# Emission Spectroscopy: Biophysics and FRET

## Introduction and Background

For this lab in particular, many students have trouble seeing the forest for the trees. Thus, to aid in a better understanding of the overall goal, the important points are presented here.

* Cytochrome *c* (cyt-c) is a protein. It contains one **tryptophan residue**, and one **heme cofactor** (among other things).
* By themselves, tryptophan molecules are **fluorescent**. They absorb ultraviolet light (<300 nm), gaining excess energy and becoming “**excited**” for a short period of time, and then release this energy by emitting longer wavelength light (~350 nm).
* When an excited tryptophan is close enough to a heme cofactor, its excess energy may be transferred to the heme *instead* of being emitted as fluorescence. This occurs by a process known as **Förster resonance energy transfer**, or **FRET**.
* When the cyt-c protein is in its “native” or folded state, the tryptophan and heme are close together and FRET is very efficient. This causes the fluorescence of the tryptophan residue to be decreased or “**quenched**” by the heme.
* As urea is added to the cyt-c solution, it disrupts the forces that hold the protein in its shape, causing it to unfold or “**denature**”. Thus, the tryptophan and heme move apart, the FRET efficiency decreases, and the measured fluorescence increases.

## Outline

We will only be doing parts A, B, and C in the manual. Part D is simply a repeat of Part A using a different chemical denaturant (guanidinium hydrochloride instead of urea) so we won’t worry about it. The goals of these three parts are as follows:

**Part A**. Study the thermodynamics of cyt-c folding in the presence of urea by a fluorometric analysis. Examine how urea changes the free energy of unfolding, and determine the free energy of unfolding in pure water. *Knowledge of FRET is not necessary*.

**Part B**. Study the kinetics of FRET between tryptophan and heme. Describe how the FRET efficiency changes with the intermolecular distance, and find the “**Förster distance**” at which the efficiency is 50%. *Knowledge of protein structure is not necessary*.

**Part C**. Apply the results of Part B to the data obtained in Part A. This is where we will analyze the protein folding in terms of the FRET mechanism in order to obtain information about the structure of cyt-c.

## Part A – Unfolding Curve and Free Energy Change

In this part we will generate an “unfolding curve” for the denaturation of cyt-c by urea. Like any physical or chemical process, there is a change in free energy associated with this unfolding, denoted . The “U” subscript indicates that the change in free energy is a function of the urea concentration. **Our goal in this part is to find , the change in free energy of unfolding in pure water, *i.e.* when [Urea] = 0.**

In pure water, we know that cyt-c remains folded, which means must be positive (*i.e.* non-spontaneous). As you add urea, the unfolding becomes spontaneous, meaning that decreases until it becomes negative. The relationship between and the urea concentration is given in Equation 9 in your manual, and reproduced here:

|  |  |  |
| --- | --- | --- |
|  |  | ( 9 ) |

In Eq. 9, the variable C is the concentration of urea. If you look at this as the equation of a line that describes how changes when the concentration of urea is varied, it should be clear that the parameter *m* is the slope of the line. In other words, it describes how quickly decreases as you add urea. It should also be obvious that is the y-intercept of the line at the point where C = 0. In other words, it is the free energy change in pure water, when there is no urea present. The x-intercept of this graph is the point at which . This occurs when the urea concentration is just sufficient to create equal populations of folded and unfolded proteins. We call this particular concentration Cm, and it must satisfy the following equation (Eq. 10 in your manual).

|  |  |  |
| --- | --- | --- |
|  |  | ( 10 ) |

Since these equations can be somewhat abstract and difficult to interpret, we will instead look at an “unfolding curve” – a graph of the fraction of unfolded proteins vs. urea concentration. This curve clearly has a more direct physical interpretation, and we will see how it varies from 0 (when there is no urea) to 1 (when there is excess urea). It should be obvious that the fraction of unfolded molecules, *f*, is given by the following equation (Eq. 11 in your manual).

