Spring 2016

Protein Sequence-Structure-Function Relationship: Testing KE-50 Modification on Recombinant Green Fluorescent Protein (AcGFP)

Anna E. Burkey
University of Akron, aeb45@zips.uakron.edu

Protein Sequence-Structure-Function relationship: Testing KE50 Modification on Recombinant Green Fluorescent Protein (AcGFP)

Anna Burkey

Department of Chemical Engineering

Honors Research Project

Submitted to

The Honors College

Approved:	Accepted:
_____ Date 4/20/2016	_____ Date 28 APR 2016
Honors Project Sponsor (signed)	Department Head (signed)
Gang Cheng	H. MICHAEL CHEUNG
Honors Project Sponsor (printed)	Department Head (printed)
_____ Date 4/21/2016	_____ Date 28 APR 2016
Reader (signed)	Honors Faculty Advisor (signed)
Lingyun Liu	H. MICHAEL CHEUNG
Reader (printed)	Honors Faculty Advisor (printed)
_____ Date 4-28-16	_____ Date _____
Reader (signed)	Dean, Honors College
Bi-min Zhang Newby	
Reader (printed)	

PROTEIN SEQUENCE-STRUCTURE-FUNCTION RELATIONSHIP: TESTING KE-50 MODIFICATION ON RECOMBINANT GREEN FLUORESCENT PROTEIN (AcGFP)

COURSE NUMBER: 4200: 497-002

ANNA BURKEY

DR. GANG CHENG

2016-04-29

CONFIDENTIAL

Executive Summary

Problem

Many therapeutic proteins, including interferon-beta-1b, a protein used in the treatment of multiple sclerosis, create immune system responses in the patients, which is called immunogenicity. The immune system attacks the protein and removes it from the body of the patient, making the treatment much less effective and requiring large doses of protein to the patient. Protein engineering and modifications can be utilized to reduce the immune response to the protein, making protein therapy more effective. Zwitterions are especially promising due to their highly hydrophilic nature and similarity to other naturally occurring proteins. This paper investigates the addition of a string of 50 alternating lysine and glutamic acid residues (also called KE50), which may eventually improve the performance of interferon-beta-1b as a therapeutic protein. The initial step to investigate the possibility of the KE50 modification is to ensure that a protein with this modification will retain its function (the modification may denature the protein and keep it from functioning correctly). To this end, the KE50 modification was tested on a green fluorescent protein (AcGFP) which can be easily analyzed due to its fluorescence which is only present when the protein is in its properly folded state.

Results

The KE50 modification was tested on a green fluorescent protein (AcGFP), and the fluorescence of the modified protein (AcGFP KE50) was compared to the fluorescence of the original protein (AcGFP). The original protein had a quantum yield of 0.80, an extinction coefficient of 32,200 $M^{-1}cm^{-1}$, and a brightness of 25,600. The modified protein had a quantum yield of 1.0, an extinction coefficient of 16,600, and brightness of 16,600. The modified protein had 65% of the

brightness of the original protein. This demonstrates that the KE50 modification, while reducing the function of the AcGFP, does not completely denature the protein. This is promising for future work as the modification may be usable for medical proteins. Creating a KE50 modification for interferon-beta-1b was also attempted, but the protein was not successfully grown in an *E.Coli* host within the allotted time.

Broader Implications

From a personal standpoint, completing this work allowed me to gain many skills and learn about techniques used in genetic engineering and cell culture. Some of these techniques include cell transformation, use of agar plates, batch cell culture, inducing protein expression, and protein purification and analysis. I also gained more basic biology lab skills such as properly using a pipettor, using an inoculation loop, and learning about sterilization procedures. This work also allowed me to gain more skill in communication, working independently, and using resources such as online protocols and previous research to create a research plan.

This work has broader applications to society in the medical applications. Therapeutic proteins are used in the treatment of several diseases, and if the KE50 modification is effective for AcGFP and interferon-beta-1b, it may also be effective in improving the use of other therapeutic proteins by reducing immunogenicity. More effective medical treatments are an important societal benefit of biological research and genetic engineering. While other work is being completed to improve immunogenicity of proteins, most of the methods involve processing the natural protein in some way, while this method changes the protein itself without any additional modification necessary after harvesting the protein.

Future Work

Future work in this area should include finding effective growth conditions for interferon-beta-1b in *E.Coli* (including selection of the plasmid and host cell strain), growing the natural and modified forms of the protein, and testing the functionality of the modified and natural protein to test if the modified protein can still be effective as a treatment. Next, the immunogenicity of the modified protein should be tested and compared to the natural protein (this can be tested by measuring the immune response (immunoglobulin production) of rat models upon injection with the modified and natural protein).

For other students completing an Honors project, I would recommend starting early, making a clear plan for the scope and schedule of the work, and to start writing the report as early as possible. Other good advice for students completing an Honors project is to ask a lot of questions and try to understand the purpose of all procedures before writing the report.

Introduction

Protein engineering has been an area of increasing research interest as the technology improves and it is possible to create modified proteins which are more useful than their native forms. Protein modifications have potential to improve protein characteristics such as thermostability, pH-insensitivity, lower cell toxicity, and higher blood circulation times. This could extend to the therapeutic use of proteins and create protein therapies which are more effective and less harmful to the patient.

Green fluorescent protein (GFP) is a protein native to the jellyfish *Aequorea victoria*; GFP is very useful in protein engineering studies due to its bright green color and fluorescence properties which make the protein relatively easy to detect and quantify. The protein is fluorescent in its properly folded state, and loses its fluorescence when denatured (Stepanenko et al. 2013), which means that fluorescence of the protein can be used to measure to what extent the bioactivity of GFP has been retained after modification. A variation of GFP, called AcGFP, was used in this experiment: this mutant protein has a higher fluorescence than the natural strain.

Interferon-β-1b (IFN-β-1b) is a recombinant protein used in the treatment of multiple sclerosis. Like many of the related interferon signaling proteins, IFN-β-1b has a short residence time in the body after treatment (known as serum half-life), which is partially due to immune response to the protein. IFN-β-1b is also hydrophobic, making the protein difficult to purify due to the propensity to form aggregates (Basu et al, 2006).

The modification tested in this paper is the addition of a chain of 50 alternating lysine and glutamic acid residues added near the N-terminus of the protein; this will be referred to as a KE50 modification. The intended purpose of this modification is to improve blood circulation time for medical proteins. The benefit of the KE50 modification comes from the zwitterionic nature of the sequence; a zwitterion is a molecule that has both positive and negative charges, and the KE50 modification alternates positively and negatively charged amino acids, creating a zwitterionic molecule. Zwitterions can increase the aqueous solubility of large molecules, and reduce immunogenicity of proteins. The overall goal of a KE50 modification is to create protein molecules with higher stability and longer serum half-life which would make them more effective for therapeutic use. Green fluorescent protein (AcGFP) and AcGFP KE50 (green fluorescent protein with a KE50 modification) were created using a recombinant method with E.Coli.

Background

Several methods have been attempted for the improvement of IFN-β-1b and other proteins for therapeutic use. A method used commercially for other proteins and attempted successfully for IFN-β-1b is the PEGylation method, which is the addition of polyethylene glycol polymer to the protein. This modification was shown to decrease the immunogenicity of the protein, and PEGylation is used in several commercial therapeutic proteins (Basu et al, 2006). Some issues with this treatment method, including the formation of anti-PEG antibodies upon repeated

exposure have raised concerns about the long-term viability of PEGylated proteins as a treatment method (Zhang et. al. 2015).

A more recent method attempted for the modification of proteins for therapeutic use is the use of zwitterions. Zwitterions have an advantage over PEGylation in that they more closely imitate natural proteins (since natural proteins are zwitterionic) and they are more hydrophilic, which might make them more effective in therapeutic applications. One method attempted using a zwitterionic polycarboxybetaine (PCB) gel encapsulation on a protein uricase which increased the circulation half-life of the polymer. Additionally, no evidence of immune reaction was found in the rat model to either the protein or the polymer gel (Zhang et. al. 2015).

The KE50 modification attempted in this research has the some of the same advantages found in the 2015 research by Zhang et al. by using zwitterions to increase hydrophilicity and protect the protein from immune response. However, instead of encapsulating the native-type protein in a zwitterionic gel, a modification was made to the genetic information of the protein. Since many therapeutic proteins are recombinant, genetic modification of the protein sequence can easily be integrated into the large-scale manufacturing process of the protein. This would allow the creation of a zwitterion-protected protein without the additional processing step of zwitterionic gel encapsulation. This method also has the advantage of being completely made of natural protein and amino acids,which might avoid some of the immune reaction issues encountered with the use of PEGylation.

