Promoter Strength Prediction Model

iGEM Costa Rica

October 20, 2021

1 Introduction

Our main motivation to attempt to do this model was because we couldn´t find in literature the promoter strength of Pnrd, which plays an essential role in our modeling. However, data about its fluorescence is available, so we figured out a way in which we can translate this fluorescence into real promoter strength, which will be applicable to all promoters.

2 Mathematical derivation of the technique

Following the central dogma of molecular biology, we can write down the following pair of ODE´s, describing the process of synthesis and degradation of a fluorescent protein:

$$\frac{d[mRNA]}{dt} = K - \gamma_1 * [mRNA] \tag{1}$$

$$\frac{d[FP]}{dt} = \alpha * [mRNA] - \gamma_2 * [FP] \tag{2}$$

Where K is the promoter strength, γ_1 the mRNA degradation constant, α the protein synthesizing constant and γ_2 the protein degradation constant.

Solving equation 1:

$$\frac{1}{K-\gamma_1*[mRNA]} * d[mRNA] = dt$$

Integrating on both sides, and setting $u = K-\gamma_1*[mRNA]$ and $\frac{du}{dt} = -\gamma_1 * \frac{d[mRNA]}{dt}$, we have that:

$\frac{-1}{\gamma_1} * ln(K - \gamma_1 * [mRNA]) = t + c$

$ln(K - \gamma_1 * [mRNA]) = -\gamma_1 * t - \gamma_1 * c$

Taking the exponential function on both sides:

$K - \gamma_1 * [mRNA] = e^{-\gamma_1 * t - \gamma_1 * c}$

$K - \gamma_1 * [mRNA] = e^{-\gamma_1 * t} * c_1$

By mathematical clearance we get that:

$$mRNA(t) = \frac{-e^{-\gamma_1 * t}}{\gamma_1} * c_1 + \frac{K}{\gamma_1} \tag{3}$$

Plugging in equation 3 in 2 and solving, we have that:

$$\frac{d[FP]}{dt} + \gamma_2 * [FP] = \frac{-\alpha * c_1}{\gamma_1} * e^{-\gamma_1 * t} + \frac{\alpha * K}{\gamma_1} \tag{4}$$

Which is a first order linear differential equation.

It can be derived and it is known that for this type of differential equation the solution can be found explicitly. Given the generic linear ODE: $\frac{dy}{dt} + P(t) * y = Q(t)$, the solution comes by: $y = e^{-\int P(t)dt} * [\int Q(t) * e^{\int P(t)dt} dt + c_2]$

Using this fact we have that:

$FP(t) = e^{-\int \gamma_2 dt} * [\int (\frac{-\alpha * c_1}{\gamma_1} * e^{-\gamma_1 * t} + \frac{\alpha * K}{\gamma_1}) * e^{\int \gamma_2 dt} dt + c_2]$

$FP(t) = e^{-\gamma_2 * t} * [\int \frac{-\alpha * c_1}{\gamma_1} * e^{(\gamma_2 - \gamma_1) * t} + \frac{\alpha * K}{\gamma_1} * e^{\gamma_2 * t} dt + c_2]$

$FP(t) = e^{-\gamma_2 * t} * [\frac{-\alpha * c_1}{(\gamma_2 - \gamma_1) * \gamma_1} * e^{(\gamma_2 - \gamma_1) * t} + \frac{\alpha * K}{\gamma_1 * \gamma_2} * e^{\gamma_2 * t} + c_2]$

By distributing, we finally found the expression for FP:

$$FP(t) = \frac{-\alpha * c_1}{(\gamma_2 - \gamma_1) * \gamma_1} * e^{-\gamma_1 * t} + c_2 * e^{-\gamma_2 * t} + \frac{\alpha * K}{\gamma_1 * \gamma_2} \tag{5}$$

So we came up with a general formula that relates the fluorescent protein being expressed in terms of the original parameters of the ODE´s.

From equation 5, we can see that it consists of two decreasing exponential functions and one constant term, therefore, we can take the limit when the time tends to infinity to predict what will be the maximum fluorescence. The two decreasing exponential functions will cancel out as they tend to zero, giving that the maximum fluorescence is completely determined by the constant term.

$$FP_{max}(t) = \frac{\alpha * K}{\gamma_1 * \gamma_2} \tag{6}$$

We chose 5 different points (time,fluorescence) to calculate promoter strength. As we can see from equation 5 , we only need 3 points to calculate the 3 coefficients c_1, c_2, K of the formula, with special

interest in K which is the promoter strength. To check reproducibility and statistical significance, we made all the possible combinations of 3 points in a set of 5, which is determined by the binomial coefficient:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

3 Results and Discussion

Here we present the process of obtaining and validating the data with our model.

