Generating, expressing, and purifying a recombinant GST-EGFP fusion protein

Jesus Quiroz

Section # (01E)

Section 1: pET41a(+) Is Used For Protein Expression and its Regulation

Generating a recombinant protein of interest efficiently for extraction, purification, and for eventual research studies is done through the use of genetically modified bacteria and plasmids. Protein expression through plasmid pET41a(+) creates an N-terminal glutathione s-transferase (GST) tag which can be followed by a fusion protein(Fig. 1A). pET41a(+) served as a backbone for an enhanced green fluorescent protein (EGFP) insert through molecular means to achieve a recombinant plasmid and subsequently a fusion protein(Fig.1B). Benchling is used to generate expected sequences of nonfusion and fusion proteins (Table S1).

pET41a(+) contains a multi-cloning site for insertion of the gene of interest through the use of the two restriction enzymes, NotI and NcoI (Fig. 1C). Additionally, two primers flank the multi cloning site and include part of GST for use in further assays, and the plasmid has a selectable marker through the kanamycin resistance gene(Fig. 1C). The plasmid's protein expression depends on the presence of the T7 RNA polymerase which is regulated through the *lac* operator which depends on the availability of glucose and lactose. When lactose is absent, *lacI* inhibits the *lac* operator, and when glucose is absent but lactose is available, transcription of T7 RNA polymerase can occur to express pET41a(+) proteins through a special promoter for the polymerase. BL21(DE3) competent *E. coli* carry a modified *lac* operator that is insensitive to glucose availability which allows the presence of lactose to regulate when plasmid expression is induced. IPTG is an analog of lactose used to induce protein expression. Subsequent studies attempt to verify the identity of an unknown plasmid as recombinant or non-recombinant plasmids, as well as quantifying fusion protein expression levels.

Section 2: Generating recombinant pET41a(+)-EGFP plasmids

Ligation of the eGFP insert into the pET41a(+) backbone's multi-cloning happened under three different restriction digest conditions. Conditions are as follows, no restriction enzyme, *NcoI* single-digest, and *NotI/NcoI* double digest. The recombinant

and nonrecombinant plasmids behaved as expected producing no aberrant results in each restriction digest treatment(Fig. 2A). With that, the unknown plasmids identity can be concluded to be recombinant.

The same conclusion can not be drawn for the PCR. All PCR samples including the negative and positive controls, produced the same results when there should've been differences(Fig. 2B). Recombinant PCR products were expected to visualize at 1156 bp and nonrecombinant at 505 bp. Since the negative control produced results, there was most likely cross contamination in the laboratory which under normal circumstances would cause a redo to firmly confirm or deny the identity of the unknown plasmid.

Section 3: Recombinant fusion protein expression in BL21(DE3) E. coli

Recombinant, non-recombinant, and unknown plasmids were induced with IPTG for protein expression in BL21(DE3) *E. coli* and subsequent extraction of protein contents allowed for an SDS/PAGE to be conducted before nitrocellulose protein transfer. In the ponceau stain, there is an abundance of staining as all protein products from the cell lysates were stained, so no conclusions on the nature of the unknown protein could be made except for some notable stains especially in purified GST-EGFP(Fig. 3A). In the antibody stain, anti-GFP stained for the presence of EGFP at a specific epitope, and there were no abnormal results(Fig. 3B). The visualized anti-GFP stained at the expected size for its sequence of the fusion protein of 58.58kDa (Table S1). However, some degradation products lower of the expected size were present in all samples except the purified GST-eGFP because the samples come from cell lysates which contain bountiful amounts of proteins that interact with each other dynamically leading to possible degradation of the fusion gene product while the purified GST-EGFP contains nothing but the whole protein. The unknown protein is a fusion protein.