|  |  |  |
| --- | --- | --- |
|  |  | ( 11 ) |

In Eq. 11, [U] is the concentration of unfolded proteins, and [F] is the concentration of folded proteins. Like any system in thermodynamic equilibrium, the fraction of molecules in a given state is also described by the Boltzmann distribution. This allows us to obtain a separate equation for *f* (Eq. 12 in your manual).

|  |  |  |
| --- | --- | --- |
|  |  | ( 12 ) |

Note that the exponents in Eq. 12 are simply , as they should be for a Boltzmann distribution. Eq. 11 is based on simply counting molecules, while Eq. 12 is derived from the laws of thermodynamics, but they must be equal to one another. If we can somehow experimentally determine the fraction *f* as a function of urea, then we can fit that data to Eq. 12 in order to extract the desired parameters, m and Cm.

In order to “count” the number of proteins in the folded and unfolded states, we must rely on a fact that we mentioned above – tryptophan in a folded cyt-c protein does not fluoresce, but in an unfolded protein it does. We don’t need to know the mechanism for this change, only that it occurs and that there is an equilibrium between the two states. Since the observed fluorescence is proportional to the concentration or number of fluorescing molecules, we can derive the following equation from Eq. 11.

|  |  |  |
| --- | --- | --- |
|  |  | ( 11A ) |

In this equation, F(C) is the observed fluorescence (at 350 nm) of a solution with urea concentration C, F0 is the observed fluorescence of the solution with no urea, and is the observed fluorescence of a solution with enough urea to fully denature all of the cyt-c (*i.e.* the solution with C = 10 M). Since Eq. 11A only involves quantities that we can measure, it allows us to generate the experimental unfolding curve by calculating *f*, and then fit it to Eq. 12.

### Nonlinear Curve Fitting

Fitting the data obtained from Eq. 11A to the model Eq. 12 requires the use of regression analysis, as described in “Nonlinear Least-Squares Curve Fitting with Microsoft Excel Solver” by Daniel C. Harris (*J. Chem. Ed.* **1998**, *75*, 119-121). The Solver add-on for Microsoft Excel is usually included in Windows versions and later Mac versions. See the following links for further help in obtaining or activating Solver:

<http://www.solver.com/excel-solver-how-load-or-start-solver>

<http://www.solver.com/welcome-mac-users-solver-now-included-excel-2011>

You are expected to read that article and understand the concept in order to implement the curve fitting procedure for your data. An outline is given here.

