Modeling the Central Dogma in *E. coli*: A Systems Biology Approach

1. Introduction

"Tout ce qui est vrai pour le Colibacille est vrai pour l'éléphant" (All that is true for *E.coli* is also true for the elephant) — this phrase by Jacques Monod beautifully captures the essence of molecular biology: the fundamental processes governing life are (quite) universal. [1] One of the main process is known as 'central dogma' of biology.

The central dogma of molecular biology, first coined by Francis Crick in 1957, describes the flow of genetic information within a biological system: DNA is transcribed into RNA, which is then translated into protein. [2] This fundamental process occurs in all living organisms, but the details of how it is carried out can differ significantly between prokaryotic and eukaryotic cells.

In prokaryotes (such as bacteria), the central dogma operates in a more streamlined and contemporaneous manner. The absence of a membrane-bound nucleus allows for a direct and simultaneous process of transcription and translation. [3] As soon as the messenger RNA (mRNA) is transcribed from DNA in the cytoplasm, ribosomes can begin translating it into protein, all occurring in the same cellular space.

In contrast, eukaryotic cells, which include plants, animals, fungi, and protists, have a more complex arrangement. Transcription occurs in the nucleus, where the DNA resides, and the resulting mRNA must then undergo various processing steps (such as splicing, capping, and polyadenylation) before being exported to the cytoplasm for translation. [2]

To create a simple model, I'll focus on a very studied bacterium, the *E. coli*. [1,3]

2. Absolute Numbers in *E. coli*

The processes of transcription and translation are carried out by specialized molecular machines: RNA polymerases (RNAP) and ribosomes. [4] Both of these machines play a crucial role in gene expression, but their number and speed of action can greatly impact cellular functions. RNA polymerases are enzymes responsible for transcribing genetic information from DNA into RNA. In *E. coli*, the total number of **RNAP** molecules typically ranges from 10^3 to 10^4 per cell. [5] These polymerases work at a speed of about **40-80 nucleotides per second**.[3,5] Ribosomes, on the other hand, are the molecular machines responsible for translating mRNA into proteins. Like RNA polymerases, the total number of ribosomes in an *E. coli* cell is similarly high, with estimates placing their numbers between 10^3 and 10^4. [5] The rate at which **ribosomes** synthesize proteins is approximately **20 amino acids per second**, which is equivalent to 60 nucleotides per second in terms of translation speed (1aa = 3 nt).[3,5] Noteworthy, the rates of the RNAP and ribosomes are similar. Indeed, if translation was faster than transcription, it would cause the ribosome colliding with RNA polymerases during their concurrent activity. [5] However, single-molecule microscopy has revealed that such collisions are relatively rare in practice and much the translation in *E. coli* is not immediately coupled with transcription. [5, 6]

A key factor that enhances the efficiency of translation is the presence of polyribosomes (or **polysomes**), which are multiple ribosomes translating the same mRNA molecule

simultaneously. [7] This phenomenon allows for an increased rate of protein production, as multiple copies of a protein can be synthesized from a single mRNA at the same time. [8] Polyribosomes improve cellular efficiency by maximizing the use of available mRNA templates, ensuring a rapid and coordinated response to cellular needs.

3. Limitations of Considering Absolute Numbers: The Role of Dilution

When discussing the number of molecules inside a cell, it's important to distinguish between absolute numbers and concentration. Focusing on absolute numbers can be problematic, particularly when considering the effects of dilution. As a cell grows and divides, its total molecular content is distributed between the daughter cells, leading to sudden decreases in the number of molecules per cell. These abrupt "jumps" can complicate the analysis of molecular dynamics over time. However, considering concentration rather than absolute numbers mitigates this issue. Concentration accounts for both the increase in the number of molecules and the expansion of cell volume, allowing for more gradual changes. Even after cell division, if the volume of the daughter cells is also halved, the concentration remains relatively stable, preventing sudden shifts. In dynamic conditions—such as cell growth, division, or responses to environmental stimuli—it is more accurate to model changes in molecular concentration over time rather than absolute molecule counts.