Experimental Methods

Transformation of E.Coli cells with Recombinant DNA

The initial step in producing a sample protein in the lab is the transformation of *Escherichia Coli (E. Coli)* using a vector containing the desired protein DNA. The *E.Coli* cells can be easily grown in the lab setting, and will express the DNA of the protein coded for in the vector. The first step, transformation, involves the uptake of the vector (in this case a recombinant plasmid containing the protein of interest) into the *E. Coli* cells.

Competent *E. Coli* cells were transformed with a recombinant plasmid pET28a (+) containing the DNA sequence to code each desired protein. A standard transformation protocol was used.

After transformation, the cells were cultivated on an LB agar plate containing kanomycin antibiotics. The native cell type does not have any resistance to kanomycin and should not be able to grow; however, the cells that were successfully transformed with the plasmid will gain resistance to the antibiotic and can grow colonies on the plate. This is how successfully transformed cells are selected from the initial batch of competent cells. The successful colonies can be grown in LB medium to increase the amount of transformed cells.

Protein Expression and Purification

After growing the required amount of transformed cells for protein synthesis, expression of the protein in the plasmid is induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG induces transcription of DNA in the lac operon, which is the operon contained in the pET 28a(+) plasmid. IPTG was added to induce expression of the protein, and the cell culture was allowed to ferment to express the protein.

After the protein expression was completed, the cells were centrifuged (to separate from the growth medium) and lysed (to break the cells and release the protein into the lysis buffer). Lysozyme, sonication, and nuclease were used to lyse the cells and break down the cell DNA to allow separation of protein from other cell components. Centrifugation was used to separate the protein (supernatant) from other cell components (precipitate).

The engineered recombinant proteins included a string of histidine residues near the N-terminus of the protein, allowing for a simple separation of the target protein from other cell proteins. A column with Ni-NTA His-Tag resin was used, which has an affinity for the histidine tag on the recombinant protein. When the protein sample is added to the column, the recombinant protein will adsorb onto the resin in the column, allowing other proteins to flow through while retaining the target protein. After several washing steps to remove other proteins, another buffer is used to flood the column with a ion with high affinity for the resin (imidazole) to remove the target protein from the column.

Confirming the Identity of Protein Samples

The presence of green fluorescence protein in a sample was confirmed using a microscope with a colored light to view the fluorescence of the AcGFP inside of the cell samples.

To detect the presence of IFN-β-1b, an SDS-PAGE method was used to separate the proteins in a sample by size. A standard protein ladder was used alongside unknown samples to estimate the size of unknown proteins. To confirm the presence of IFN-β-1b in a sample, a Western Blot protocol was used. The primary antibody used in the Western Blot was IFN IgG interferon-beta antibody, which will preferentially bind to IFN-β-1b. The secondary antibody was Goat anti-rabbit IgG, which binds to the primary antibody and allows detection using spectrophotometric methods.

Fluorescence Measurements of AcGFP and AcGFP-KE50

To determine the effect of the KE50 modification of protein folding and bioactivity, fluorescence measurements were performed to compare AcGFP and AcGFP-KE50. The measurements used were quantum yield and extinction coefficient.

Quantum yield was measured by measuring the absorbance at 488 nm and the fluorescence emission curves from 450-750 nm. These measurements were made at 5 different concentrations for the sample and for a reference of known quantum yield. In this experiment, the reference

used was fluorescein with a quantum yield of 0.92. The integral of the emission curve was plotted against the absorbance for each concentration. This plot was approximately linear, and the slope of the plot for the standard and sample was used in the following equation to calculate the quantum yield:

$$\Phi_X = \Phi_{ST}\left(\frac{M_X}{M_{ST}}\right)\left(\frac{\eta_X^2}{\eta_{ST}^2}\right) = \Phi_{ST}\left(\frac{M_X}{M_{ST}}\right)$$

Where φ_X is the quantum yield of the sample, φ_{ST} is the quantum yield of the standard, M_X is the slope of the plot from the sample, M_{ST} is the slope of the plot from the standard, and η is the refractive index of the solvent (in this case η is the same for the standard and the sample).

The extinction coefficient was calculated using the measured absorbance for 5 different concentrations of each protein, and using the Beer-Lambert law to plot the absorbance against the concentration and pathlength in order to find the extinction coefficient (the slope of the Beer-Lambert plot). The Beer-Lambert law is expressed with the following equation:

$$A = \varepsilon C l$$

Where A is the absorbance, ε is the extinction coefficient, C is the concentration of the sample, and l is the pathlength of the cuvette.

Data and Results

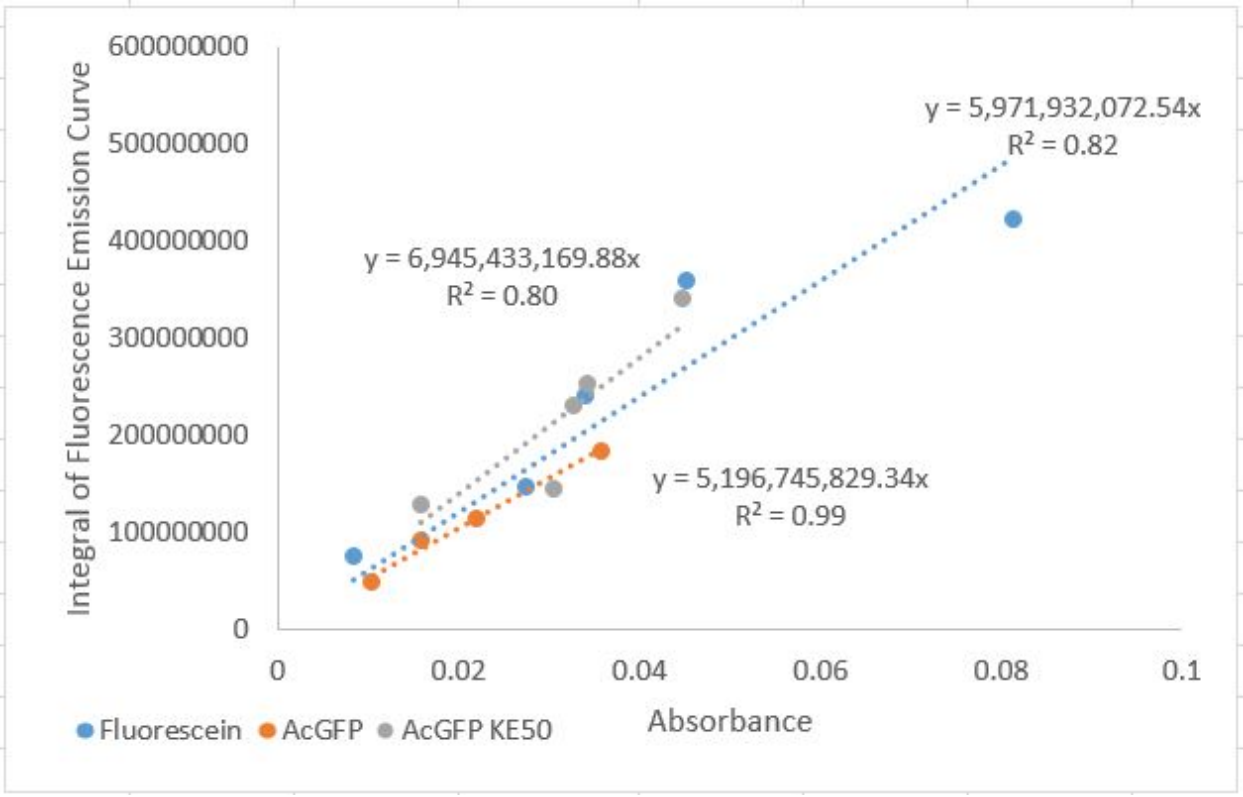

Figure 1. Integral of fluorescence emission curve from 480nm-650nm versus the absorbance of the sample at 488 nm at 5 different solution concentrations

Table 1. Slopes of plots from Figure 1 and quantum yield calculated for each sample using the slope