1. For each of your solutions, the **UV-Vis absorbance spectra must be corrected** by subtracting the corresponding absorbance blank you recorded for that solution before adding cyt-c. Once this is done, you must calculate the **concentration of cyt-c** in each solution by taking the corrected absorbance value at 410 nm and dividing it by the extinction coefficient given in your manual (ε410 = 105,000 M-1∙cm-1). The samples should all have very similar concentrations, but there will be slight differences.
2. The **fluorescence spectra must be corrected** in the same way, by subtracting the corresponding fluorescence blank recorded before adding cyt-c. This will remove any background scattering signal, and should also get rid of the Raman scattering peak of the water.
3. After this blank subtraction, the **fluorescence must be** **normalized** by dividing each spectrum by the concentration of cyt-c for that solution calculated in step 1. This is done to account for small variations in the cyt-c protein concentration that can cause differences in the fluorescence.
4. Once these corrections and normalizations are done, start a new worksheet in Excel and make two columns labelled something like “Urea (C)” and “Fluorescence (F)”. In the first column (Column A), enter the **molar concentrations of urea** (*not* the cyt-c concentration) that you calculated for each solution. This column should vary from 3 to 10 in 0.5 M increments. In the second column (Column B), enter the corresponding value for the **corrected, normalized fluorescence at 350 nm.**
5. Label a 3rd column (Column C) something like “fraction (experimental)” and enter a formula corresponding to Eq. 11A above. You will need to have the formula refer to the correct cells that contain the value of F(C) from the second column, as well as the static parameters F0 and , which are the fluorescence values for C = 0 and C = 10, respectively.
6. Make a 4th column (Column D) labelled something like “fraction (theoretical)”. In this column you will enter a formula corresponding to Eq. 12 above. In this equation, C will refer to the corresponding cell in the first column containing the proper urea concentration, and R and T are just the gas constant and temperature in the proper units. The parameters m and Cm must refer to static cells in a separate area of the worksheet that will hold these values. For now, just enter “1” as a guess into both of these cells.
7. Make a 5th column (Column E) and enter a formula to calculate the square of the difference between Columns C and D. For example, the formula entered into cell **E2** would be “**=(C2-D2)^2**”. These values are thus the squares of the errors between the experimental fraction and the theoretical fraction (based on the m and Cm values guessed). If all of the data points aligned perfectly with the curve given by the model equation, all of these errors would be zero. The further the data is from the curve, the more positive the squares of the errors are. Give this column a heading name such as “Squared Errors”.
8. In a nearby cell, enter a formula to calculate the sum of all of the squared errors in Column E. This will give a non-negative result that measures the overall deviation of the data from the model equation. Changing m and Cm will change the theoretical fractions calculated in Column D, and thus this sum.
9. Plot both the experimental fraction (Column C) and the theoretical fraction (Column D) as a function of urea concentration (Column A) *on the same graph*. Give the curves different styles or colors to distinguish them and see how well they overlap. Since you have only estimated values for m and Cm, they will probably be very different, and your theoretical curve may even look like a straight line.
10. Invoke the Solver add-on (usually found on the *Data* tab in Excel) as described in the *J. Chem. Ed.* article and use it to minimize the sum of the squares of the errors by varying the parameters m and Cm. In the Solver window, in the field for “Set Objective” you will enter the cell containing the sum of the squared errors, and in the field that says “By Changing Variable Cells” enter the cells containing the values of m and Cm. Then make sure choose the “Min” radio button and click “OK”. The program will run for a few seconds and then a window will open up. ***Make sure you read this information window, as it will tell you if Solver was able to find a solution or not****.* If it works, you can click “OK” and it will replace the values of *m* and *Cm* in your spreadsheet with the new, optimal values it has calculated. If you just close the window, it will throw the solution away and you’ll only see your original guess again.
11. Try a few different initial guesses for m and Cm and make sure that Solver converges to the same results (or very, very close) every time. Make a note of the values for m and Cm found, and look at how well the two curves you made in the previous step now overlap.

## Part B – FRET Efficiency and Kinetics

In Part A we only needed to assume that the cyt-c protein molecules exist in an equilibrium between a folded (quenched) state and an unfolded (fluorescent) state. In this part we will examine the precise mechanism for the fluorescence quenching that occurs in the folded state. The mechanism is known as Förster Resonance Energy Transfer, or FRET, and it has a firm theoretical basis that dates back almost a century. In many ways, FRET is a phenomenon which occurs at the boundary between classical and quantum physics. In basic terms, however, it can be pictured as resonant electromagnetic coupling between two closely spaced oscillating dipoles – a donor and an acceptor. In this experiment, tryptophan act as the donor and heme as the acceptor.