3.1 Robustness of the method

For the pBAD promoter, we used two plots from igem parts, one of them is seen in figure 1. The link to iGEM parts page is: $http://parts.igem.org/Part:BBa_K584000$

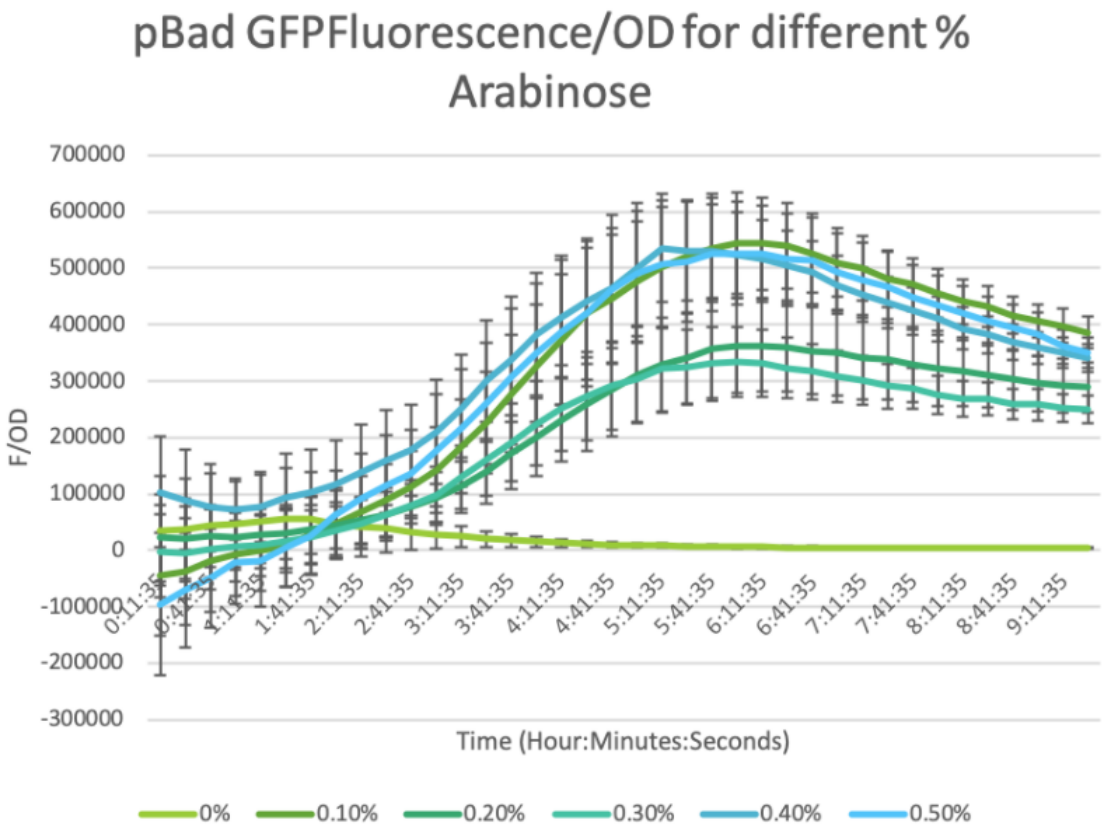

Figure 1: Measurement of arabinose Promoter Activity.

From all those points we used the graph corresponding to 0.10%, and extracted the points described in table 1.

Time(h,m,s)	Fluorescence
2:41:35	90000
3:41:35	220000
4:41:35	420000
5:41:35	520000
6:11:35	540000

Table 1: Data used for measure pBAD strength. Extracted from figure 1,using 0.10% arabinose.

