Experiment 4
Amplification of a GAF-4 Domain Sequence by the Polymerase Chain Reaction

Objectives

1. Primer design for the
    a. PCR amplification of a GAF-4 domain from the cyanobacterium, Nostoc punctiforme.
    b. directional cloning of the amplified GAF domain into an expression plasmid.
2. Amplify the GAF-4 domain DNA using PCR methodology.
3. Perform a restriction enzyme digest and agarose gel analysis on the PCR generated DNA fragment to confirm the identity of the amplified DNA.
4. Quantify DNA spectrophotometrically.

Introduction

To determine tertiary structures of proteins, and to study the protein function, it is critical to isolate a protein of interest away from other proteins (i.e. to purify the protein of interest). Methods and strategies of protein purification are covered in Experiment 5 in this manual. In the post-genomic era, the first approach to protein purification is often to use molecular cloning techniques, which are fast and simple. Molecular cloning for the purpose of synthesizing and isolating a protein of interest involves copying the protein's gene and inserting it into an expression plasmid. The plasmid is stored and propagated in a host organism. When a cloned protein is needed for isolation and experimentation, protein expression is induced in the host organism, and later, isolated. Molecular cloning also avoids difficulties with the de novo purification of proteins, such as integral membrane proteins, transient signal transduction proteins, heterogeneous populations of a protein (isozymes expressed from gene families, or proteins that are differentially processed), or low abundant proteins. Methods of molecular cloning are rapidly evolving. Molecular cloning by means of PCR amplification and directional cloning are discussed below; plus a brief description of Gibson Cloning, which is a recent and highly popular method. However, no matter what technique is chosen, cloning a gene for the purpose of expressing a (functional) protein requires attention to the components of gene structure and the primary protein structure.

For biochemists and cell biologists, molecular cloning techniques also permit experimentation on portions of proteins, known as domains. Protein domains are identified by their tertiary structure in addition to a simple primary structural determination. Domains have specific structural and/or functional roles within a larger protein; and a specific domain may be found in many proteins of different functions. Structural motifs that are short conserved stretches of sequence found within domains; for enzymes, a motif may identify the active site. However, a particular motif found within a domain may or may not possess the function implied by the motif; in such cases, the motif may be retained to maintain the tertiary structure of the functional domain. The stretches of DNA coding for domains can be switched between different DNA sequences coding for other proteins, hence creating new proteins with modified functions. The

evolution of proteins in part involves molecular mechanisms placing one domain next to another by well-established events in the genome: transposition, rearrangements (inversions, translocations, deletions and duplications), and homologous recombination. There are many families of protein domains found in all kingdoms of life. The phytochrome family of photoreceptors is an example of a protein family constructed of domains (Figure 4.1). The goals of studying the structures and (possible) functions of proteins and their domain substructures are to understand the origin and evolution of acquired functions, to discover altered or new functions of proteins, and to engineer proteins toward specific functions for improvements in health and agriculture.

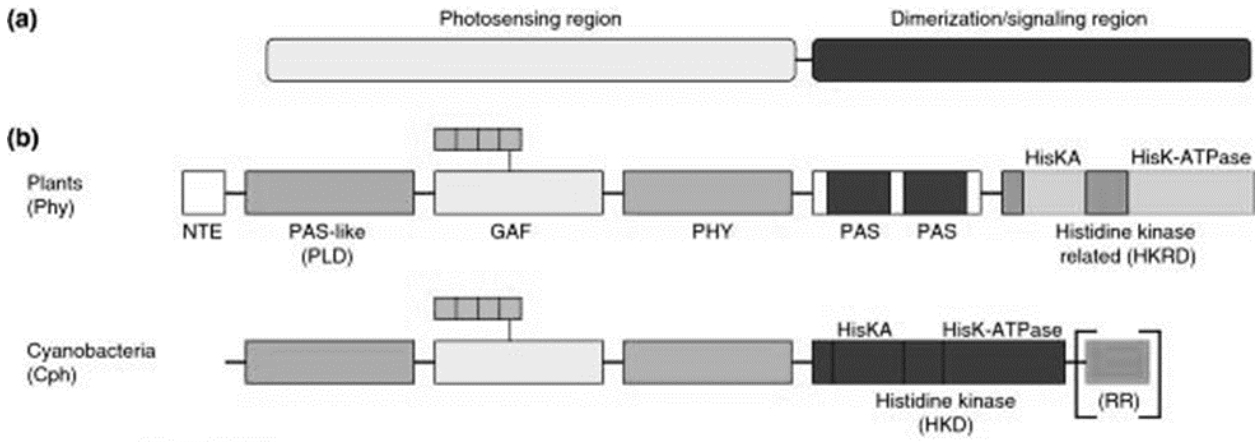

Figure 4.1 Phytochrome family of photoreceptors
(a) The photoreceptors have two functional portions: a photosensing region and a dimerization/sensing region. They exist as homodimers through associations between the dimerization/signaling regions.
(b) Each of the two regions in (a) are made of distinct domains. A bilin chromophore is covalently bound to a GAF domain, and it is the chromophore that detects light quality and quantity; light absorption triggers a conformational change in the receptor structure that is transmitted to other signal transduction pathways via the dimerization/signaling region. Note that the photosensing regions are composed of three primary domains: PAS, GAF, and PHY, while the signaling region is constructed of histidine kinase related domains (HisKA and HisK-ATPase). This is a highly simplified comparative model; for example, cyanobacterial phytochromes may contain multiple GAF domains with or without attached chromophores.

Polymerase Chain Reaction

The most direct means of cloning a protein coding gene is to make a DNA copy of the gene by the polymerase chain reaction method (PCR) and then incorporate this copied DNA segment into a plasmid. Arguably, one of the most significant breakthroughs in biochemical and molecular research efforts was the advent of the polymerase chain reaction (PCR) methodology. PCR is quick and reliable for which the sequences of target genes and genomes are known (underscoring the power of the genome projects determining the DNA sequences of many genomes from many organisms). A prokaryotic gene can be copied directly from the genome

since there are no introns. Introns are non-protein coding segments of DNA interspersed in the gene along with exons that are the protein coding segments. Eukaryotic genes typically contain introns. Hence, PCR reactions to copy a contiguous protein coding segment must be performed with mRNA for most Eukaryotic genes, which had introns removed by RNA processing mechanisms. Isolated mRNA is copied into DNA by reverse transcriptase (RT), then the DNA is copied by the PCR method, in a technique noted as RT-PCR. There are several methodologies using PCR, and all successful PCR methods require one critical aspect: **PRIMER DESIGN** (discussed below).

The PCR Reaction

DNA is always synthesized in the 5' to 3' direction by DNA Polymerases. A DNA polymerase adds single nucleotide 5'-triphosphates at the free 3'-OH group of an existing DNA segment, thus extending the length of the DNA in the 5' to 3' direction. The nucleotides that are added are complementary to the DNA strand being copied; so each strand of the double stranded DNA molecule serves as a template for synthesizing (i.e. making a copy) of the opposite strand. The double stranded DNA molecule is melted apart under high temperature in a test tube and each of the two parental DNA strands serves as templates for DNA Polymerase to copy. DNA Polymerases from thermophilic organisms are used for PCR since these enzymes remain in their native tertiary structures and enzymatic activity under temperatures needed to melt double stranded DNA. One such early and frequently used polymerase is from the organism Thermus aquaticus, called Taq Polymerase; there are many other thermos-stable DNA polymerases now in use. The PCR reaction is shown in Figure 4.2.