	Quantum Yield	Slope
GFP	0.80	5.2
GFP KE50	1.06	6.9
Fluorescein	0.92	6.0

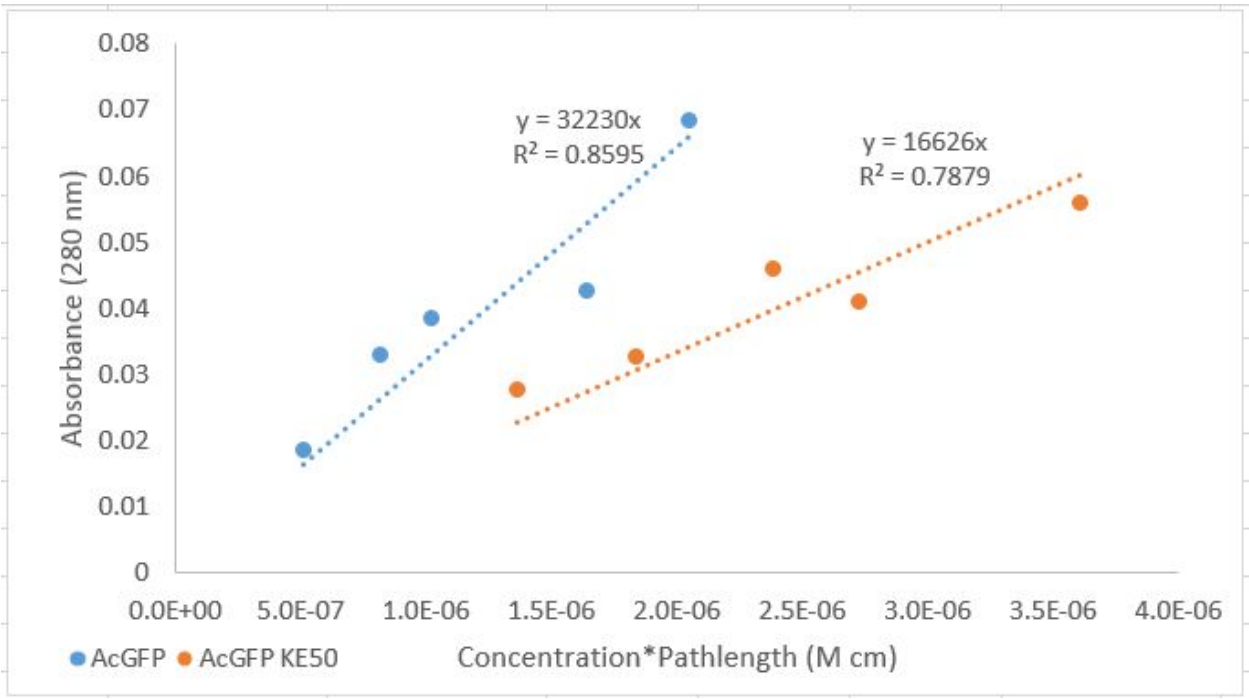

Figure 2. Absorbance plotted against the concentration of the sample multiplied by the pathlength: the slope of each linear plot is the extinction coefficient.

Table 2. Extinction coefficient found using Figure 2 and the brightness of each sample found by multiplying the extinction coefficient times the quantum yield.

	Extinction Coefficient ($M^{-1}cm^{-1}$)	Brightness (QY*EC)
GFP	32200	25674
GFP KE50	16600	16600

The modified protein had 65% of the brightness of the original protein.

Growth of IFN-β-1b and IFN-β-1b KE50 was attempted by using transformation in the same manner as for AcGFP and AcGFP KE50. After separating the protein samples with SDS-PAGE, a Western blot procedure was used to detect the interferon protein. This was attempted with the following *E. Coli* strains and the transfection was not successful: Nova Blue, BL21, C41 (DE3) and C43(DE3).

Discussion/Analysis

Transfection and growth of AcGFP and AcGFP KE50 proteins were successfully completed in *E. Coli* cells. The proteins were purified and tested to determine whether or not the protein would

successfully retain its function or if the KE50 modification would cause denaturation and loss of function.

While the quantum yield of the AcGFP KE50 was higher than the original AcGFP, the extinction coefficient of the AcGFP KE50 was lower, resulting in a lower overall brightness. The modified protein had 65% of the brightness of the original protein. The quantum yield measures the percentage of light absorbed that is emitted as fluorescence, and the extinction coefficient is a measure of how much light is absorbed by a certain concentration of the substance. Therefore, the brightness defined here as quantum yield multiplied by the extinction coefficient is a measure of the fluorescence intensity per concentration.

The results show that the KE50 variant has lower brightness, meaning that the KE50 modification deforms the protein and reduces activity. However, the modification does not completely eliminate the function of the protein, and may be an effective method to reduce immunogenicity of therapeutic proteins. These results are consistent with current theories related to sequence-structure-function relationships. Altering the sequence of the protein affected the structure of the protein, and therefore altered its function. While this paper demonstrates reduced function of the protein, further work needs to be done to determine whether the KE50 modification has the intended effect of reduced immunogenicity.

The measured quantum yield and extinction coefficient for AcGFP was relatively close to the literature value ($\varphi = 0.82$ and $\varepsilon = 32{,}500\ M^{-1}cm^{-1}$ respectively) (Clontech 2015). The measured extinction coefficient may be less accurate due to uncertainty in the concentration measurement method for the initial protein solution and uncertainty/inaccuracy of the dilutions. The quantum yield measurement should be more accurate due to less reliance on the accuracy of the known solution concentration. However, the quantum yield result for AcGFP KE50 of 1.0 is strange and does not seem correct. Several of the data points used in the calculation of the quantum yield seem to be outliers from the remaining data and may be inaccurate; further work will be completed to determine the accuracy and repeatability of these results. One way to improve the quantum yield measurement and ensure better results would be to use two fluorophore standards instead of one and cross-reference the results with known values to ensure that the equipment and measurement methods are accurate.

The transformation of the *E.Coli* strains with a pET28a(+) plasmid containing IFN-β-1b and IFN-β-1b KE50 was not successful. The failure to produce detectable levels of IFN-β-1b protein could be due to several factors such as unsuccessful transformation, spontaneous genetic mutation of the selected strains, protein toxicity to the cells, insolubility of the protein, or a combination of factors. Since each separate attempt to transform, grow cells, express proteins, and test for proteins takes several days, more time would be required to find a host cell strain and plasmid combination that is able to express a significant amount of IFN-β-1b protein using *E.Coli* cells.

Literature Cited

Basu et al. (2006) Structure-Function Engineering of Interferon-â-1b for Improving Stability,Solubility, Potency, Immunogenicity, and Pharmacokinetic Properties by Site-Selective Mono-PEGylation. Bioconjugate Chemistry. 17, 618-630

Clontech (BD Biosciences) (2015) BD Living Colors AcGFP1 Fluorescent Protein: A novel monomeric green fluorescent protein alternative to EGFP. Accessed online. http://wolfson.huji.ac.il/purification/PDF/Tag_Protein_Purification/FluorescentProteins/BD_AcGFP1MonFluorProt.pdf

Pletneva, et al. (2010) Structural Evidence for a Dehydrated Intermediate in Green Fluorescent Protein Chromophore Biosynthesis. J Biol Chem. 285 (21) 15978-15984.

Stepanenko et al. (2013) Beta-Barrel Scaffold of Fluorescent Proteins: Folding, Stability and Role in Chromophore Formation. Int Rev Cell Mol Biol. 302, 221-278.

Zhang et al. (2015) Zwitterionic gel encapsulation promotes protein stability, enhances pharmacokinetics, and reduces immunogenicity. PNAS. 112 (39) 12046-12051.

Appendices

Table A.1 Absorbance and integral of emission curve data used to calculate the quantum yield in Figure 1

Protein	Sample	Absorbance (488 nm)	Integral of Emission Curve
Fluorescein	1	0.034	241700000
Fluorescein	2	0.0452	360100000
Fluorescein	3	0.0813	422100000
Fluorescein	4	0.0085	75090000
Fluorescein	5	0.0275	148200000
AcGFP	1	0.0106	48770000
AcGFP	2	0.0159	91560000
AcGFP	3	0.0223	114000000
AcGFP	4	0.0357	184600000
AcGFP	5	0.0556	251100000
AcGFP KE50	1	0.0158	128600000
AcGFP KE50	2	0.0306	146000000
AcGFP KE50	3	0.0328	230100000
AcGFP KE50	4	0.0343	254000000
AcGFP KE50	5	0.045	339900000

Table A.2. Raw data used to create Figure 2. Absorbance measured at 280 nm and molar concentration of protein samples

Protein	Sample	Absorbance (280 nm)	Concentration (M)
AcGFP	1	0.019	5.12E-07
AcGFP	2	0.033	8.19E-07
AcGFP	3	0.039	1.02E-06
AcGFP	4	0.043	1.64E-06
AcGFP	5	0.068	2.05E-06
AcGFP KE50	1	0.028	1.36E-06
AcGFP KE50	2	0.033	1.84E-06
AcGFP KE50	3	0.046	2.38E-06
AcGFP KE50	4	0.041	2.73E-06
AcGFP KE50	5	0.056	3.61E-06