The efficiency of FRET between a donor and an acceptor depends on the characteristics of the two molecules, and specifically on their electronic energy levels in the form of the emission and absorption spectra. Since FRET is a way for an excited state donor to lose energy or deactivate, it competes with the other major deactivation pathway, fluorescence emission. While the rate of fluorescence is essentially constant for a given donor, the rate of FRET increases as the donor-acceptor distance decreases. Thus, we can define the “Förster distance”, R0, as the distance at which the rate of FRET is equal to the rate of fluorescence. When the donor and acceptor are further apart than R0, the donor decays by normal fluorescence and little to no FRET occurs. When they are closer than R0, FRET dominates and we see a decrease in the observed donor fluorescence. The expression for the Förster distance is Eq. 2 in your manual, but we will use the version shown below, because it allows us to use more convenient units in the calculation of the overlap integral, JDA.

|  |  |  |
| --- | --- | --- |
|  |  | ( 2A ) |

Note that the factor in front is raised to a different power to account for the change in units. You might also notice that the “(λ)” argument of JDA has been eliminated, because it gives the incorrect impression that the overlap integral somehow depends on the wavelength. In reality, you are integrating over a range of wavelengths and ending up with a single number. The equation for the overlap integral is the same as in the manual, as shown below.

|  |  |  |
| --- | --- | --- |
|  |  | ( 3 ) |

The “(λ)” arguments have been left in for FD and εA because they are, in fact, wavelength dependent quantities. In Eq. 3, FD is the fluorescence of the donor molecule (tryptophan) by itself, and εA is the molar extinction coefficient of the acceptor (heme) by itself. **FD has arbitrary units that are cancelled out by the normalization factor in the denominator, εA has typical units of (M∙cm)-1, and λ has units of nanometers. When these units are used for JDA, the value of R0 obtained with Eq. 2A will be in angstroms (Å).**

### Calculating JDA

In order to calculate JDA, we will approximate the integrals in Eq. 3 as Riemann sums. Recall that an integral is the area under the curve defined by the integrand, and can be approximated by splitting it up into rectangles of width dλ and height equal to the value of the integrand at that point. The way to do this depends on the way the data was obtained on the spectrometers, but a brief outline is given here.

1. In any spreadsheet program, open the **fluorescence data for the NATA solution** that you obtained, so that you have two columns – Column A is the wavelength in nm, and Column B is the fluorescence intensity in arbitrary units. This data should range from about 305 nm to about 505, and will be used as FD.
2. Open the **corrected absorbance data for the cyt-c solution with 10.0 M urea**. Since the absorbance of cyt-c in the visible region is solely due to the heme cofactor, we can use this data to obtain εA, the acceptor absorbance.
3. The NATA fluorescence and cyt-c absorbance must be combined into one spreadsheet in such a way that the data for corresponding wavelengths match up. Many times the FD data will go from low wavelength to high, while the absorbance data will be reversed. There may also be differences in the interval between the data points for the two sets. **Do not simply cut and paste the data from one sheet into the other without carefully checking that they match up.** Any wavelengths that have only FD or only absorbance, but not both, should be deleted. When you are done combining them, you should have 3 columns: wavelength, FD, and “cyt-c absorbance”.
4. To convert from the cyt-c absorbance to the actual extinction coefficient (εA), you will need to find the absorbance of your sample at 410 nm. For now we will call that specific value A410. Recall from the Beer-Lambert law that , where the pathlength has been left out because we used 1 cm cuvettes. Thus, to find the extinction coefficient at any particular wavelength, you must divide the absorbance at that wavelength by the molar concentration of the solution. In order to find the exact concentration of the solution, we can use the known extinction coefficient at 410 nm, and rearrange the Beer-Lambert law to get the expression . Thus, the overall expression for the extinction coefficient is . Make a new column (Column D) labelled “extinction coefficient” and enter that formula, referring to the corresponding absorbance value and putting in the value of A410 that you found for your solution. For example, if you found A410 = 0.00123, then cell **D5** should hold the formula “**=(C5/0.00123)\*105000**”.
5. For the numerator, the integrand is FD(λ)∙εA(λ)∙λ4, which corresponds to the “height” of the rectangles used for the Riemann sum. This height must be multiplied by the “width”, dλ, to find the area of the rectangle at each value of λ. You should examine your wavelength column to determine the interval between data points. Usually the UV-Vis spectrometer is set to take a measurement every 1 nm, so dλ = 1. If not, determine what the correct value of dλ is for your data set. In a 5th column (Column E), enter a formula for the value of the integrand multiplied by dλ. For instance, cell **E23** should have a formula like “**=B23\*D23\*(A23^4)\*1**”. Obviously, if your value of dλ is 1, it can be left out, but otherwise it must be included. The sum of this column is the numerator.
6. For the denominator, the integrand is simply FD(λ), so in the 6th column (Column F), enter a formula for the product of column B (*i.e.* FD) and the value of dλ that you found. The sum of this column is the denominator. Again, if dλ = 1, this is unnecessary and you can just directly sum column B.
7. Divide the sum of column E by the sum of column F (or B) to find JDA.