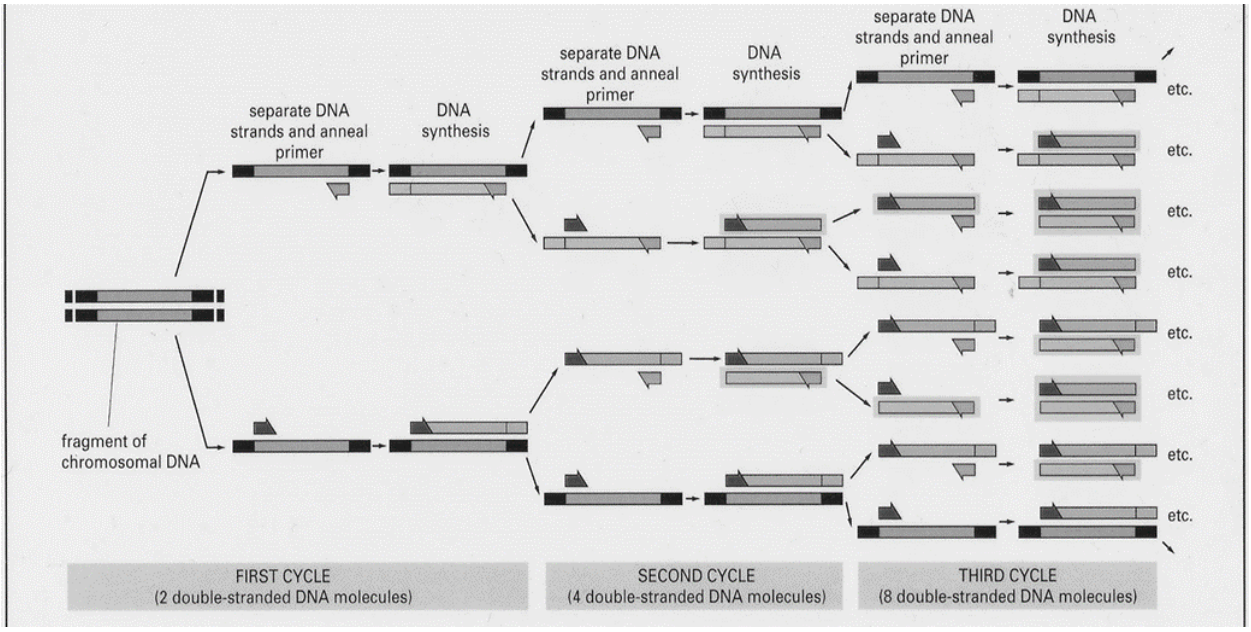

Figure 4.2 The PCR Cycle (from: Molecular Biology of the Cell, 4th ed, Alberts *et al*, Garland Science)

DNA Polymerases require a pre-existing segment of DNA with a free 3'-OH group from which to extend the newly synthesized strand in the 5' to 3' direction, which are provided by short DNAs

or RNAs basepaired with the strand to be copied, called “primers”. Synthetic primers are used for PCR reactions and they are short segments of DNA whose sequences are designed to be complementary to the segment of DNA to be copied. Two primers are constructed for PCR reactions, one binds at one end of the DNA segment to be copied and a second binds at the other end of the same segment. When the primers bind to the template DNA strands, the Taq DNA Polymerase will synthesize a new copy of each of the two template strands. The PCR reaction contains the template DNA that has a segment of DNA to be copied (such as a gene), primers to bind to specific sequences of the template strands, nucleotide 5’- triphosphates, and magnesium ions (substrates for DNA Polymerases are the magnesium salts of the nucleotide triphosphates). The PCR reaction has three steps: (1) melting the template DNA strands at > 95^{o}C, (2) binding the primers at 37^{o}C to 60^{o}C (called annealing), and (3) DNA synthesis at 72^{o}C (the temperature optimum for Taq DNA Polymerase). These steps are repeated many times with each successive cycle doubling the number of DNA molecules being copied. Typically, between 25 and 30 cycles are performed by a thermo-cycler machine that very quickly raises and lowers the temperature required for each step. Hence, between 225 (33,554,432) and 230 (1,073,741,824) copies of a segment of DNA is made! Also, notice that with each cycle, the length of the template DNAs shorten to just that of the segment being copied (the original parent strands and the early extended strands that were copied still remain, but they are unobservable.

Agarose Gel Analysis

The identity of the resulting DNA segment from a PCR reaction needs to be confirmed to prevent the perpetuation of an artifact (i.e. copying the wrong segment of DNA). The analysis of the DNA segment is initially performed by assessing its size in base-pairs (bp) by agarose gel electrophoresis. Electrophoresis is the separation of molecules in solution passing through a porous matrix under an electric force. Agarose is a polysaccharide of galactose units, and galactose derivatives cross- linked by hydrogen bonding create a gel when hydrated. An agarose gel matrix of 0.7% to 1.2% (w/v) separates DNA fragments of about 20000 to 200 bp. At physiological pH, nucleic acids are polyanions, due to the ionized phosphate groups in their phosphodiester backbone. As such, nucleic acids migrate towards the anode in an electric field. Also, each DNA strand has the same charge-to-mass ratio due to the regular intervals of phosphate groups along the DNA molecule; hence, the relative electrophoretic mobility of linear DNA is determined by its length alone. A full discussion on electrophoresis is in Experiment 6 of the manual The DNA in an agarose gel is visualized as bands by staining with any number of dyes that intercalates between the bases of DNA molecules and fluoresces when excited with ultraviolet light: some common dyes are ethidium bromide, SYBR Safe, GelRed and RedView.

The length of a linear piece of DNA (e.g. a PCR reaction product) is estimated by its mobility compared to the mobility of a set of linear DNAs of known sizes (aka “markers”). A typical standard curve created from the known size standards (Figure 4.3) shows mobility of linearized DNA versus size; the mobility of DNA is non-linear and the length (size) is typically referred to as base-pairs (bp) or kilobase-pairs (kbp), hence the units are log bp or log kbp, to achieve a linear standard curve. Relative mobility refers to the ratio of the mobility of a band to the mobility of the dye front (a small molecular weight anionic dye is included to follow the progress of electrophoresis, the mobility of the dye is much faster than the DNA). In addition to size, stained

gels can be used to estimate the amount of DNA in a given band by visually comparing its fluorescence intensity to that of a known amount of marker DNA, often to one or more of the size standards.

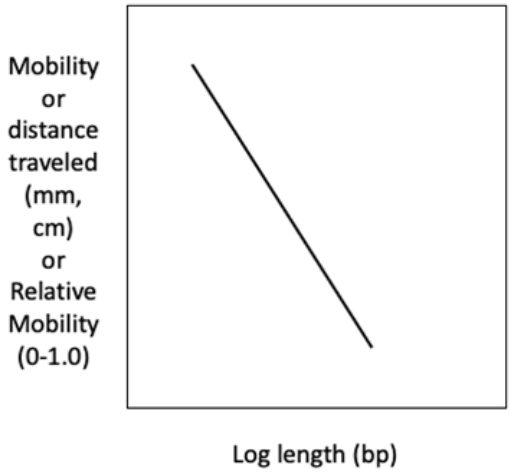

Figure 4.3 A standard curve for the electrophoresis of DNA

Restriction Enzyme Mapping

The size alone of a piece of DNA does not verify that it is the correct sequence. The product could be a random amplification that happens to be similar in size. A quick analytical verification of sequence performed in the lab is restriction enzyme mapping. Restriction enzymes catalyze the cleavage of double stranded DNA molecules generating smaller DNA fragments. These enzymes are found in a wide variety of organisms and hundreds of purified restriction enzymes are commercially available. Restriction enzyme names typically consist of an abbreviation for the source organism (e.g. Eco for E. coli) followed by a Roman numeral or number to indicate the enzyme (Table 4.1).