Emission curve data of protein samples integrated to obtain value in Table A.1

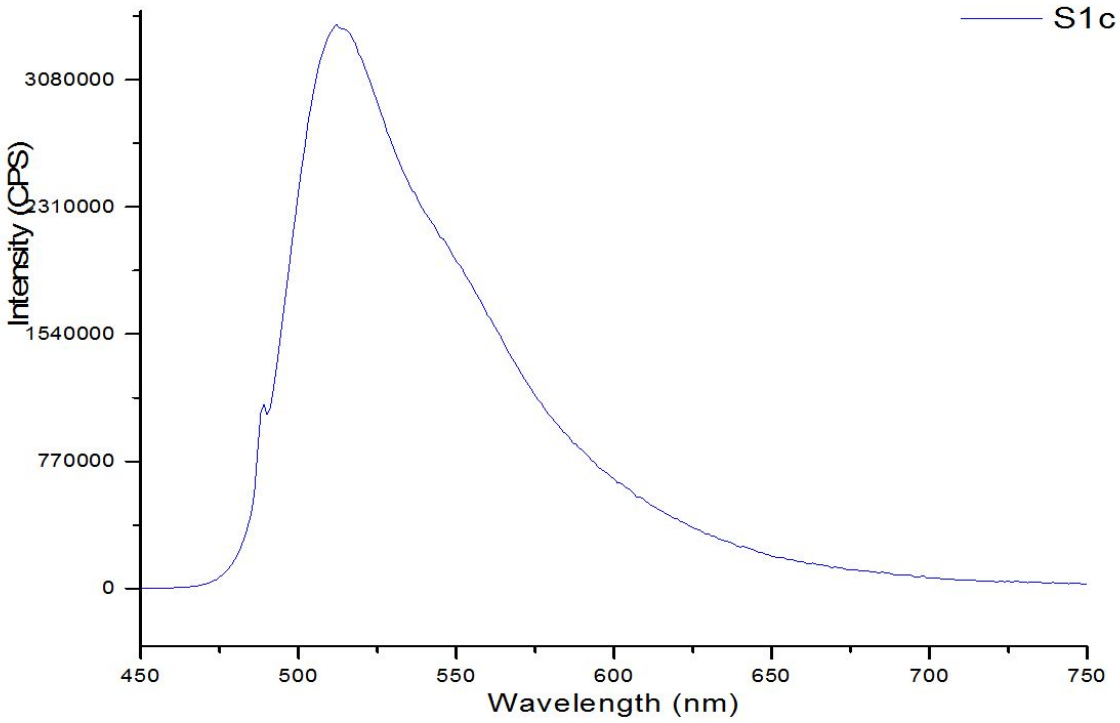

Figure A.1. Emission curve of fluorescein sample 1

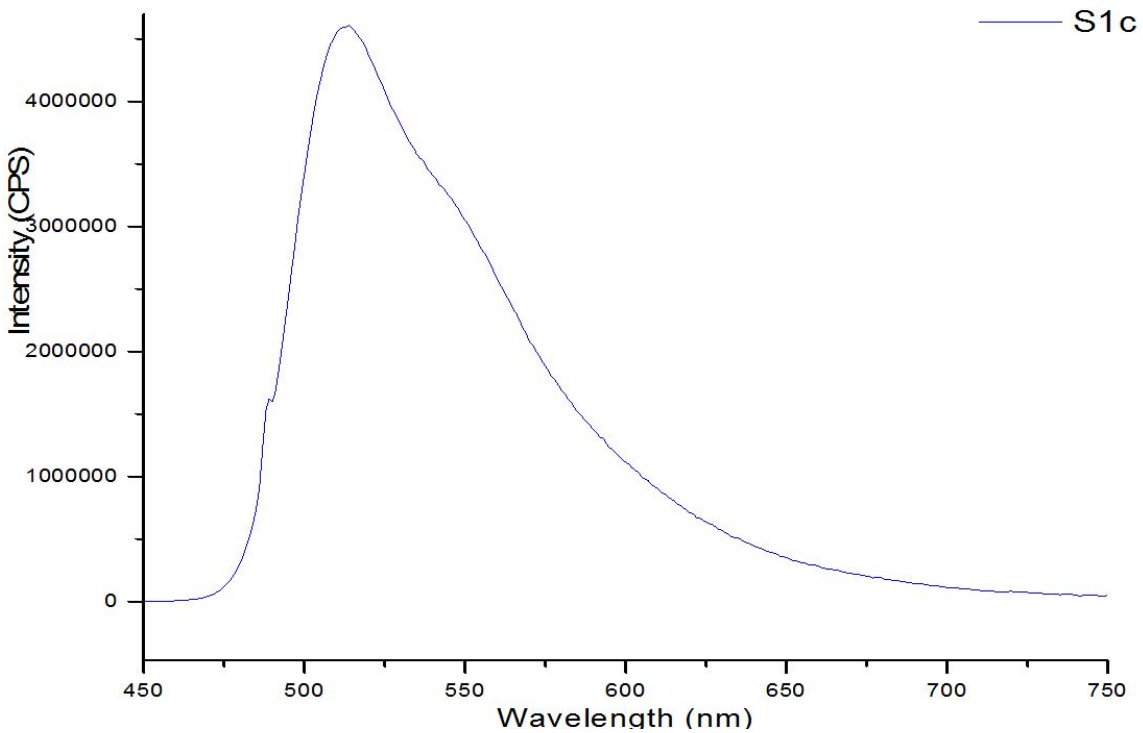

Figure A.2. Emission curve of fluorescein sample 2

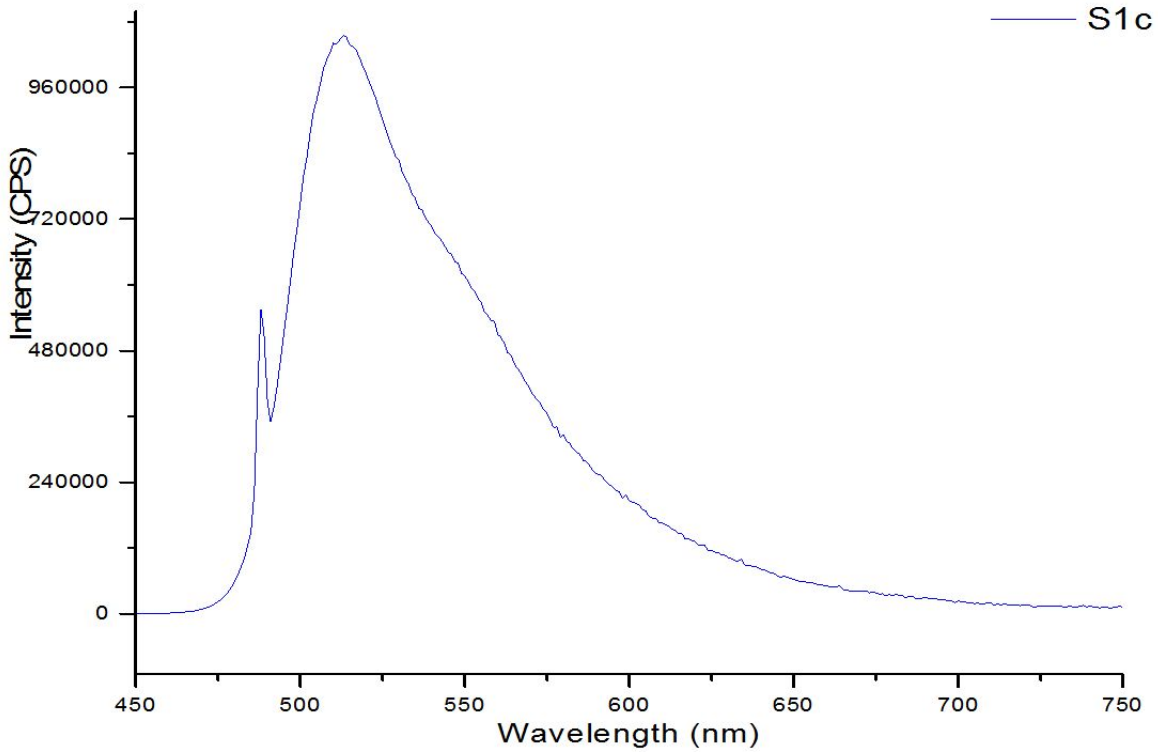