Now you can calculate R0 using your value of JDA and Eq. 2A above, along with the following constants (from the manual): κ2 = ⅔, n = 1.4, and ΦD = 0.13. This number is the Förster distance in angstroms, as long as you keep the units used above. Keep in mind that this distance was derived without any reference to the details of the cyt-c protein – it is characteristic of the tryptophan-heme pair.

## Part C – Structural Analysis of Cyt-C

Now that we know R0, we essentially have a nanoscale “ruler” that we can use to measure the actual distance between the tryptophan and the heme embedded in a cyt-c protein. For instance, if we found that the fluorescence of the folded protein is exactly half the fluorescence of the unfolded protein, we would know that the two molecules are exactly R0 angstroms apart in the folded structure, because that is the point where the FRET efficiency is 50%. Equation 4 in the manual tells us how the FRET efficiency depends on the tryptophan-heme distance, r, and the Forster distance, R0.

|  |  |  |
| --- | --- | --- |
|  |  | ( 4 ) |

It should be clear that when r = R0, the efficiency is 0.5 or 50%, which is how we defined it. In addition, you can see that when r << R0, E ≈ 1, and when r >> R0, E ≈ 0.

Spectroscopically, the FRET efficiency can be related to the fluorescence intensity of the donor in the presence of the acceptor (FDA), relative to the intrinsic fluorescence of the donor when there is no acceptor (FD). The expression for this is Eq. 5 in your manual.

|  |  |  |
| --- | --- | --- |
|  |  | ( 5 ) |

Eq. 5 can be used to see how the FRET efficiency varies with the urea concentration through FDA. The FDA term is the corrected, normalized fluorescence intensity (at 350 nm) of the solutions we made in Part A, and it will therefore depend on the urea concentration. On the other hand, FD will be the fluorescence intensity of the 10.0 μM NATA solution that you obtained in Part B, which will act as a model for tryptophan fluorescence in the absence of a heme acceptor.

It is also instructive to find how the donor-acceptor distance, r, changes with the urea concentration for our protein. To do that, Eq. 4 must be inverted to find r as a function of E. Then we can put in the values of E calculated from Eq. 5 for each concentration to obtain a plot of r *vs.* C.

The data analysis for this part is outlined below.

1. First generate a graph of E *vs.* r according to Eq. 4. This can be done simply in Excel. Open a new worksheet and enter a series of r values between 1 and 100 in the first column. In the second column enter a formula corresponding to Eq. 4. The formula will include the value of R0 that you calculated in Part B, and a reference to the current r value in the first column. Plot this data with the first column (r) as the x-axis, and the second column (E) as the y-axis. Format the data series so that the data markers are hidden but connected by a smooth line with a width of 1 pt or less. This way you’ll have a smooth curve instead of a collection of points.
2. Now we will want to plot the efficiency as a function of urea concentration. On a new worksheet, in the same file, enter the experimental urea concentrations in the first column (Column A). These values should range from 0.0 to 10.0 in 0.5 M increments, as before.
3. In the next column (Column B) enter the corrected, normalized, fluorescence intensity of each solution at 350 nm. This column will be FDA.
4. In the next column (Column C) enter a formula for the efficiency corresponding to Eq. 5. The formula will refer to Column B for the FDA value. For the value of FD, you will need to use the fluorescence intensity of the NATA solution at 350 nm, divided by the concentration of that solution. Plot E (Column C) vs. C (Column A). The efficiency should be the y-axis and the urea concentration should be the x-axis.
5. Invert Eq. 4 to find r as a function of E. Enter this equation as a formula into the 4th column (Column D). Wherever the equation includes an E, have the formula refer to the corresponding datum in Column C. You can now graph r (Column D) as the y-axis *vs.* urea concentration (Column A) as the x-axis on a new chart.
6. On the same graph of E *vs.* r that you made in step 1, add a new data series that plots the experimental E (Column C) *vs.* the calculated r (Column D). Format this new data series so that there is no connecting line, but the data markers are clearly visible. You should see that they fall perfectly along the existing line from step 1.