Section 4: Column chromatography of GST-EGFP fusion protein

The cell lysate of cells containing pET-41a(+)/EGFP recombinant plasmid were spiked with IPTG to induce protein expression. Various fractions of flow through and

elution were collected from a column containing agarose beads conjugated with glutathione, visualized under UV, and analyzed through absorbance and comparison to a BSA standard curve.The BSA standard curve behaved as expected by producing a linear relationship between absorbance and protein concentration(Fig. 4A). The expectation for the fractions was that the flow through would not glow while elution would. However, flow through fractions glowed while elutions did not (Fig. 4B). The reason for this could be because the linker between the two proteins broke separating GST and EGFP into two. This follows as the EGFP can not interact with glutathione conjugated to agarose bead to immobilize it, and it would have left into the first few flow through fractions. Subsequently, it is unclear if the absorbance and protein content data accurately quantify the amount of fusion protein present based on the preceding reasoning(Fig. 4C). In fact, the data suggests the gain of total protein from the cell lysates from starting of ~1147µg to a final of ~1351µg which is a protein loss of -17.79% and a % fusion protein of 0.00%(Table 1).

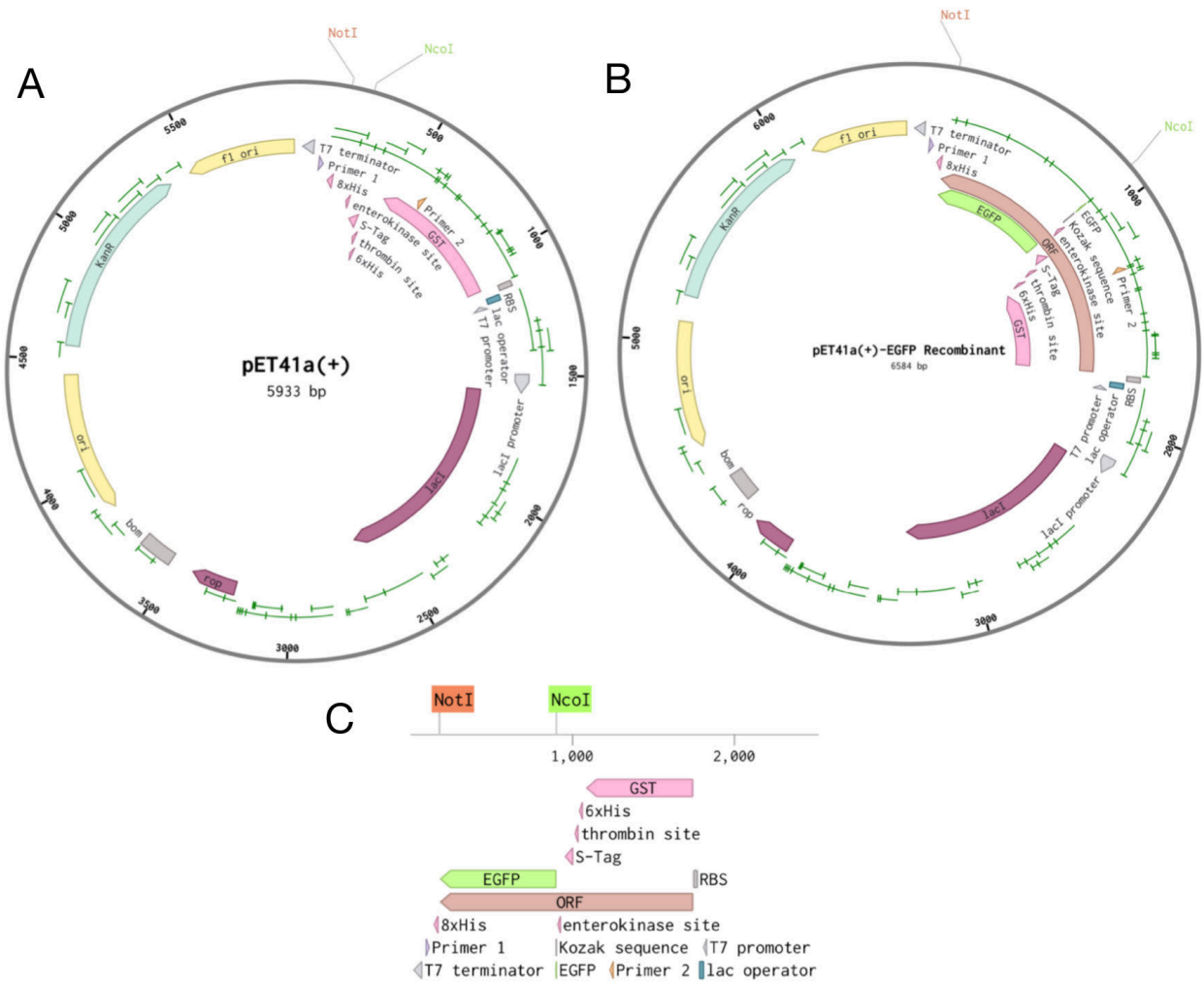


Figure 1. Recombinant and Nonrecombinant Plasmid Maps Used For Protein Expression

(a) pET41a(+) Nonrecombinant Plasmid Map For GST Protein Expression

(b) pET41a(+) Nonrecombinant Plasmid Map For Protein Expression of Fusion Protein GST-EGFP

(c) Reverse Linear Map of GST & EGFP Fusion Protein Transcription Region Regulated by Lac Operon With NotI/NcoI Multi-cloning Sites Flanked by Primer 1 & 2

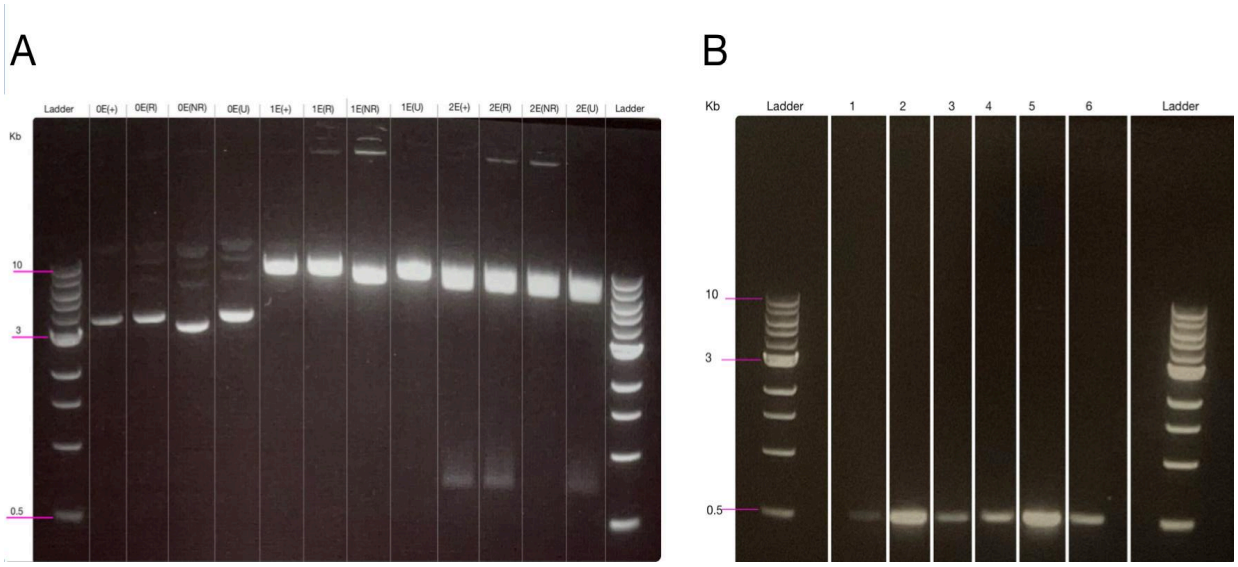


Figure 2. Gel Electrophoresis of Restriction Digest Reactions and PCR Products Used To Identify an Unknown Clone

(A) 0-2 types of restriction enzymes used on different plasmids to characterize restriction digest patterns