Figure A.3. Emission curve of fluorescein sample 3

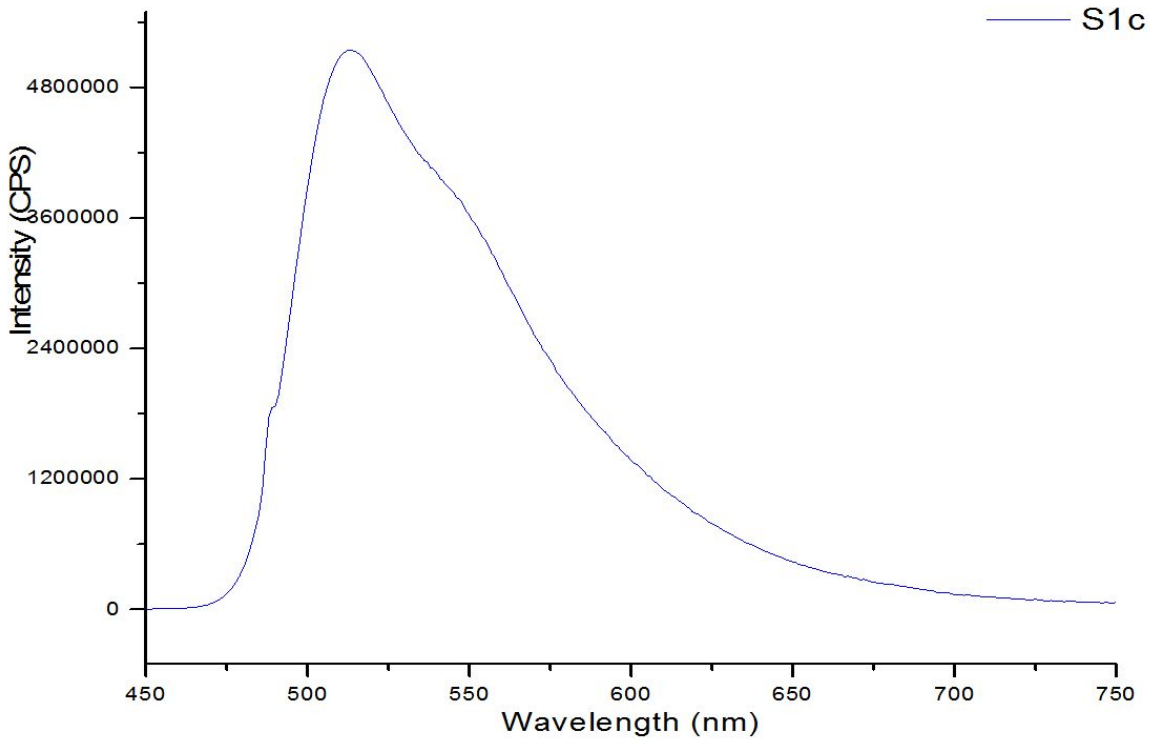

Figure A.4. Emission curve of fluorescein sample 4

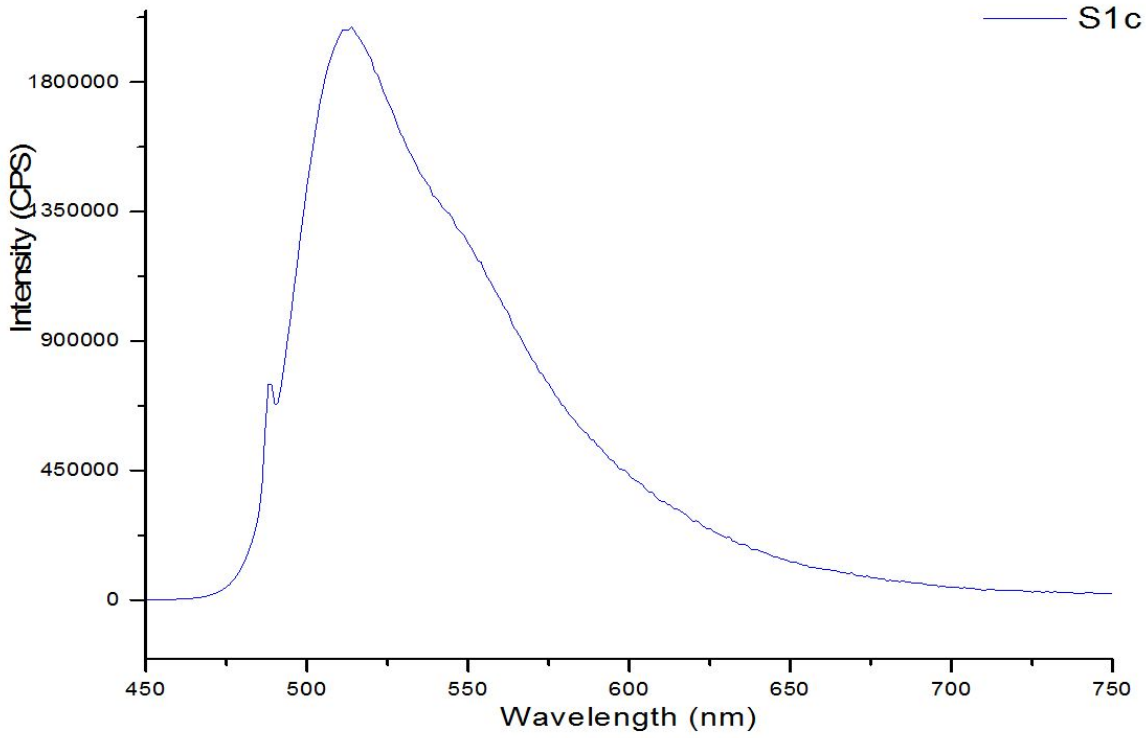

Figure A.5. Emission curve of fluorescein sample 5

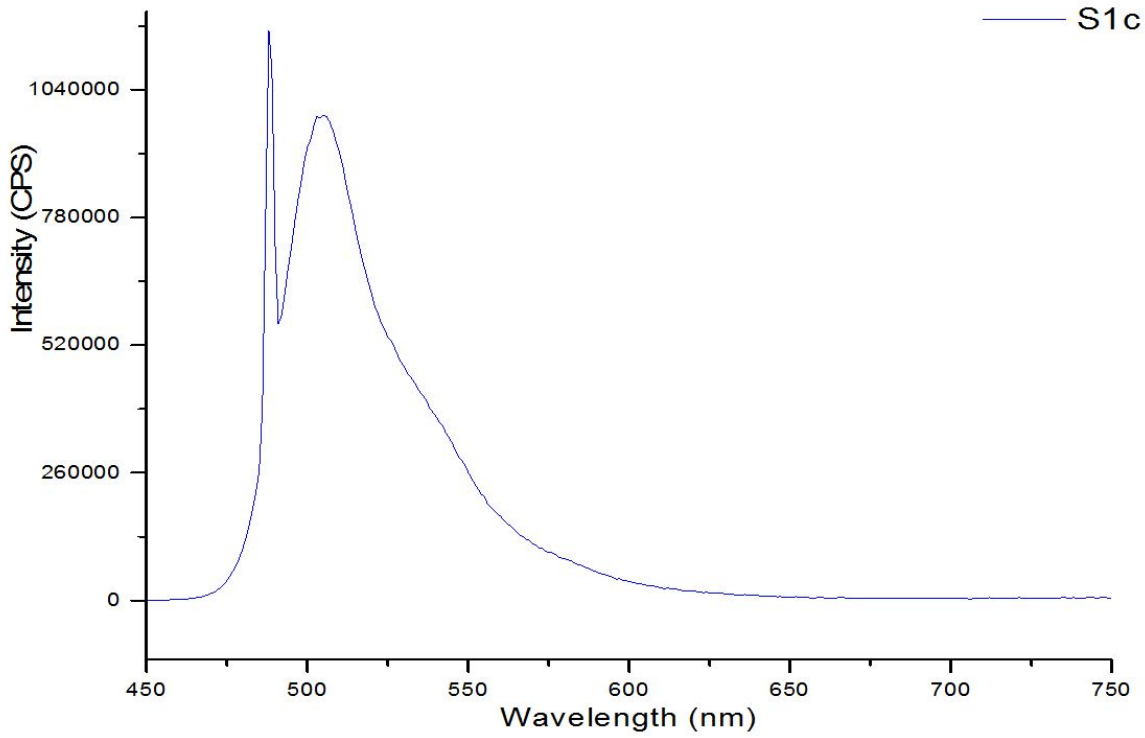