Restriction enzymes recognize and cleave DNA at specific sequences that are typically palindromes. The importance of this enzymatic specificity is that the digestion of a DNA molecule produces a set of discrete, specific sized fragments. The pattern of restriction enzyme sites in a DNA segment is a fingerprint of that molecule, and is referred to as a restriction map. Hence, if a PCR reaction has amplified the correct gene, then digesting the PCR product by an enzyme whose recognition sequence is present will generate smaller sized fragments of known sizes. Figure 4.4 shows the result of a restriction enzyme mapping of a hypothetical PCR generated DNA fragment. If a 1000 bp fragment is expected, the agarose gel confirms a 1000 bp band in Lane 1 is produced. The sizes of the smaller fragments from restriction enzyme digests with PstI and SstI also confirm the identity of the PCR product.

Name	Recognition Site	Organism
BamHI	G^GATCC	*Bacillus amyloliquefaciens* II
EcoRI	G^AATTC	*Escherichia coli* RY13
HindIII	A^AGCTT	*Haemophilus influenzae*Rc
NcoI	C^CATGG	*Nocardia corallina* (ATCC 19070)
NdeI	CA^TATG	*Neisseria denitrificans* (NRCC 31009)
PstI	CTGCA^G	*Providenica stuarti* 164
SalI	G^TCGAC	*Streptomyces albus* G
SmaI	CCC^GGG	*Serratia marcescens*
SstI	GAGCT^C	*Streptomyces stanford*
XbaI	T^CTAGA	*Xanthomonas campestris pv badrii*

Table 4.1 A short list of restriction endonucleases. ^ indicates specific cleavage sites

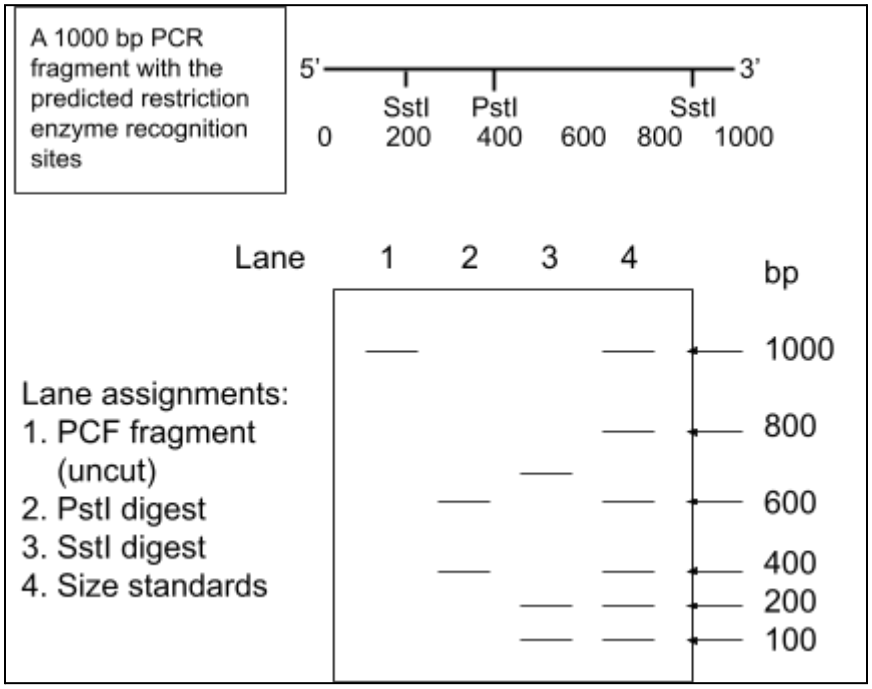

Figure 4.4 Restriction enzyme mapping of a DNA fragment analyzed by agarose gel electrophoresis.

Another important consequence of restriction enzyme digestion is the formation of cohesive ends. DNA fragments from different sources have the same sticky ends (cohesive ends) if they have been generated by digestion with the same restriction enzyme, e.g. Figure 4.5 shows the result of digestion with EcoRI. Thus, two different DNA fragments can be annealed together by complementary base pairing of the cohesive ends under appropriate conditions of temperature and ionic strength. DNA ligase catalyzes the formation of a phosphodiester bond between the 3'-hydroxyl and 5'-phosphoryl groups of adjacent nucleotides held in position by hydrogen bonding of the complementary base pairing. The reaction is called ligation. Blunt end fragments, such as those generated by SmaI, are much more difficult to ligate since they do not have complimentary cohesive ends; however, blunt ends do allow the ligation of any two fragments of DNA generated by different blunt-end forming restriction enzymes. This is recombinant DNA: the recombining of DNA segments not naturally found together.

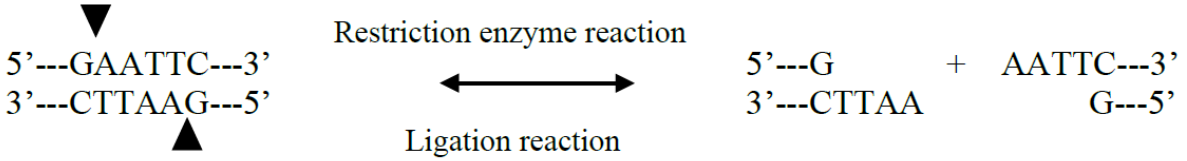

Figure 4.5 Cohesive ends (sticky ends) produced by EcoRI cleavage of double stranded DNA (top). Combining DNA fragments with EcoRI ends by Ligation (bottom), reformation of the EcoRI recognition site.

Nostoc GAF-4 domain

Nostoc punctiforme is a cyanobacterium, which is a photosynthetic organism, hence sensing changes in light intensity and wavelength, and making appropriate metabolic and physiologic responses, is critical. In cyanobacteria, there are many light sensing molecules that collectively span the visible spectrum called cyanobacteriochromes. Cyanobacteriochromes sense light by a linear tetrapyrrole that is attached onto a cysteine belonging to a protein domain known as GAF. The GAF domain units are universally found in many proteins in all kingdoms, which contain the structural motifs for cyclic GMP regulated nucleotide phosphodiesterase, Adenylyl cyclase, and the bacterial transcription factor FhlA. The precise mechanistic details of the GAF sensing and signaling events in the cyanobacteriochromes remain unknown. One Nostoc cyanobacteriochrome sensory protein, from the NpR6012 gene, contains four GAF domains in tandem. The fourth domain, Nostoc GAF-4 domain, is cloned in this experiment (Figures 4.6 and 4.7). The GAF-1 domain of NpR6012 does not contain an attached bilin.

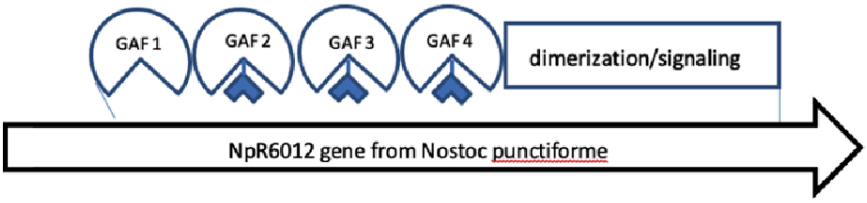

Figure 4.6 The NpR6012 gene from Nostoc punctiforme. The fourth GAF domain is designated NpR6012g4; phycocyanobilin is attached via a Cysteine.