The other plot is also from igem parts, as seen in figure 2. The link to iGEM parts page is: $http://parts.igem.org/Part:BBa_K808000$

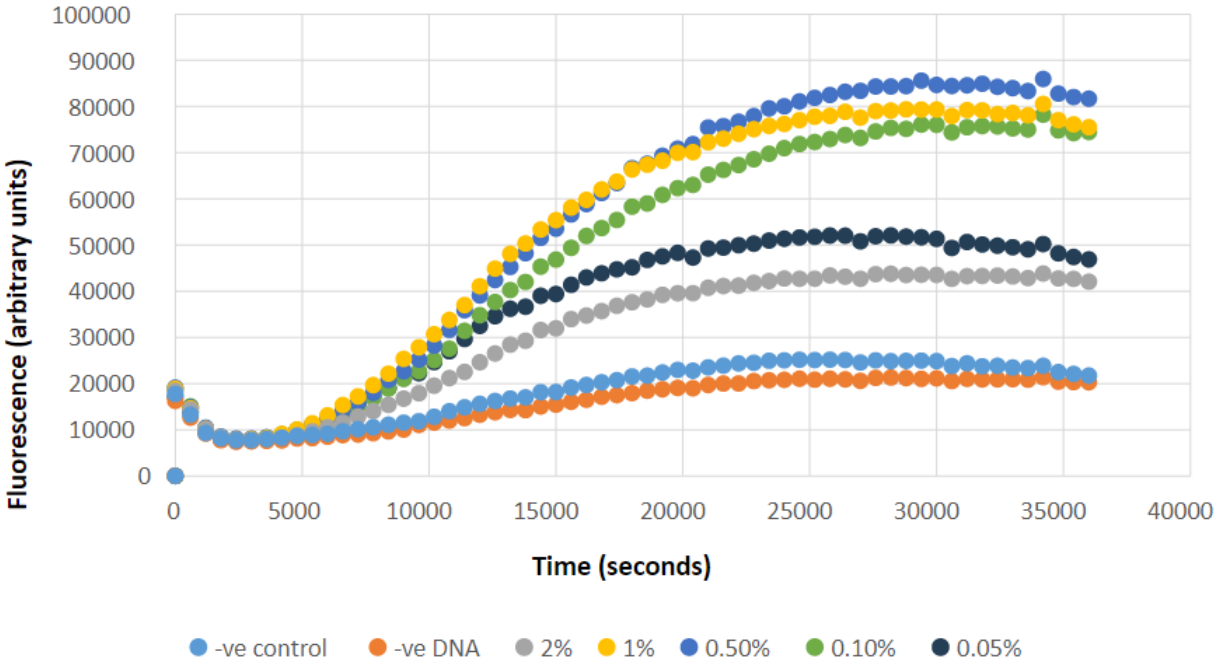

Figure 2: Measurement of arabinose Promoter Activity.

From all those points we used the graph corresponding to 0.10%, and extracted the points described in table 2.

Time(s)	Fluorescence
10000	25000
15000	48000
20000	62000
25000	71000
30000	77000

Table 2: Data used for measure pBAD strength. Extracted from figure 2, using 0.10% arabinose.

For the Pnrd promoter, we also used a plot from igem parts, as seen in figure 3. The link to iGEM parts page is: $http://parts.igem.org/Part:BBa_K2070012$

Time(minutes)	Fluorescence
0	25000
18	26500
45	27500
63	28500
81	31000

Table 3: Data used for measure Pnrd strength. Extracted from figure 3.

Now we report the 10 scenarios for which we calculated promoter strength with our method, corresponding to all the different ways of choosing 3 combinations out of 5 data points for each promoter, as well as the mean and standard deviation.