Figure A.6. Emission curve of AcGFP sample 1

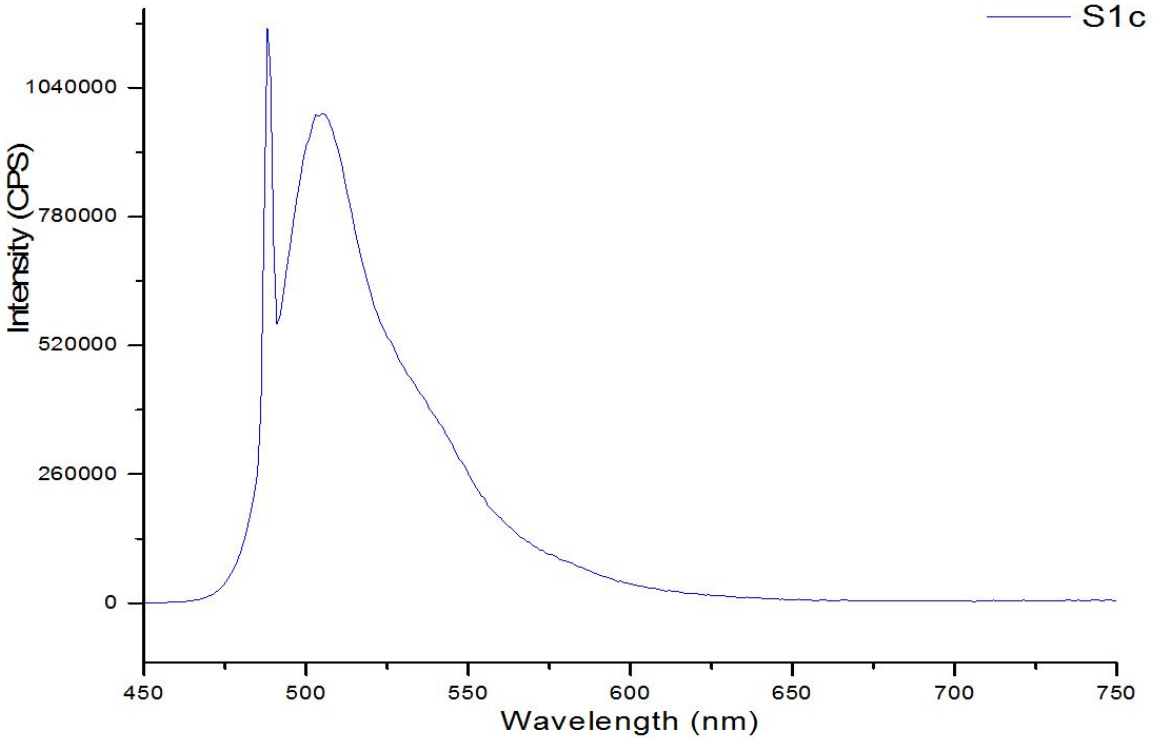

Figure A.7. Emission curve of AcGFP sample 2

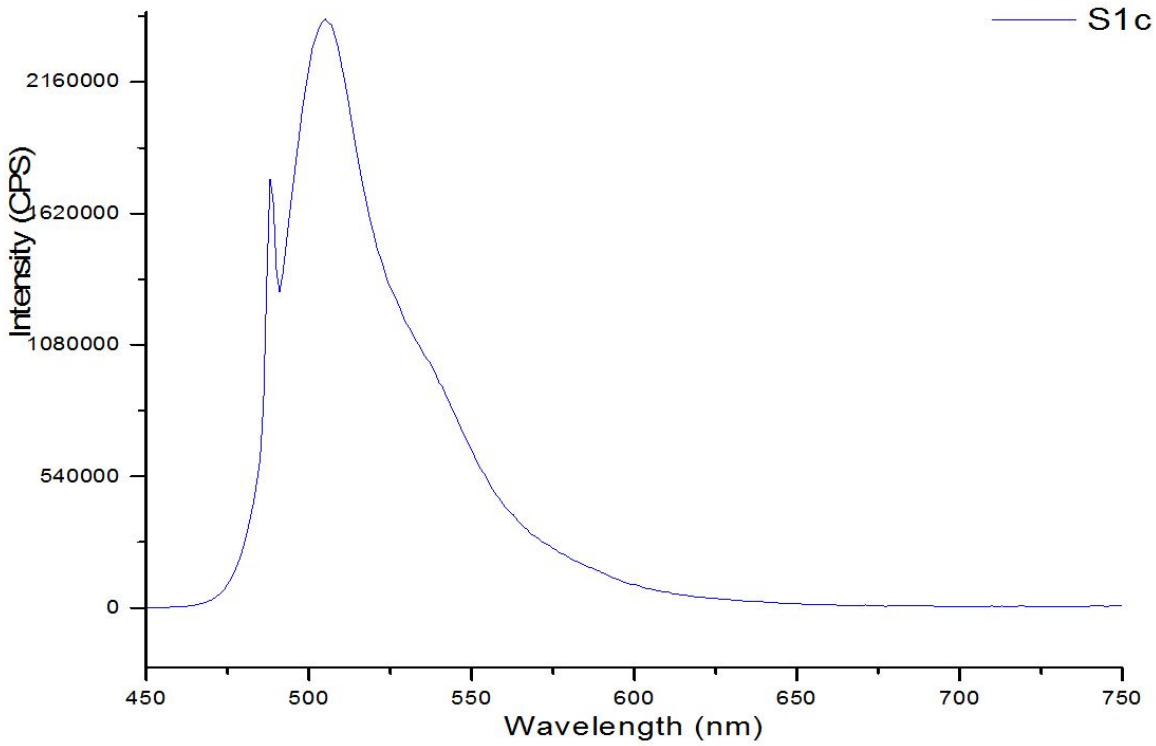

Figure A.8. Emission curve of AcGFP sample 3

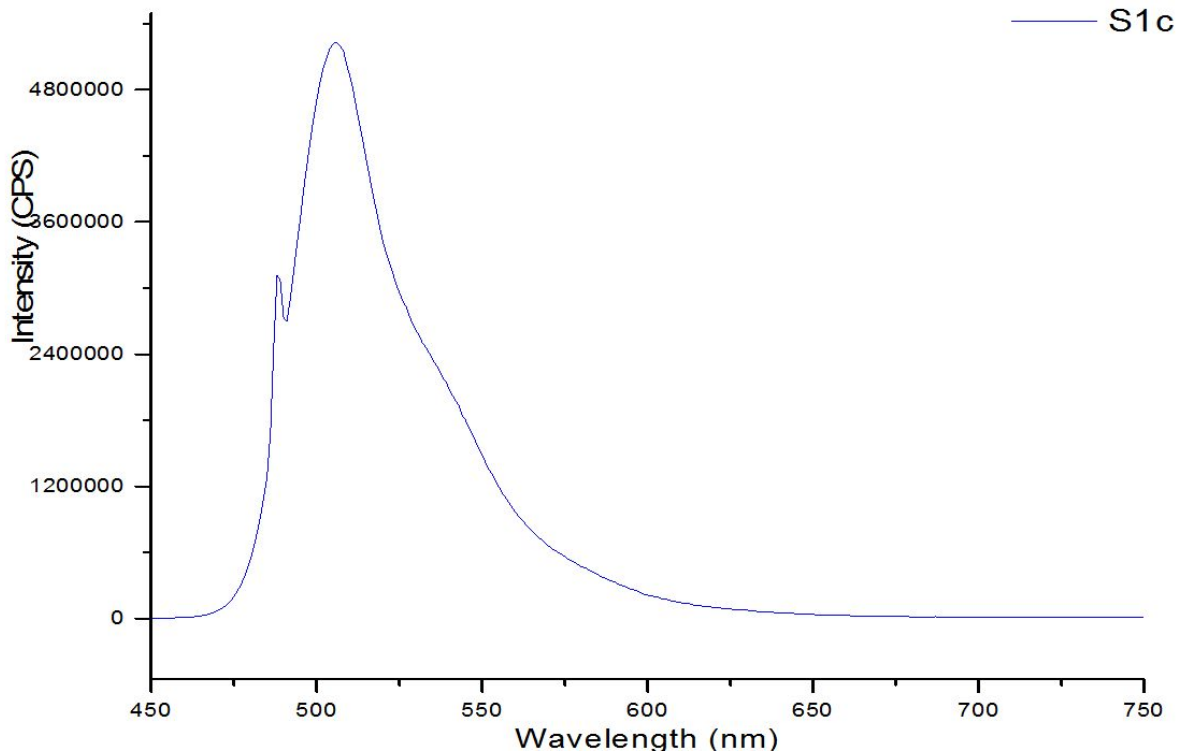