The GAF-4 domain peptide sequence is shown in Figure 4.7. The peptide is 176 amino acids. The designation for the beginning and end of the domain is based on the tertiary structure (not shown). The attached linear tetrapyrrole is a bilin called, phycocyanobilin, which appears blue in color to the eye, and has a molar absorption coefficient of 132,000 M-1 cm-1 at 652 nm (Figure 4.8).

```
                                                    EKAVTK ISNRIRQSSD 600
VEEIFKTTTQ EVRQLLRCDR VAVYRFNPNW TGEFVAESVA HTWVKLVGPD IKTVWEDTHL
660 QETQGGRYAQ GENFVVNDIY QVGHSPCHIE ILEQFEVKAY VIVPVFAGEQ
LWGLLAAYQN 720 SGTRDWDESE VTLLARIGNQ LGLALQQTEY LQQVQGQSAK
760
```

Figure 4.7 The sequence for amino acids 585 – 760 of the Nostoc GAF-4 domain (NpR6012g4). The underlined C-687 is the bilin attachment site (UniProtB accession number: B2IU14)

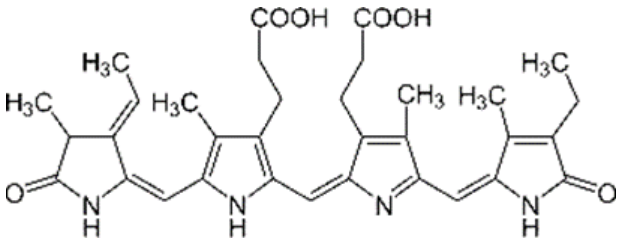

Figure 4.8 The structure of phycocyanobilin

In order to dissect the biochemical, molecular, and cellular function(s) of GAF domains, these domains are cloned and expressed apart from their native biological environments. The molecular structure of a cloned GAF domain can be altered, for example, by site-directed mutagenesis. The resulting recombinant mutated GAF gene then can be returned to its native organism (Nostoc in this example) to examine the effect such mutations may have on the sensory and signaling pathways.

Directional Cloning

The isolation of a protein coding gene is accomplished primarily by PCR, followed by the directional insertion of the PCR product into a plasmid. Plasmids are extra-chromosomal circular double stranded DNA molecules found in numerous prokaryotes and yeasts; and is one type of "vector" DNA molecule used for cloning DNA sequences of interest. Cloning vectors are bio-engineered from plasmids or viruses that are autonomously replicated in a host cell, thus also replicating the inserted DNA fragment. They range in size from two to several hundred kilobase-pairs (kbp). Plasmids, and other cloning vectors, have been extensively engineered to where many plasmids are available and a specific plasmid is chosen for a specific purpose. Expression plasmids are chosen to express a (functional) protein in host cells.

General Plasmid Functions

In this experiment, the vector pBAD/Intein-CBD (Figure 4.9) was chosen to express the cloned Nostoc GAF-4 domain. Plasmids have common features that allow them to persist in host cells and for experimental use. The plasmid pBAD/Intein-CBD is an example of a bio-engineered plasmid, meaning the functional portions of the plasmid are derived from multiple sources and pieced together.

Origin of replication

It is essential for a vector to have a sequence that permits its autonomous replication and retention in host cells. For bacterial hosts, the ori sequence fills this function: shown as the pBR322 ori in Figure 4.9.

Antibiotic resistance gene

The presence of an antibiotic resistance permits only those host cells that have taken up a plasmid to grow in media containing an antibiotic. The ampicillin resistance gene encodes -lactamase, an enzyme conferring resistance to the antibiotic, ampicillin.

Multiple cloning site

A multiple cloning site (MCS) is typically included, which contains several restriction enzyme recognition sites not found anywhere else within the plasmid sequence (i.e. unique single sites). The collection of restriction enzyme sites in the MCS gives several possibilities for the insertion of foreign DNA segments into the plasmid. The pBAD/Intein-CBD MCS contains an NcoI site and the 5' end of the MCS, and a SmaI site at the 3' end of the MCS; constructed from the pBAD MCS and the pTYB2 MCS for the specific purpose of cloning the Nostoc GAF protein.

Expression plasmids

Expression plasmids must contain four essential DNA elements to transcribe and translate an inserted protein coding gene. These elements might be included with an inserted gene or

provided by the plasmid for convenience. These elements are located in or near the MCS to reconstruct a functional recombinant gene.

1. **Promoter**. A promoter is required for RNA Polymerase to bind. This is the pBAD promoter 5' to the MCS on pBAD/Intein-CBD, so the inserted gene is under its control.

2. **A ribosome binding (RBS)**. A RBS site located downstream of the promoter and directly upstream (5-15 bp) of the initiator methionine codon allowing the translation machinery to bind.

3. An initiator **methionine codon**. All (nearly all) organisms begin translation of mRNAs with a start methionine codon (AUG). Often a plasmid will provide a start methionine followed by a short set of other codons within the MCS if the start methionine of an inserted gene is not present. Importantly, the codons of an inserted gene must be in-frame with the plasmid initiator methionine codon.

4. A **stop codon**. A translation stop codon is included with an inserted gene. Plasmids might provide a stop codon along with a preceding short set of other codons. Again, the codons of an inserted gene must be in-frame with the plasmid provided stop codon if used. The pBAD/Intein-CBD provides a stop codon at the end of the Intein-CBD peptide sequence (the Intein-CBD peptide sequence begins at the Cysteine).

   Plasmids are often engineered to contain other features included for experimental convenience, later protein isolation, controlling expression, etc. The pBAD/Intein-CBD has two such additional elements.

5. **An affinity or epitope tag**. An epitope tag is an amino acid sequence recognized by antibodies, high affinity binding proteins, small molecules or molecular complexes that can be used to detect (epitope tag) and/or to purify (affinity tag) the translated cloned protein. This is important for non-enzyme proteins or where enzymatic activity may be compromised or poor, or where the production level is low. The pBAD/Intein-CBD contains the Intein-CBD sequence adjacent to the MCS. A cloned gene whose protein is translated in-frame with Intein-CBD sequence will create a fusion protein with the Intein-CBD sequence located the C-terminal end of the translated protein. The actual affinity tag is the CBD portion of the Intein-CBD peptide; CBD means Chitin Binding Domain. The fusion protein is detected (ie reports its existence) by its direct physical connection to the CBD, which is used to affinity purify the recombinant fusion protein. Intein is a self-cleaving peptide under high concentrations of a thiol reducing agent such as dithiothreitol; the Intein-CBD fusion is removed from the expressed cloned protein fusion leaving a purified protein (see Experiment 5).

6. **Transcriptional regulatory protein**. Genes cloned into the MCS of pBAD/Intein-CBD are under control of the pBAD promoter. Transcription is repressed in the presence of the araC regulator protein, encoded by the araC gene on pBAD/Intein-CBD itself.

Transcription is induced in the presence of arabinose, which binds to the araC regulator protein, altering the promoter region allowing RNA polymerase to bind and begin transcription. In brief, arabinose induces protein expression. Such a system prevents the over-expression of a cloned protein that may be harmful to the host cells until expression is induced for an experiment.

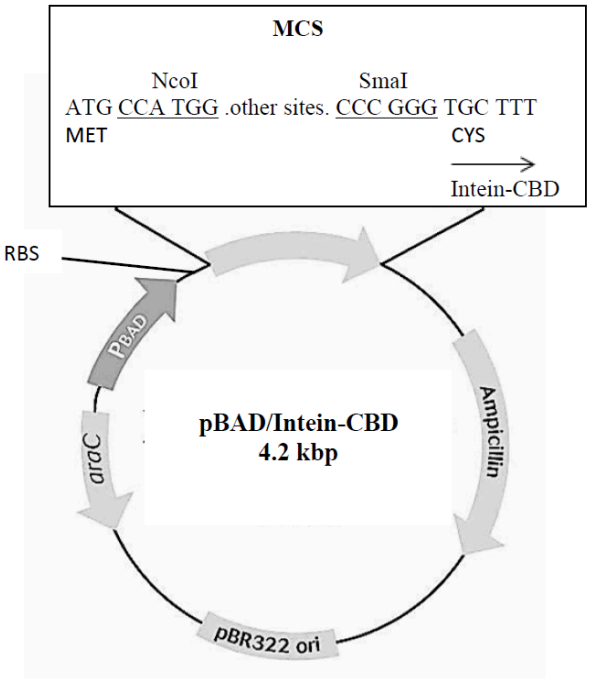

Figure 4.9 The pBAD/Intein-CBD plasmid (pBAD™ is from Invitrogen; the SmaI site and Intein-CBD sequence is derived from pTYB2™ from NEB)