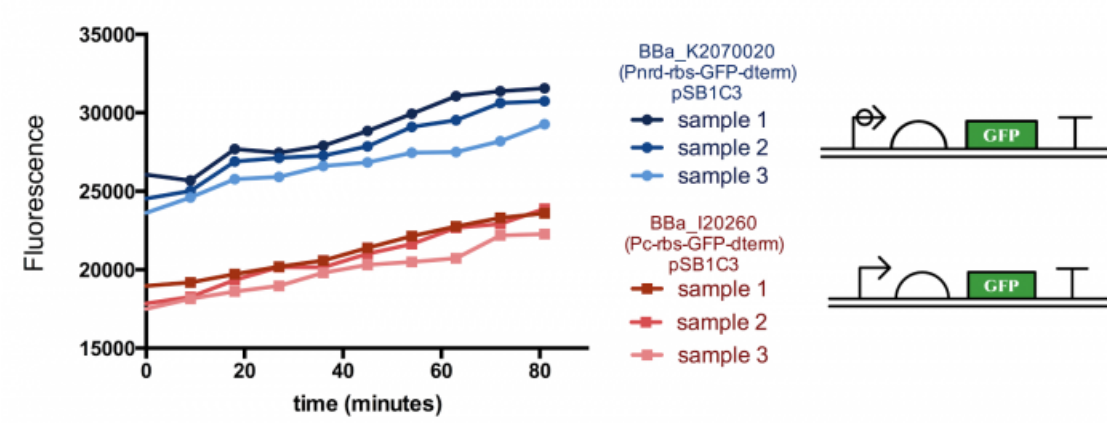

Figure 3: Measurement of nrd Promoter Activity

Combination	Value(protein/s)
1	0.0348
2	0.0402
3	0.0414
4	0.0410
5	0.0418
6	0.0422
7	0.0410
8	0.0418
9	0.0422
10	0.0422
Mean	0.040858
Standard Deviation	0.002216

Table 4: Results of the value of promoter strength for pBAD, for all the possible 3-combinations for 5 data points, data from table 1.

Combination	Value(protein/s)
1	0.0048
2	0.0054
3	0.0059
4	0.0055
5	0.0059
6	0.0059
7	0.0055
8	0.0059
9	0.0059
10	0.0059
Mean	0.005654
Standard Deviation	0.000361

Table 5: Results of the value of promoter strength for pBAD, for all the possible 3-combinations for 5 data points, data from table 2.

Combination	Value(protein/s)
1	0.0022
2	0.0022
3	0.0024
4	0.0023
5	0.0025
6	0.0026
7	0.0023
8	0.0025
9	0.0026
10	0.0027
Mean	0.002424
Standard Deviation	0.000181

Table 6: Results of the value of promoter strength for Pnrd, for all the possible 3-combinations for 5 data points.

As we can see from tables 4,5 and 6, very low standard deviations were obtained, at least one order of magnitude lower than the mean, which means that the promoter strength obtained through all the different combinations was very homogeneous, suggesting that it would be okay if we only use 3 points to calculate the promoter strength. It also make us think that the method is robust.

3.2 Relative Strength Comparison

We already proved in last section the robustness of the method, meaning that indeed we can use only 3 data points to calculate promoter strength, without fear of a bad generalization. Now we are going to try to prove its applicability by comparing the outputs given by the method with real lab data.

As a first approach, we built upon the work of iGEM Tacoma team of 2019. They collected characterization data on the relative RFP expression rates of Anderson Promoters *Bba_J*23100, *Bba_J*23101, *Bba_J*23102, *Bba_J*23105, *Bba_J*23106, and *Bba_J*23112 in the RFP expression plasmid *Bba_J*61002 (samples provided in the 2019 iGEM Distribution Kit).

Their protocols were adapted from the iGEM 2018 Interlab protocol.

$AFU AVG \pm SD$	0 hours	3 hours	5 hours
J23100	116.5 ± 9.2	159.4 ± 9.2	210.6 ± 6.0
J23102	97 ± 6.9	151.1 ± 3.4	157.8 ± 15.5
J23101	103 ± 5.9	143.0 ± 3.2	178.7 ± 15.9
J23106	80 ± 8.2	118.8 ± 9.5	131.9 ± 8.6
J23105	82 ± 11.6	101.2 ± 7.3	123.8 ± 23.4
J23112	67 ± 4.0	86.3 ± 5.1	83.0 ± 8.7
Neg	57.4 ± 2.4	65.3 ± 5.5	80.2 ± 3.3

Table 7: Measured average arbitrary fluorescence units of Anderson promoters by iGEM Tacoma team 2019.

$OD660 AVG \pm SD$	0 hours	3 hours	5 hours
J23100	0.006 ± 0.006	0.123 ± 0.018	0.294 ± 0.011
J23102	0.007 ± 0.007	0.152 ± 0.015	0.354 ± 0.018
J23101	0.009 ± 0.006	0.206 ± 0.016	0.395 ± 0.028
J23106	0.008 ± 0.004	0.198 ± 0.010	0.376 ± 0.010
J23105	0.017 ± 0.005	0.250 ± 0.013	0.436 ± 0.010
J23112	0.0198 ± 0.009	0.247 ± 0.012	0.415 ± 0.013
Neg	-0.009 ± 0.004	0.007 ± 0.012	-0.011 ± 0.003

Table 8: Optical Density of cultures measured at Abs660nm by iGEM Tacoma team 2019.