0E = 0 Restriction Enzymes were added to the plasmid
1E = Single restriction enzyme(NcoI) was added to plasmid
2E = Double restriction enzyme (NcoI & NotI)
(+) = Control plasmid ligation known to be made up of pET-41a(+) backbone + EGFP insert
(R) = Recombinant plasmid picked from transformed colonies glowing under fluorescent light
(NR) = Nonrecombinant plasmid picked from transformed colonies not glowing under fluorescent light
(U) = Plasmid picked from transformed colonies randomly without fluorescent light
*All samples contain 1X loading dye
*Gel made out of 90mL 0.9% agar gel of 1X TAE Buffer

(B) PCR products amplified by flanking primers 1&2

1 = PCR Product of Experimental Recombinant

2 = PCR Product of Experimental Nonrecombinant

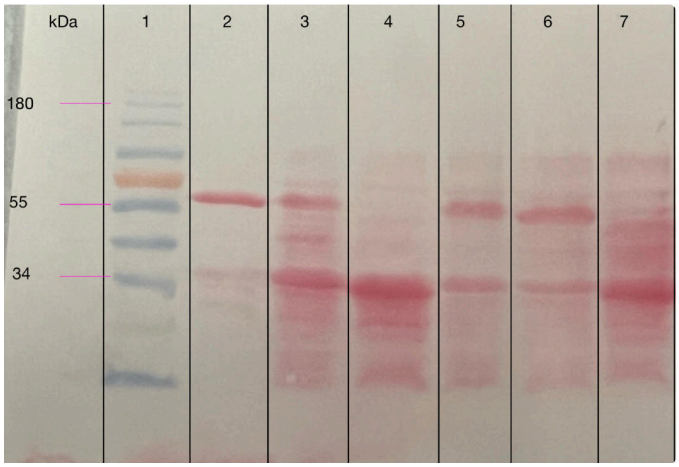
3 = PCR Product of Experimental Unknown

4 = PCR Product of Verified Recombinant

5 = PCR Product of Verified Nonrecombinant

6 = PCR Product of Negative Control

A



B

Figure 3. Ponceau stain and Antibody Stain of Protein Transfer from SDS/PAGE

1= NEB 10-250 kDa Protein Ladder

2= Purified GST-eGFP Protein

3= Experimental Recombinant Plasmid Protein Products

4= Experimental Non-Recombinant Plasmid Protein Products

5= Experimental Unknown Plasmid Protein Products

6= Verified Recombinant Plasmid Protein Products

7= Verified Non-Recombinant Plasmid Protein Products

(A) Ponceau Stain of all protein products on nitrocellulose paper from protein transfer of an SDS/PAGE

(B) Antibody stain using antibody which binds to EGFP epitope to visualize eGFP including fusion GST-EGFP