Primer Design

A recombinant protein coding gene must be inserted into the MCS of a plasmid properly so that the elements of an expressed gene are in the correct orientation:

-5'-promoter > RBS > initiator methionine codon > protein coding region > translation stop codon-3'-

Inserting a protein coding gene in the correct orientation is known as directional cloning. Directional cloning is accomplished by adding different restriction enzyme recognition sequences on the 5'-ends of the PCR primers. The choice of recognition sites depends on: (i) the availability of the same sites in a plasmid MCS, and (ii) the chosen sites are not present within the gene to be cloned to avoid the destruction of the inserted gene when digested with the restriction enzymes to generate compatible sticky ends. The sequences of the primers used in this experiment are shown in Figure 4.10. These sites are chosen in conjunction with the engineered NcoI and SmaI sites in the pBAD/Intein-CBD MCS, and are not present in the Nostoc GAF-4 domain gene sequence.

The Forward primer (sense primer)

Since the GAF-4 domain peptide sequence is internal to the larger cyanobacteriochrome protein, it lacks a start Methionine. The NcoI site allows the inclusion of a start Methionine in-frame to the 5'-end of the inserted Nostoc GAF-4 domain sequence. This adds three amino acids to the N-terminus of the Nostoc GAF-4 domain: Met (ATG), Pro (CCA), and Trp (TGG) (Figure 4.10)

```
GAGAAAGCTGTCACCAAGATCAGTAACCGCATCCGGCAATCTTCAGATGTAGAAGAAATCTTCAAA    66
ACAACCACTCAAGAAGTACGACAATTACTGCGATGCGATCGCGTCGCAGTCTATCGCTTCAACCCT   132
AATTGGACTGGTGAATTTGTTGCAGAATCAGTAGCTCACACTTGGGTAAAACTGGTAGGCCCCGAT   198
ATCAAGACTGTCTGGGAAGATACCCACTTACAAGAAACTCAAGGAGGTCGATATGCCCAAGGGGAG   264
AACTTTGTTGTAAATGACATTTATCAGGTAGGTCATTCTCCTTGTCACATTGAAATTTTAGAGCAA   330
TTTGAAGTCAAAGCTTATGTAATTGTTCCTGTATTTGCCGGGGAACAATTGTGGGGATTGCTAGCA   396
GCTTATCAAAATTCTGGAACTCGTGATTGGGATGAATCGGAAGTCACCTTGTTAGCACGCATTGGC   462
AACCAGTTAGGGCTAGCATTACAACAGACTGAATATTTGCAGCAAGTACAAGGGCAGTCAGCCAAA   528
```

Figure 4.10 The nucleotide sequence for the Nostoc GAF-4 domain (NpR6012g4)
Forward primer: 5'-CCCCATGGGAGAAAGCTGTCACCAAG (NcoI)
Reverse Primer: 5'-CCCCCGGGTTTGGCTGACTGCCCTTGT (SmaI)
TGT (307-309) is the Cysteine codon that is the bilin attachment site.

The Reverse primer (anti-sense primer)

The SmaI site allows the in-frame C-terminal addition of the Intein-CBD peptide to the 3'-end of the of the inserted Nostoc GAF-4 domain sequence. This adds Pro (CCC) and Gly (GGG) to the C-terminus of the Nostoc GAF-4 domain (Figure 4.11).

For both primers, the 5' addition of two cytosine nucleotides improves ability of the restriction enzymes to bind to their recognition sites, and each has 18 nucleotides complementary to the GAF-4 sequence.

```
        NcoI                                                    SmaI
..ATGCCATGGGAGAAAGCTGTCACCAAG...  ...ACAAGGGCAGTCAGCCAAACCCGGGTGC...
    GGTACCCTCTTTCGACAGTGGTTC....  ...TGTTCCCGTCAGTCGGTTTGGGCCC
  M  P  W  E  K  A  V  T  K ..    ..  Q  G  Q  S  A  K  P  G
```

Figure 4.11 The insertion of the GAF-4 domain PCR product into the NcoI and SmaI sites of the pBAD/Intein-CBD. The nucleotide and amino acid sequences in red are plasmid derived, whereas the sequences in black are GAF-4 derived.

Ligation, Transformation and Selection

Ligation

After the PCR amplification of the Nostoc GAF-4 domain gene, the resulting DNA fragment is inserted into the pBAD/Intein-CBD plasmid where it can be stored in host E. coli cells and used for later expression. The PCR amplified Nostoc GAF-4 domain gene, and the pBAD/Intein-CBD plasmid, are both digested with NcoI (cohesive ends) and SmaI (blunt ends). These two DNAs are joined together by the binding of the NcoI cohesive ends; the enzyme ligase covalently joins the SmaI blunt ends and the NcoI cohesive ends forming a recombinant plasmid: pBAD/GAF-Intein-CBD. Figure 4.12 illustrates the construction of the recombinant plasmid.

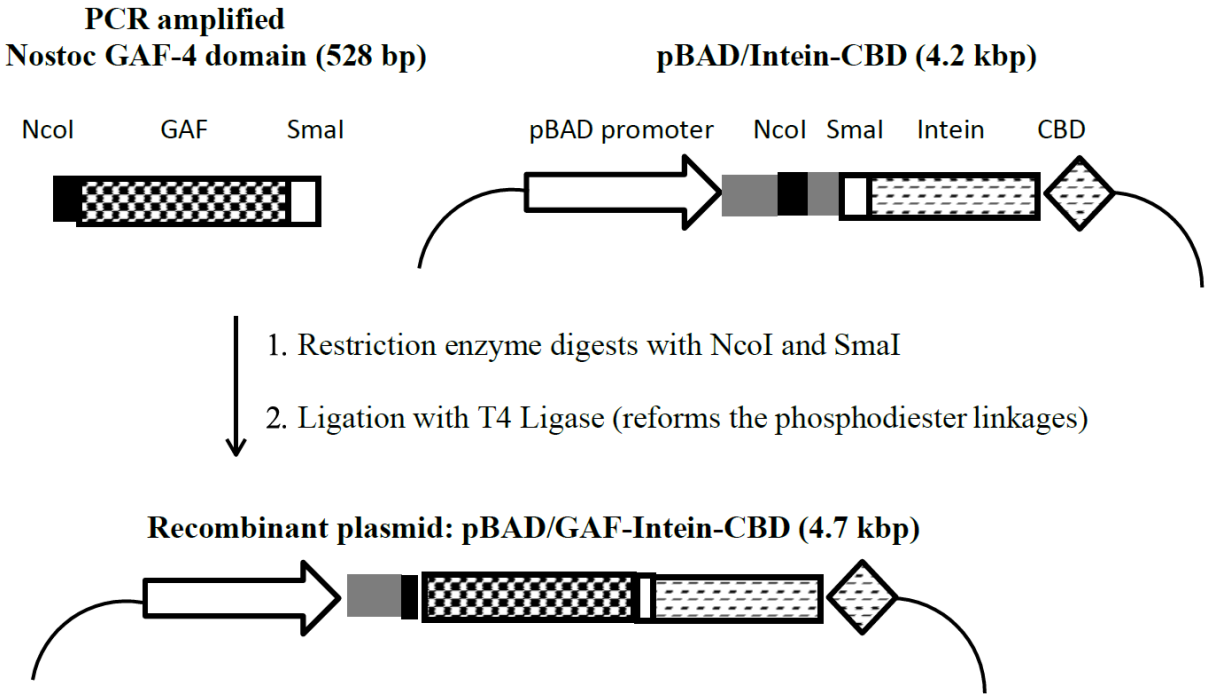

Figure 4.12 The construction of the recombinant plasmid, pBAD/GAF-Intein-CBD.