Figure A.9. Emission curve of AcGFP sample 4

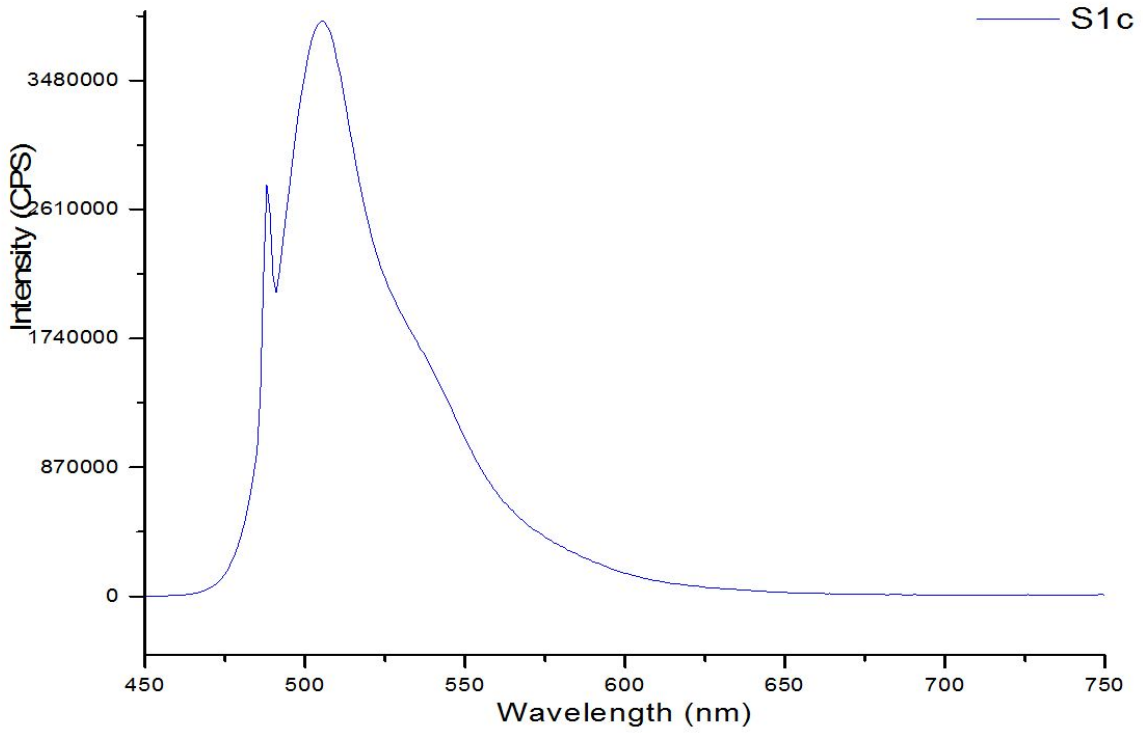

Figure A.10. Emission curve of AcGFP sample 5

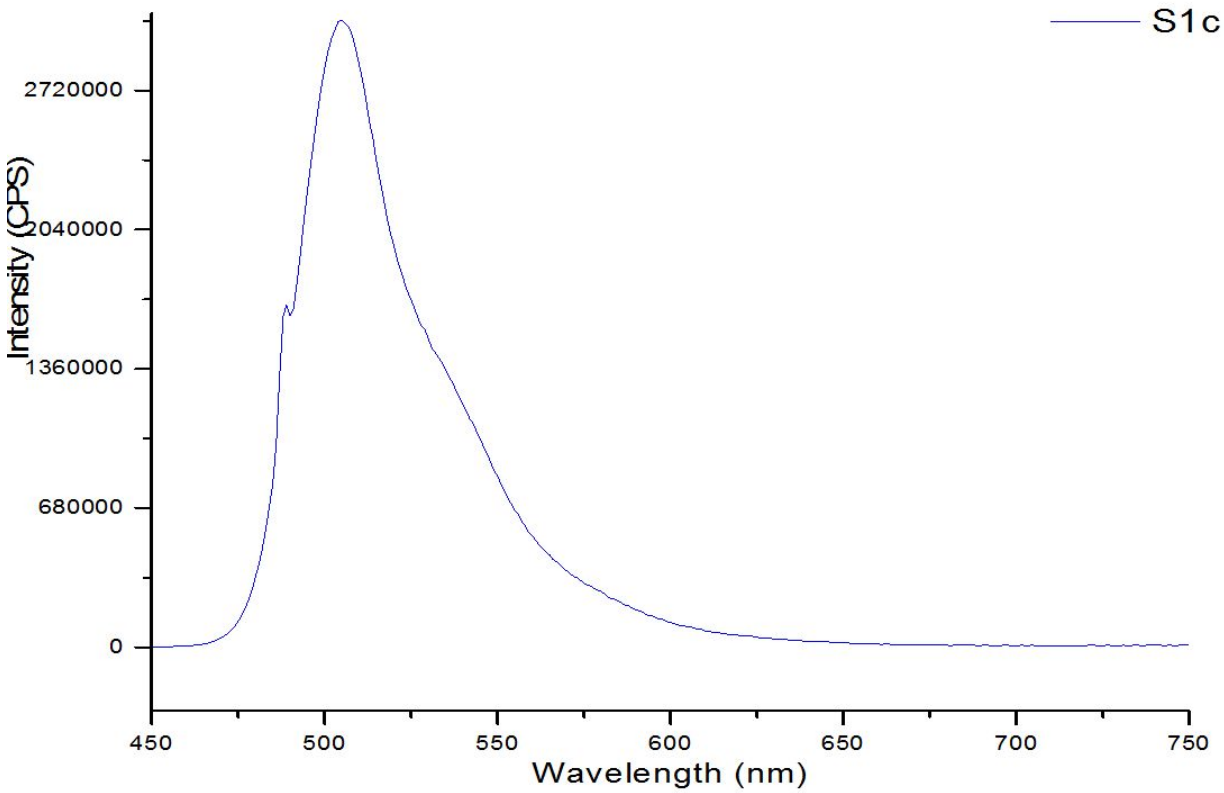

Figure A.11. Emission curve of AcGFP KE50 sample 1

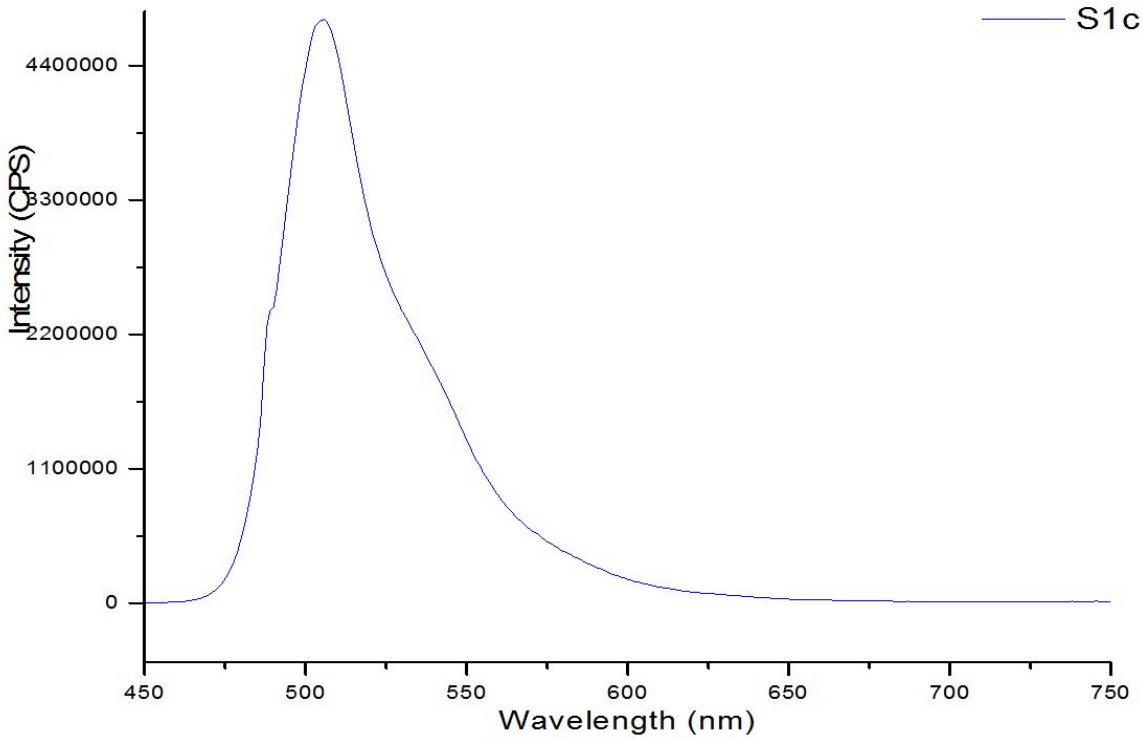

Figure A.12. Emission curve of AcGFP KE50 sample 2