AFU/OD	0 hours	3 hours	5 hours
J23100	19416.67	1295.93	716.33
J23102	13857.14	994.08	445.76
J23101	11444.44	694.17	452.41
J23106	10000	600	350.8
J23105	4823.53	404.8	283.94
J23112	3383.84	349.39	200

Table 9: Arbitrary Fluorescent Units normalized to Optical Density.

As the original characterization of the Anderson promoters and this study performed by the Tacoma iGEM team were done using the red fluorescent protein(RFP), we adapted our model with the new parameters corresponding to this protein, as previously we just took into account the green fluorescent protein for the method.
We found in literature different values for the parameters of our equations described in equation 1 and 2, reported by the iGEM teams of Nanjing-China 2017($http://2017.igem.org/Team:Nanjing-China/Model/ch2o$) and Alma-USA 2019($https://2019.igem.org/Team:Alma/Model$):

Meaning	Symbol	Team Nanjing-China(s^{-1})	Team Alma-USA(s^{-1})
mRNA RFP degradation rate	γ_1	0.0022	0.0337
RFP protein degradation rate	γ_2	0.0012	0.0048
RFP translation rate*	α	0.072	0.00077

Table 10: Parameters values for RFP. * Equivalently the translation rate was found to be 17.1 amino acids per second.

We present the results obtained of applying our method for predicting promoter strength applied to data from table 9, with the different combinations of the parameters of mRNA degradation rate, protein degradation rate and translation rate for the RFP, described in table 10.

Promoter	Promoter Strength(protein/s)	Relative Strength
J23100	$1.106 * 10^{-4}$	1
J23102	$6.880 * 10^{-5}$	0.62
J23101	$6.984 * 10^{-5}$	0.63
J23106	$5.403 * 10^{-5}$	0.49
J23105	$4.383 * 10^{-5}$	0.40
J23112	$3.087 * 10^{-5}$	0.28

Table 11: Predicted promoted strength using Nanjing parameters, with RFP translation rate 17.1 ntd/s

Promoter	Promoter Strength(protein/s)	Relative Strength
J23100	0.0263	1
J23102	0.0163	0.62
J23101	0.0166	0.63
J23106	0.0128	0.49
J23105	0.0104	0.40
J23112	0.0073	0.28

Table 12: Predicted promoted strength using Nanjing parameters, with RFP translation rate $0.072s^{-1}$

Promoter	Promoter Strength(protein/s)	Relative Strength
J23100	$6.831 * 10^{-3}$	1
J23102	$4.251 * 10^{-3}$	0.62
J23101	$4.314 * 10^{-3}$	0.63
J23106	$3.338 * 10^{-3}$	0.49
J23105	$2.708 * 10^{-3}$	0.40
J23112	$1.907 * 10^{-3}$	0.28

Table 13: Predicted promoted strength using Alma parameters, with RFP translation rate 17.1 ntd/s

Promoter	Promoter Strength(protein/s)	Relative Strength
J23100	15.170	1
J23102	9.440	0.62
J23101	9.581	0.63
J23106	7.412	0.49
J23105	6.013	0.40
J23112	4.236	0.28

Table 14: Predicted promoted strength using Alma parameters, with RFP translation rate $0.00077s^{-1}$

As seen in tables 11, 12, 13 and 14, even though we used different parameters for the original equations, surprisingly the relative strength of the promoters when normalizing by the promoter $J23100$ were all the same in the different scenarios.

	Grown to saturation
J23100	1.00
J23102	0.86
J23101	0.70
J23106	0.47
J23105	0.24
J23112	0.00

Table 15: Summarized Anderson lab data for comparison, analyzed relative to BBa_J23100 expression.

It is important to notice that(as seen in table 16) the comparison against BBa_J23100 is consistent with the original Anderson characterization, except for position 2 and 3, where it is supposed that

Promoter	Anderson Lab	Relative Strength Predicted	%Error
J23100	1	1	0
J23102	0.86	0.62	27.9
J23101	0.70	0.63	10
J23106	0.47	0.49	4.3
J23105	0.24	0.40	66.7
J23112	0.00	0.28	>100

Table 16: Comparison between relative strength of the promoters, been the reference J23100.

the promoter $J23102$ is stronger than $J23101$. A possible error that could explain this difference is that the OD660 values were still increasing from 3 hours and 5 hours, which indicates the cultures were still growing. Allowing the cultures to grow further would allow a better determination regarding the difference in promoter strength of BBa_J23101 and BBa_J23102, as Anderson promoters were measured until they reached saturation.