Transformation and Selection

The transfer of plasmids into E. coli host cells is called transformation. The host E. coli cells are transformed by either the CaCl2/heat-shock method or the electroporation method. The $CaCl_2$/heat shock method requires placing approximately 50 femtomoles of plasmid into a dense solution of E. coli prepared in a high concentration of $CaCl_2$ (known as 'competent' cells); this mix is incubated at 42°C for 30 seconds to 2 minutes and the plasmid is taken-up by the E. coli. The transformed E. coli cells are typically selected on an agar plate with an antibiotic; only those E. coli cells containing a plasmid with the antibiotic resistance gene (transformants) will survive. The electroporation method again requires placing approximately 50 femtomoles of plasmid into a dense solution of E. coli, but prepared in 10% glycerol (known as 'electro-competent' cells); the mix is place in a electroporation cuvette and the plasmid is transferred into the E. coli with a burst of electric current (in an instrument called an electroporator). Again, the transformed E.

coli cells are selected on agar plates containing the appropriate antibiotic. The electroporation method is superior but requires the purchase of the electroporator and electroporation cuvettes. (Figure 4.13)

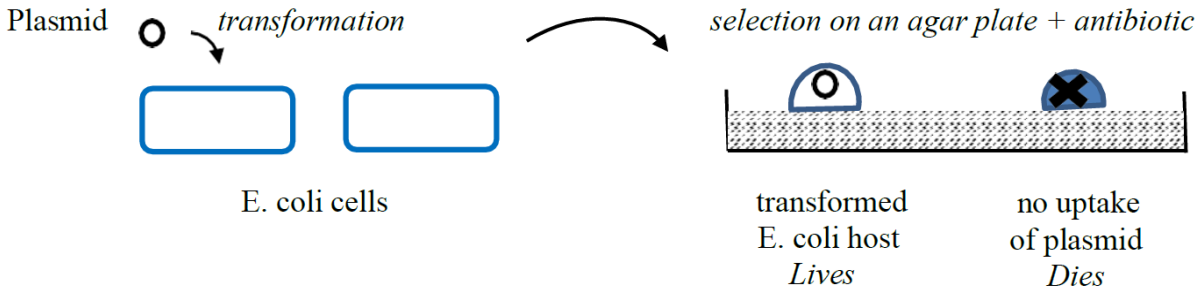

Figure 4.13 Transformation and selection of host E.coli cells

PCR and Gibson Cloning

The cloning method discussed above has been a powerful and widely used method for cloning. However, as noted, the cloning of the GAF-4 gene segment introduced five additional amino acids to the expressed domain (a Met codon, and four amino acids, two each from the use of the restriction sites for directional cloning). Amino acids introduced in this manner may or may not impact the tertiary structure or the function of the protein; empirical experimentation will provide such data. Of course, in the case of the GAF-4 domain, the Met codon must be introduced for translation to begin. More recently, a new cloning method, Gibson Cloning (aka Gibson Assembly), allows for the insertion of a DNA segment into a vector without the need of restriction sites.

Gibson cloning avoids the use of restriction sites for inserting a PCR generated segment of DNA into a vector by (i) adding approximately 15 nucleotides onto the 5' end of the primers that are complementary to the vector sequence, and (ii) introducing two enzymes in the ligation step: a 5' exonuclease and DNA polymerase. The longer length of complimentary nucleotides on the primers form a much more stable base- pairing with the vector than the 2 to 5 base-pairing of compatible restriction sites. To allow the primers to bind to the vector, the circular vector is opened (ie made linear), usually by a single restriction digest at an unique site (eg for currently available and useful expression plasmids, simply choose a site in the MCS). The PCR copied DNA segment and opened vector are mixed together with the Gibson reaction mixture (ie the reaction mixture contains the 5' exonuclease, DNA polymerase, and ligase). The 5' exonuclease chews back DNA only from the exposed 5' ends on both the vector and the DNA segment to be inserted. See Figure 4.14 below for a comparison between cloning via the use of restriction sites, panel A, and the Gibson modification, panel B. The exonuclease results in single stranded stretches of DNA that allow the complementary sequences of the vector and DNA segment to base-pair (ie anneal). The gaps remaining after the exonuclease activity and annealing of the complementary sequences are filled in by DNA polymerase (recall, all DNA polymerases extend DNA from free 3'-OH ends). Lastly, the ligase restores the phosphodiester bonds. The final vector with insert is used for transformation of a host; growth and selection of transformants are as described above.

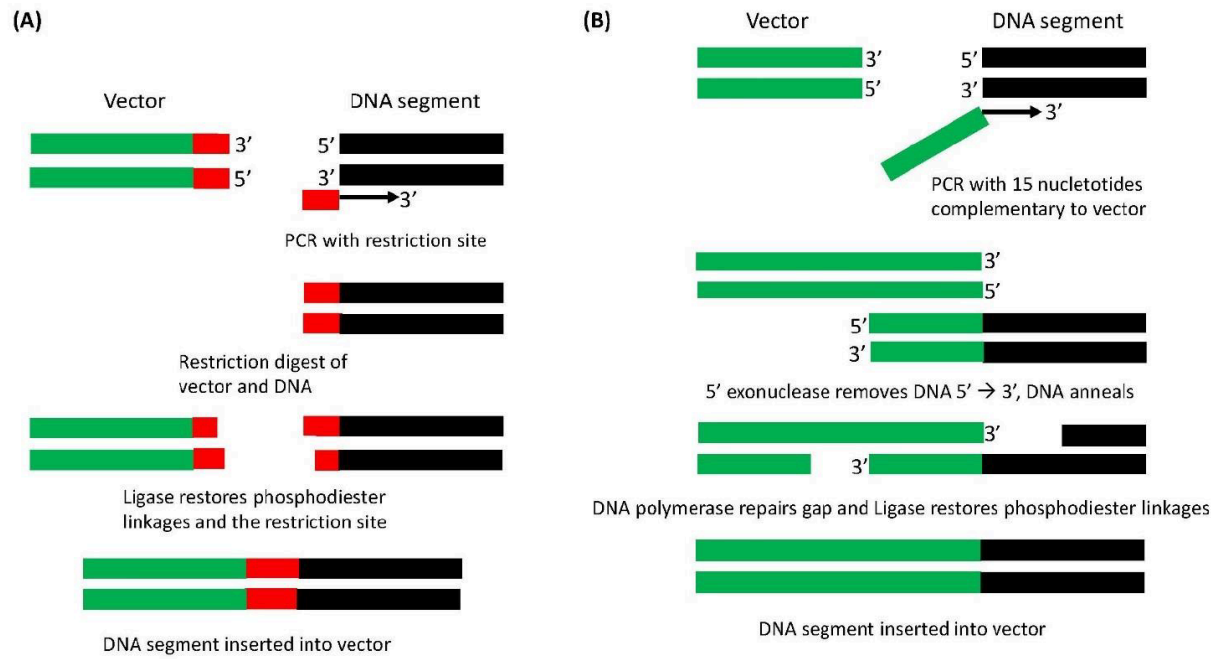

Figure 4.14 Cloning PCR generated DNA segments into plasmids. (a) Steps requiring restriction sites. (b) Steps in Gibson cloning avoiding restriction sites.

The primers used in Gibson cloning do not introduce added amino acids to cloned and expressed protein sequences. Whether or not Gibson cloning is used, the re-creation of a functional gene for protein expression via primer design is very similar. As an example, if GAF-4 were cloned using the Gibson cloning, how would the primer design change? And, how would this affect the protein sequence of the expressed GAF-4 protein?