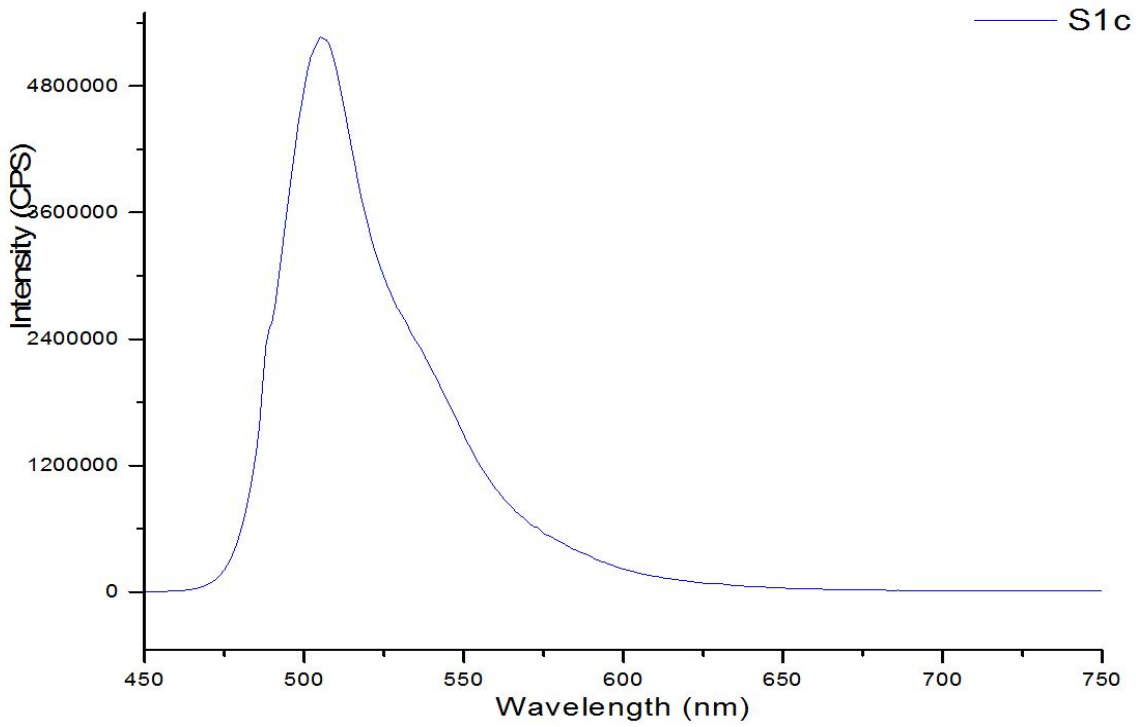

Figure A.13. Emission curve of AcGFP KE50 sample 3

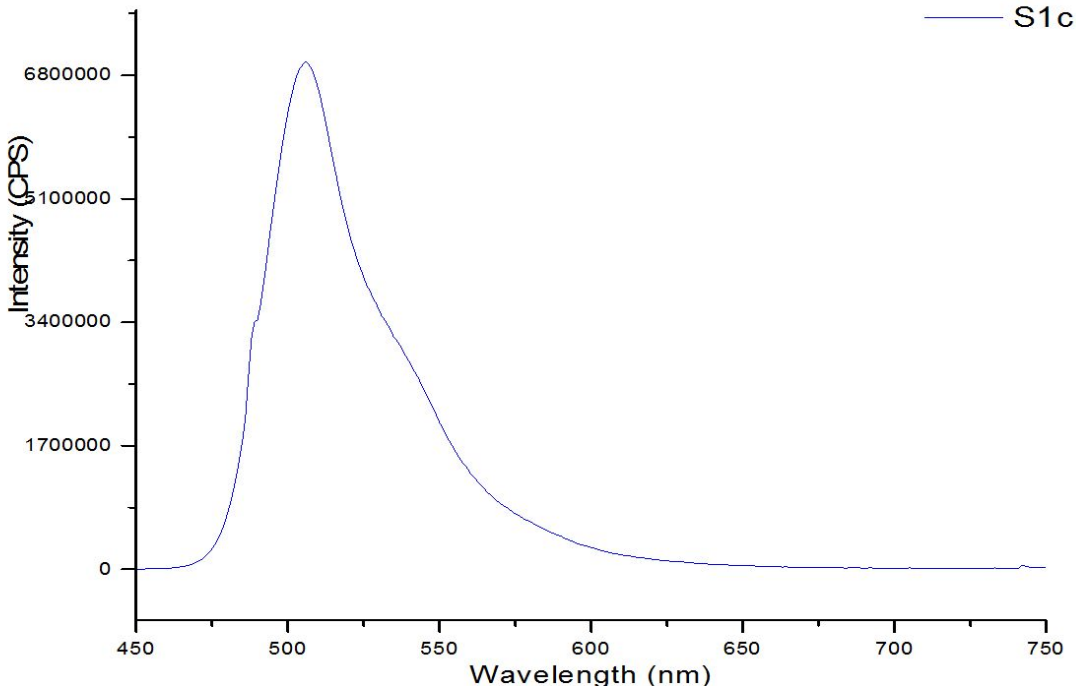

Figure A.14. Emission curve of AcGFP KE50 sample 4

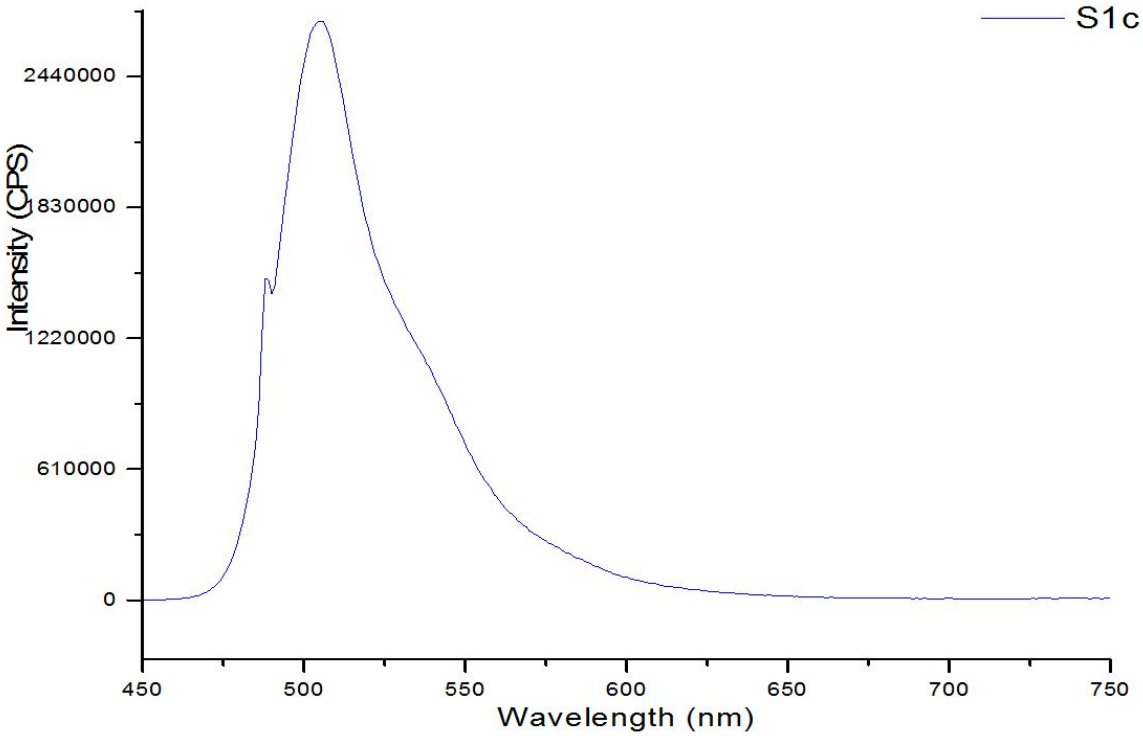

Figure A.15. Emission curve of AcGFP KE50 sample 5