Also the relative values obtained by our method present a good consensus with the 3 data points in time the iGEM Tacoma team measured.

3.3 Direct Comparison

We tried to make a direct comparison between values of promoter strength predicted by our model and the corresponding reported literature values, however, we have failed until now mainly because of lack of biological units on fluorescence data(the vast majority is reported in arbitrary units) and the great variance that exists between labs because of lack of standardized methods of measurement and calibration.

We have tried to compare the values obtained by our method of promoter J23101 and J23106(precisely were the ones which presented the least percentage relative error, as seen in table 16) against the values in biological units established in the iGEM Interlab study, which also take into account these two promoters. We have data in arbitrary fluorescence units of these promoters, as well as their optical density, however we have failed to convert that into the proposed units of Molecules of Equivalent Fluorescein(MEFL)/cell in order to be able to compare them.

Other sources of possible error in our method and things to improve could be:

- Our equation describes fluorescent protein, and data available and used to calculate promoter strength is fluorescence, so there is a slight mismatch. It is known that the relation between the fluorescent protein and its fluorescence is linear, however this has to be measured and computed by the lab that performed the measurements.
- It is important to take into account and report the RBS strength of the RBS used in the experiments.
- It is very important to reach an agreement about what biological unit to use to report promoter strength. We propose the unit of Molecules of Equivalent Fluorescein(MEFL)/cell, as proposed in the iGEM Interlab study.
- Calibration methods and materials should be reported along with fluorescence data, because it describes the conditions in which the experiment was performed and this data could be useful in the comparison process.

4 Software Tool

Lastly, we created a software tool to translate fluorescence into promoter strength in order that other iGEM teams could use it, as seen in figure 4. The app can receive data in GFP or RFP, but we plan to include other fluorescent proteins.

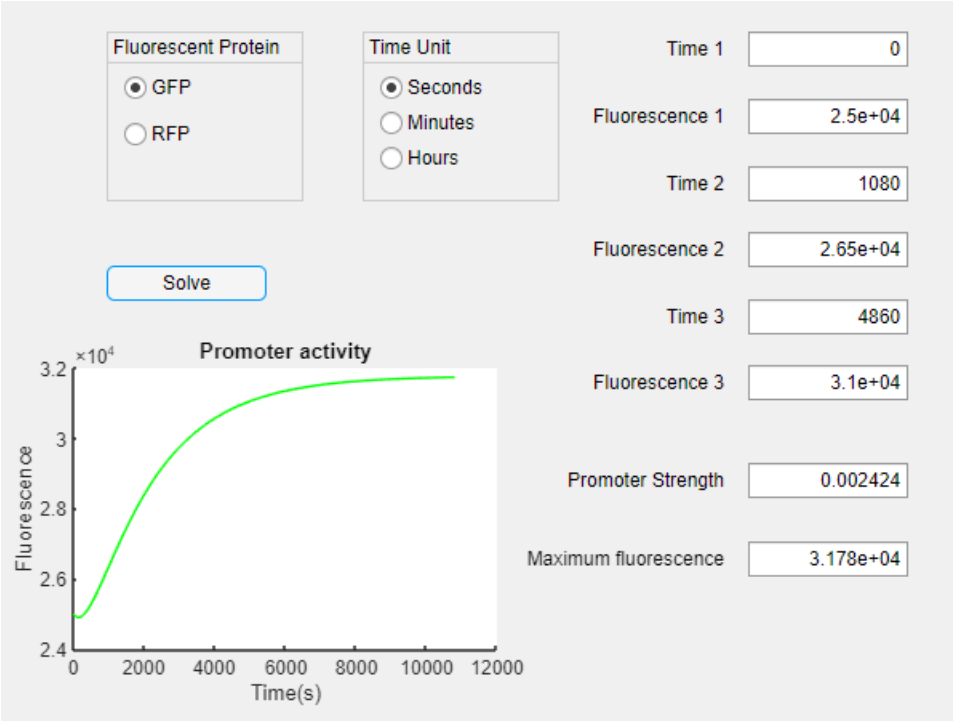

Figure 4: Example of the app working, designed with Matlab App Designer.