The primers from Figure 4.10 above
Forward primer: 5'-CCCCATGG***GAGAAAGCTGTCACCAAG*** (NcoI)
Reverse primer: 5'-CCCCCGGG***TTTGGCTGACTGCCCTTGT*** (SmaI)

The primers that would be used for Gibson cloning
Forward primer: 5'- 12 nucleotides ATG***GAGAAAGCTGTCACCAAG***
Reverse Primer: 5'- 12 nucleotides GCA***TTTGGCTGACTGCCCTTGT***

The 15 nucleotides of the forward primer ending with ATG are from the plasmid; the ATG must be introduced for translation to begin. The 15 nucleotides of the reverse primer ends in GCA, which is complementary to the Cys codon, TGC, required to synthesize the fusion protein with the Intein/CBD.

Protein sequence from Figure 4.11 above

NcoI SmaI

..*ATG*CCATGGGAGAAAGCTGTCACCAAG… ...ACAAGGGCAGTCAGCCAAACCC*GGGTGC*...

*GGTACC*CTCTTTCGACAGTGGTTC… …TGTTCCCGTCAGTCGGTTTGGGCCC

M P W E K A V T K .. .. Q G Q S A K P G C

Protein sequence that would result from using Gibson cloning

..*ATG*GAGAAAGCTGTCACCAAG… …ACAAGGGCAGTCAGCCAAA*TGC*...

CTCTTTCGACAGTGGTTC… …TGTTCCCGTCAGTCGGTTT

M E K A V T K .. .. Q G Q S A K C

In the GAF-4 protein expression resulting from Gibson cloning the proline and tryptophan at the N-terminal end are eliminated, and the proline and glycine at the C-terminal end are eliminated.

Concluding remarks regarding primers

The databases of submitted genome and protein sequences, and PCR technology, allowed for the experimentation on many previously intractable proteins such as transient signal transduction proteins. Primers for PCR are chemically synthesized, which are inexpensive and widely available. The researcher submits the sequence information to a company and the primers are delivered within a couple of days. The most critical step in PCR methodologies is primer design, especially for cloning functional genes for protein expression. Just one incorrect nucleotide can have dire consequences on the structure of the expressed protein: one added or skipped nucleotide will shift the reading frame resulting in a completely different primary sequence from the point of the error onward; an incorrect notation (eg C instead of a G) might result in a point mutation. However, many research labs are by-passing the de novo amplification of genes by having entire genes commercially synthesized from sequence information available from the online accessible data banks. The synthetic genes are then inserted into plasmids or other vectors as described above. Alternately, entire plasmids are synthesized containing the specific functions and gene to be expressed: referred to as in silico synthesis.

References

1. http://www.neb.com/nebecomm/products/productn6702.asp (pTYB2 plasmid)
2. http://tools.invitrogen.com/content/sfs/manuals/pbad_man.pdf (pBAD plasmid)
3. Ho Y-S et al (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. The EMBO Journal 19: 5288 – 5299.
4. Gambetta GA, Lagarias JC (2001) Genetic engineering of phytochrome biosynthesis in bacteria. PNAS 98 (19), 10566-10571.
5. Ikeuchi M, Ishizuka T (2008) Photochem Photobiol Sci 7: 1159-1167.
6. Rockwell NC et al (2008) Biochemistry 47(27):7304-7316.
7. Narikawa R et al (2008) J Mol Biol 38(8): 844-855.
8. Hirose Y, Narikawa R, Katayama M (2010) PNAS 107(19): 8854-8859.
9. Rockwell NC, Lagarias JC (2010) Eur J ChemPhysChem 11(6): 1172-1180.
10. Kehoe D (2010) PNAS 107(20): 9029-9030.
11. Rockwell NC et al (2011) PNAS 108(29): 11854-11859.

12. Ishizuka T et al (2011) Biochemistry 50(6): 953-961.
13. Rockwell NC, Martin SS, Lagarias JC (2012) Biochemistry 51: 3576-3585.
14. Rockwell NC, Martin SS, Lagarias JC (2012) Biochemistry 51: 9667-9677.
15. "PCR optimization: primer design" (1997) http://www.qiagen.com/literature/qiagennews/0597/975pcro.pdf
16. Jacob F (1977) "Evolution and Tinkering". Science 196(4285): 1161-1166.
17. Campbell I, Downing A (1994) "Building Protein Structure and Function from Module Units" Trends in Biotechnol 12(5): 168-172.
18. Bork P, Doolittle R (1992) "Proposed Acquisition of Animal Protein Domain by Bacteria" PNAS 89(19): 8990-8994.
19. Wikipedia: Protein domains
20. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York (1989).
21. Watson JD, Gilman M, Witkowski J, Zoller M: Recombinant DNA. Second Edition. Scientific American Books, W.H. Freeman and Co., New York (1992), 260 pp.
22. Mullis, KB, and FA Faloona (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Meth. Enzymol. 155: 335-50.
23. Saiki, RK, DH Gelfand, S. Stoffel, SJ Scharf, R Higuchi, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239(4839): 487-91.
24. Tirabassi, R. (2018) Foundations of Molecular Cloning – Past, Present and Future. New England Biolabs. https://www.neb.com/tools-and-resources/feature-articles/foundations-of-molecular-cloning-past-present-and-future

Materials

agarose	6X DNA loading dye	1.0 X TAE	10,000X SybrSafe (Invitrogen)
DNA Mass	Ruler (100 bp to 1000 bp)	0.1 M NaCl	

Fermentes GeneJet™ PCR Purification Kit. Columns and accompanying solutions:
Binding-Buffer, Wash-Buffer, and 10 mM Tris pH 8.5 (elution buffer).

PCR ingredients:
2X PCR Master Mix (2X PCR buffer, 400 uM dNTPs, 0.05 units/ µl)*
Forward GAF primer, 10 pmol/µl
Reverse GAF primer, 10 pmol/µl

Restriction enzyme reaction mix cocktail for one reaction: 15 µl cocktail volume (20 µl final volume)
12.5 µl H2O
2 µl 10X buffer (specified for the enzyme by manufacturer)
0.5 µl restriction enzyme, **HindIII**

Experimental Procedures

Day 1
Part A. PCR amplification of a Nostoc GAF-4 domain sequence

1. The **final** concentrations or quantities of the following components of a single PCR reaction in a 50 µl reaction volume are:

   1X PCR buffer solution (contains final concentration of: 10 mM Tris-HCl pH 9.0, 50 mM KCl, and 1.5 mM $MgCl_2$)
   200 µM each dNTP
   50 pmol Forward GAF Primer
   50 pmol Reverse GAF Primer
   1.25 units Taq DNA Polymerase

   To simplify the reaction, the buffer solution, dNTPs, and Taq DNA polymerase were pre-mixed into a 2X PCR master mix for you.

Pre-lab: Calculate the volumes of each component required for a single PCR reaction (50 ul total volume):

The **stock** concentrations of the PCR reaction components are:
2X PCR Master Mix (2X PCR buffer + 400 uM dNTPs + 2.5 units Taq DNA Polymerase)
10 pmol/µl Forward GAF primer
10 pmol/µl Reverse GAF primer

5 µl target DNA solution (containing the Nostoc GAF-4 sequence of interest)
_____ 2X PCR Master Mix
_____ Forward GAF primer, stock: 10 pmol/µl
_____ Reverse GAF primer, stock: 10 pmol/µl
_____ H_2O

2. Obtain and label the top of the lid of a **0.2 ml thin-walled PCR tube**, not the microcentrifuge tubes in your drawer.

3. Transfer 5 µl of the target DNA solution into the labeled PCR tube (the target DNA is approximately 0.4 - 1.0 ng per reaction).

