serve as the unknown for further experiments, so it was necessary to acquire unknown from another source. This may have been due to improper plating technique, there may have been alcohol still on the spreader, or the spreader may have been too hot from the flame used to sanitize it. The second negative control that was on the Luria broth only plate was entirely covered in the colonies. This control showed how necessary kanamycin was to the experiment, since the kanamycin resistance was part of the pET-41 vector it only allowed E. coli that had taken in the pET-41 vector or the recombinant plasmid to grow.

In the gel containing the transformation all sample tracks exhibited several bands, which suggests that the higher bands may have been dimers or other polymers of the plasmids, this is because each track should have contained

either the vector plasmid or the recombinant plasmid containing the vector and the EGFP gene. The known recombinant plasmid sample and the unknown sample show the same bands as each other, which suggests that the unknown was in fact a recombinant plasmid. The recombinant and unknown samples can be compared to the non-recombinant sample which showed similar banding, but was shifted down a small amount compared to the other tracks. This shift would be due to how it did not carry the EGFP gene was lighter and able to progress through the gel quicker; this helps confirm that the unknown sample likely contained the EGFP gene.

The restriction digest gel for the singly and doubly digested non-recombinant plasmid showed similar banding, which was to be expected since the use of NcoI and NotI would have meant only section less than 100 base pairs of the sequence was cut from the plasmid, making it difficult to differentiate from them the singly cut plasmid that retained all of its base pairs. All of the undigested samples showed several bands, which suggest there were some dimers and other polymers being formed between the plasmids as well as some supercoiling of the plasmids. The singly and doubly digested recombinant plasmids showed more significant differentiation than the non-recombinant, this was to be expected since the use of both NcoI and NotI would have cut away the EGFP sequence from the plasmid, making it significantly lighter and able to pass through the gel quicker than the singly digested recombinant plasmid which retained all of its base pairs. If you compare the single digest recombinant plasmid to the single digest non-recombinant plasmid the non-recombinant it is further along the gel, which suggested the recombinant is heavier and contains the EGFP gene. Further,

the double digest recombinant shows a fainter band further along the gel, which is where one would expected the much lighter EGFP gene to be located. The unknown samples exhibited similar banding to the recombinant sample, but much lighter. This suggests that the unknown was a recombinant plasmid, this can be seem especially in the double digest of the unknown where a very faint band can be seen at the same location where the double digest recombinant likely had its EGFP gene. This suggests that the unknown contained the EGFP gene.

For the PCR gel there were heavy signs of outside contamination. This contamination was possibly from debris falling into the solution or contamination in the micropipette, cross contamination from using the same micropipette tip was not a factor; however it may have been possible that some of the micropipette tips became contaminated when the box containing them was left open. However, the unknown, recombinant, and positive control all displayed similar banding when compared against the non-recombinant or the negative control samples. The PCR may have been partially successful because for the unknown, recombinant, and positive control samples the heaviest banding occurred at about 1100 base pairs and the expected PCR product was 1, 182 base pairs. The non-recombinant however also showed a band at about that range also, but the most significant band for the non-recombinant took place at about the same location as the negative control which showed bands confirming the presence of the primers only. A second attempt of the PCR under cleaner conditions would likely yielding better results without ambiguity, but it

seems most likely that the PCR product for the unknown contained the EGFP gene when compared to the recombinant or positive control.

Through the course of running these several experiments and the analytical methods used the presence of EGFP gene in the unknown sample can be reasonably confirmed despite the lack of any visible expression. The several gel electrophoresis runs that were performed during the different stages of the experiment showed repeatedly that the unknown plasmid was characteristically identical to the known recombinant sample in size and quality of bands formed. The only area of concern during the experiment was the significant contamination of the PCR product, despite the PCR producing in both the unknown and known recombinant banding that was predicted by the bioinformatics for sequence containing the EGFP gene. A second PCR under cleaner conditions is necessary since the non-recombinant also produced a faint band in the location where the EGFP PCR product was expected to occur, which should not have been possible. The use of other restriction enzymes following the PCR could potentially salvage the experiment, or at least identify if the non-recombinant was indeed contaminated with the EGFP gene at that stage by seeing if fragments of the sequence at sizes that would be expected from using certain restriction enzymes. Further experimentation could isolate the expressed EGFP protein in the transformed E. coli cells and further characterize the proteins and identify them, perhaps expressing the EGFP gene in the unknown sample E. coli and comparing their proteins to the known recombinant.