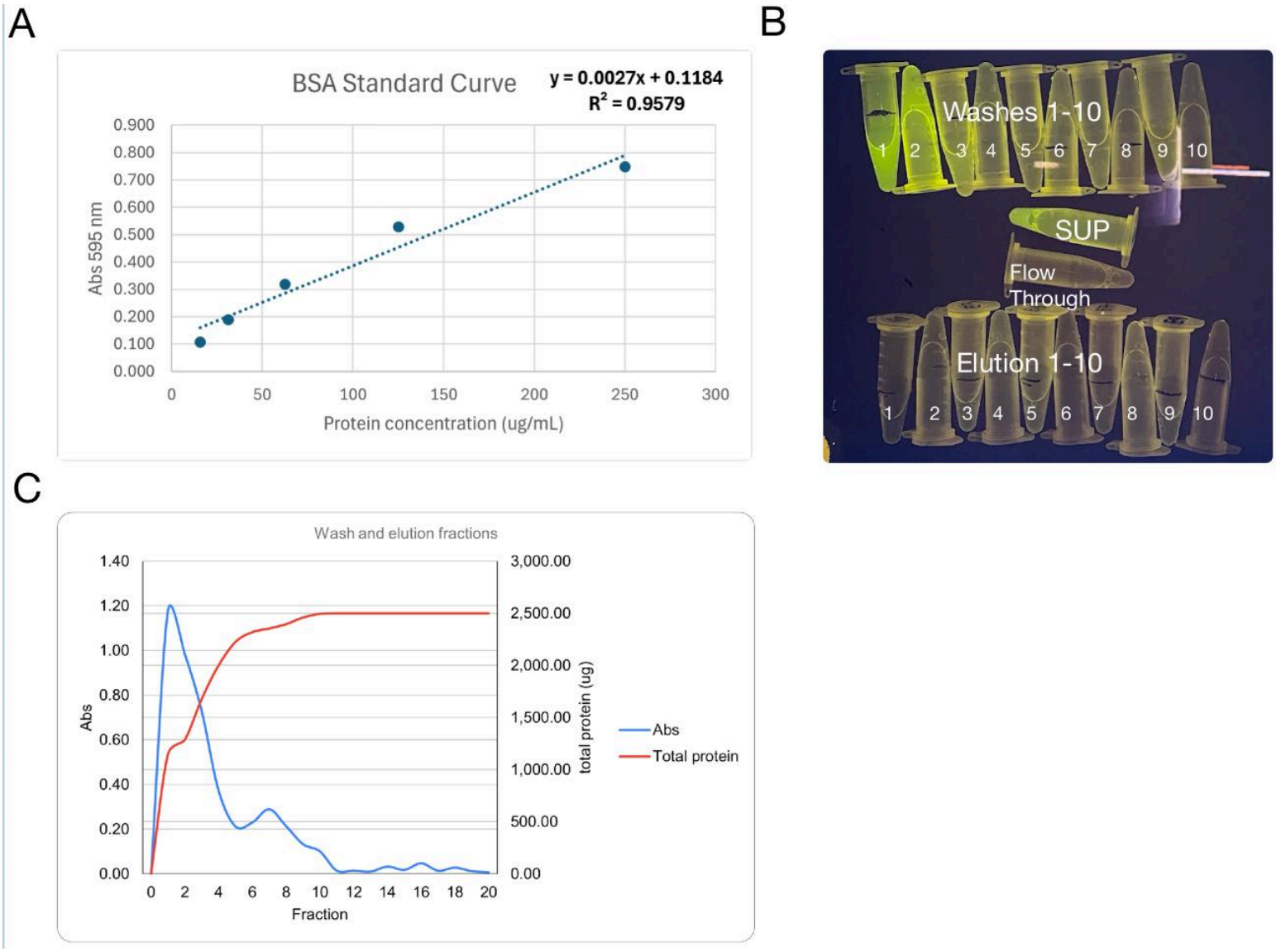


Figure 4. Affinity Chromatography for GST & EGFP Fusion Protein

(A) Protein Absorbance Standard Curve established with Bovine Serum Albumin

(B) Wash Fractions(1-10) to collect all proteins that don't interact with agarose beads conjugated with glutathione and Elution Fractions(1-10) to collect fusion protein through reduced visualized with UV

(C) Protein content in each wash and elution fraction calculated with absorbance values from a Bradford assay and BSA standard curve

Fraction	protein (ug)
Starting protein est	1147.101235
Flow through total	138.745679
Wash total	1212.37037
Elution total	0
FT + W + E	1351.116049
Protein lost	-17.79%
% fusion protein	0.00%

Table 1. Total Protein Content in Flow Through & Wash Fractions

	Sequence	Sequence Length	Expected Size(kDA)
Non-recombinant	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDE GDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIR YIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIA YSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGD HVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKK RIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPP KSDGSTSGSGHHHHHHSAGLVPRGSTAIGMKETAA AKFERQHMDSPDLGTGGGSGDDDDKSPMGYRGS EFCTGLGAPAGELRRQACGRTRAPPPPPPPLID	318	35.73
Recombinant	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDE GDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIR YIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIA YSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGD HVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKK RIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPP KSDGSTSGSGHHHHHHSAGLVPRGSTAIGMKETAA AKFERQHMDSPDLGTGGGSGDDDDKSPMVSKGEE LFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKL TLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVK FEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNT PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLE FVTAAGITLGMDELYK	518	58.58

Supplemental table 1. Protein sequences for non-recombinant protein (GST) and fusion protein(GST-EGFP)