4. Add the 2X PCR Master Mix, forward and reverse primers, and water into the same PCR tube and mix well.

5. Place the tube in a rack at the front of the lab.

The instructor or TA will start the PCR thermocycler when every group is ready.

Programmed PCR reaction cycles:

Step 1. 95^{O}C for 30 seconds
Step 2. 55^{O}C for 30 seconds
Step 3. 72^{O}C 1 minute
Step 4. Repeat steps 1 – 3, for 24 additional cycles
Step 5. 72^{O}C for 10 minutes
Step 6. RT

6. The PCR reaction will take approximately 1.5 hours.

Part B. GeneJet™ PCR Purification Kit clean-up of the PCR product

The following protocol is designed to remove primers, nucleotides, polymerases, and salts from the PCR reaction mixture, thereby purifying DNA (i.e. PCR product) ranging from 100-100,000 bp using a GeneJet™ PCR Purification Kit (Fermentes).

1. Retrieve the tube with the PCR reaction mixture.

2. Transfer the reaction using a P-200 to a 1.5-ml microcentrifuge tube. Add an equal volume of Binding-Buffer (i.e. 50 µl) and mix well.

3. Place a GeneJet column onto the Vacuum manifold, load the entire PCR sample in Binding-Buffer onto the column and leave the column-cap open.

4. Turn on the vacuum. It may be necessary to press down on the lid of the manifold to achieve a good seal. After the sample has passed through the column, turn off the vacuum.

5. Add 750 µl of Wash-Buffer to the column and restart the vacuum, to wash any remaining contaminants from the bound DNA, and leave the vacuum on for about two minutes to ensure the removal of the Binding-Buffer.

6. Remove the column from the manifold and insert it into a clean 1.5-ml microcentrifuge tube labeled "CLEAN PCR".

7. Add 30 µl of 10 mM Tris-HCl, pH 8.5 (elution buffer), directly onto the center of the column and centrifuge the column in the 1.5 ml tube for 1 min at maximum speed. The DNA (PCR product) is now in the liquid that has eluted from the column (~ 30 µl). Discard the column.

8. Proceed to Part C, and **SAVE** the remaining Clean PCR product. TA will store it in the refrigerator for Day 2.

Part C. Quantifying the PCR product

Nanodrop Spectrophotometer

1. On the Nanodrop startup menu, select “Nucleic Acid”. The first user will initialize the instrument by cleaning the pedestals and loading a water sample. Initialization should be necessary only at the beginning of the lab period.

2. In the upper left hand corner of the screen, note the display: Nucleic Acid. Before measuring a sample, wipe the pedestals of the Nanodrop briskly with a Kimwipe.

**Please treat the Nanodrop gently! DO NOT lift the arm by the flexible fiber optic cable! DO NOT bang the arm down onto the bottom pedestal. Lower the arm gently!

3. Measure the absorbances of the Clean PCR product using 4 ul.
    a. Reminder: **SAVE** the remaining Clean PCR product for day 2!

Table 4.4 Nanodrop measurements

	A260	A280	A260/A280	Conc of DNA ng/ul	Con of DNA calculated using A260
PCR product					

Calculated Concentration of DNA

When DNA is sufficiently isolated from other cellular components, the concentration of DNA is easily determined by spectrophotometry. The Nanodrop readout automatically adjusts the absorbance for a pathlength of 1cm. An absorbance value of 1.0 corresponds to 50 µg/ml of double stranded DNA at 260 nm in a 1 cm path length; so, using the Beer’s Law expression: $a\ (260nm) = 0.020\ (\mu g/ml)^{-1}\ (1\ cm)^{-1}$.

Day 2

Part D. Restriction enzyme digest

1. Add 20 ul water to the PCR product to make ~50 ul sample.

2. Label a 0.5 ml microfuge tube for the HindIII restriction enzyme reaction.

3. Transfer 5 µl of the Clean PCR product to the labeled tube.
(**SAVE** the remaining ~25 µl Clean PCR product).

.

4. Transfer 15 µl of the restriction enzyme reaction cocktail ("master mix") to the labeled tube and mix well. The reaction mix contains the following reagents per reaction:
   12.5 µl H_2O
   2 µl 10X Restriction enzyme buffer solution compatible with HindIII (e.g. NEB Buffer 2) (1X NEB Buffer 2: 10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl2, 1 mM Dithiothreitol, pH 7.9 at 25°C)
   0.5 µl Hind III (final: approx. 20 units)
   The final reaction volume is 20 µl.

5. Incubate at 37°C for 60 minutes.

(While the reaction is incubating, pour the agarose gel, part E - 1 gel is shared by 3 groups)

Part E. Agarose gel analysis of the PCR product and the restriction enzyme digests.

*****3 groups will share **one** agarose gel

1. The agarose gel casting tray is flat-bottomed, walled on two sides, and open at both ends. Place the tray in the apparatus so that the open ends are "closed" by the walls of the apparatus.

2. Agarose gels are made by dissolving agarose in TAE buffer in % weight/volume. For example, 1% gel is made by dissolving 1 g of agarose into 100 ml of TAE buffer. You will be making a 1.5% agarose gel of 80 ml volume. Calculate the amount of agarose needed.

   Pre-lab: Dissolve _____ g of agarose in 80 ml of 1.0 X TAE by heating the mixture in a microwave oven for approximately 1-2 minutes. Watch closely to avoid boiling over the mixture, swirl to fully dissolve the agarose (The gel will be 1.5% agarose that is dense enough to see DNA bands below 1000 bp).

3. Allow to cool to a "comfortable" temperature, then add 8 µl SybrSafe (DNA dye) and swirl to mix evenly throughout the melted agarose. TA will oversee the SybrSafe step.

4. Pour the gel. Insert the thicker side of one comb and let sit until the agarose has solidified.
   a. When finished, clean the flask really well by scrubbing out any leftover gel with a bottle brush, rinse with dH_2O and place back in the green tray, not the dishwash bin, along with the sponge.

5. Once the agarose has solidified, remove tray with the gel and reorient it 90^0 so that the top of the open end of the gel (i.e. the end with the comb/wells) is near the cathode (negative electrode) and the bottom of the gel is near the anode (positive electrode).

6. Add 400 ml 1.0 X TAE buffer (just enough to cover the gel completely, add more if needed). Then remove the comb; the buffer helps the comb slip out easier without tearing the gel.

7. Retrieve the restriction enzyme reaction when finished and add 4 µl 6X loading dye.

8. Transfer 5 µl of the Clean PCR product into a 0.5 ml microfuge tube, add 5 µl dH_2O, and then add 2 µl 6X loading dye.

9. Obtain 5 µl of the DNA mass ruler standard premixed with dye (see Table 4.2 below).

Table 4.2 DNA mass ruler standards

DNA length (bp)	**Amount of DNA loaded on the gel**
1000	100 ng
700	70 ng
500	50 ng
200	20 ng
100	10 ng

10. Load the agarose gel according to Table 4.3 below.

Table 4.3 Sample Loading for the 1.5% Agarose Gel

Lane	**Sample**	
1		
2	Uncut Cleaned PCR (12 ul)	Group 1
3	HindIII digested (24 ul)	
4	Standards (5 ul)	
5		
6	Uncut Cleaned PCR (12 ul)	Group 2
7	HindIII digested (24 ul)	
8	Standards (5 ul)	
9		
10	Uncut Cleaned PCR (12 ul)	Group 3
11	HindIII digested (24 ul)	
12	Standards (5 ul)	

11. Connect the electrodes with the lid of the apparatus and run at 125 V.

12. When the leading orange dye has migrated ~70% the distance of the gel, turn the power supply off.

13. Photograph the gel using the LiCOR imager following the directions provided in the instructions file in Canvas.