## Lab Report

Your lab report should consist of the following parts:

* **Title, Author, Lab Partner and Date**
* **Introduction and Objective** – A paragraph describing what we hope to find in this experiment, and how. Make sure to cover all three parts.
* **Experimental Procedure** – This should be an outline of the procedure, written out in paragraph form, **NOT** as a list of steps. Describe how you prepared the solutions and ran the spectrometers. Give the make and model for any major instruments you used, as well as any important settings. Don’t put any complex calculations in the main text; save them for the appendix. **DO NOT** copy directly from the lab manual.
* **Results & Discussion** – This should consist of the following:
  + **Part A** – Include graphs of the experimental and theoretical unfolded fractions, *f*, plotted together as a function of the urea concentration, C. Along with these graphs, give the values for m and Cm that you calculated using Solver. Make sure the graphs are presentable and have appropriate axes, labels, units, etc. Describe the data analysis, including how you corrected and normalized the spectra, calculated the concentrations, calculated *f*, and fit the curves using Solver to find m and Cm. Give relevant equations that pertain to calculations or graphs discussed.
  + **Part B** – Include a graph of the NATA fluorescence and the cyt-c absorbance together. The y-axis scale doesn’t really matter, so you will want to adjust the data by a constant multiple so that they have comparable heights and can both be seen. An easy way to do this is to divide each spectrum by its maximum value. Give the values for JDA and R0 that you calculated, along with the equation used. Make sure you include units.
  + **Part C** – Give the equation that you derived for r as a function of E in step 5. Describe how you calculated the efficiency E using Eq. 5. Include your graph of E *vs.* r, which should have both the smooth line from step 1, and the individual data points from step 6. What is the value of r for which the efficiency drops to approximately 0.5? Also include the graph of E *vs.* C from step 4. How does the efficiency change as you add urea? What is the efficiency when C = 0 M? What is the efficiency when C = 10.0 M? What is the urea concentration when the efficiency is halfway between these two extremes? Does that agree with your other results from Part A? Finally, include the graph of r *vs.* C from step 5. What is the value of r in the folded protein, *i.e.* when C = 0 M? What is the value of r in the unfolded protein, when C = 10.0 M? Are these values within the range where FRET can provide accurate distance measurements? Do those distances agree with the literature concerning the structure of Cytochrome C?
  + All of your graphs should have properly labelled axes with units and a caption briefly describing what they show.
  + Include a couple of paragraphs describing your results, your calculated values (m, Cm, R0, etc.) *with proper units*, and the graphs you have made.
  + What was the assumption we made concerning the folding and unfolding in Part A? Is that assumption consistent with the analysis of the intramolecular distances in Part C? Why or why not?
  + Identify what you believe to be the most significant source of error for each part (A, B, and C). Try to give a quantitative estimate of that error, and discuss the effect it has on the uncertainty of your final answers.
* **Conclusion**
* **References**
* **Appendix** – At the very end of your report, include examples of any calculations that you did by hand. Provide digital copies of the Excel (or other) files that you used to generate your graphs.