4. mRNA Time Degradation: A Comparison of *E. coli* and *H. sapiens*

The lifetime of mRNA molecules is generally much shorter than the time scale of cell biology defined by the cell division cycle. In *E. coli*, most **mRNA** molecules have **lifetimes** between **3 and 8 minutes**, while in yeast, the mean lifetime is around 20 minutes, and in human cells, it is approximately 600 minutes. [5] This trend correlates with the duration of the **cell cycle**, which is about **30 minutes** in *E. coli*, 90 minutes in yeast, and 3000 minutes in humans under rapid exponential growth conditions. [5] These findings suggest that the mRNA degradation timescale is approximately one-fifth of the cell cycle duration in these organisms. Furthermore, because mRNA degradation occurs much faster than the redistribution of molecules during cell division, the impact of cell division can often be neglected when modeling mRNA dynamics in rapidly growing cells.

5. Protein Time Degradation: A Comparison of *E. coli* and *H. sapiens*

The lifetime of proteins varies significantly between organisms and is generally much longer than that of mRNA. In *E. coli*, most proteins have half-lives ranging from a few hours to over a day, while in yeast, the median protein half-life is around 90 minutes, and in human cells, it can extend from several hours to days. [5].

Proteins can be actively degraded through specific mechanisms. In prokaryotes, degradation is primarily mediated by cytoplasmic proteases such as Lon and ClpXP, which recognize specific signals in target proteins. [9] In eukaryotes, the main degradation pathway is the ubiquitin-proteasome system, where proteins destined for degradation are tagged with ubiquitin and subsequently broken down by the proteasome. [10]
However, for particularly stable proteins, whose half-life is much longer than the cell cycle duration, degradation occurs primarily through passive dilution due to cell division. In these cases, protein concentration is regulated more by cell growth than by active degradation mechanisms.

6. The Dynamic Nature of Protein Production from a Single mRNA

The **protein-to-mRNA ratio** in *E. coli* is around **1000**, meaning that at any given time, there are roughly 1000 proteins for every mRNA in the cell. [5] However, since mRNA degrades slower than translation occurs, this high ratio cannot result from a single mRNA producing hundreds of proteins before being degraded. Instead, it suggests that gene expression is maintained through continuous transcription and translation, compensating for the rapid turnover of mRNA.
A sanity check with known values supports this idea. If mRNA degradation occurs in about 3 minutes, while the cell cycle lasts around 30 minutes, each mRNA exists for only one-tenth of a cell cycle. Given the total protein-to-mRNA ratio of 1000, this means that a **single mRNA produces** only **10 to 100** proteins before degrading. [5] This range is consistent with translation dynamics in bacteria, where ribosomes synthesize proteins at a rate of approximately 10-20 amino acids per second. Given an average protein length of 300 amino acids, a single translation event takes about 15-30 seconds. [5] Since an mRNA lasting 3 minutes can be translated multiple times by ribosomes operating in parallel (polyribosomes), it is reasonable to estimate that each mRNA generates between 10 and 100 proteins before degradation.

7. Modeling the Central Dogma in *E. coli*: A Systems Biology Approach

To construct a simple model of the central dogma in *E. coli*, we summarize the key numerical values reported in previous sections:

Parameter	Approximate Value	Units
RNAP Speed	40-80	$nt \cdot sec^{-1}$
Ribosome Speed	20	$aa \cdot sec^{-1}$
mRNA lifetime	3-8	min
Cell Cycle	30	min
Protein-mRNA ratio	1000	-
Proteins synthesized from single mRNA	10-100	-

Specifically, let's assume we want to model the central dogma of an E. coli gene such as LacZ. LacZ encodes for the enzyme β-galactosidase which plays a key role in breaking down lactose into glucose and galactose. It is part of the lac operon, a regulatory system in bacteria that controls the metabolism of lactose: when lactose is present, LacZ is expressed and produces β-galactosidase, allowing the cell to use lactose as a carbon source. [11].
The LacZ gene size is about 3075 base pairs while the protein contains 1021 amino acid residues (about one third of nucleotides as we expected). [12]
The mRNA for LacZ is typically short-lived in *E. coli*, with a half-life of around 3-8 minutes, while the β-galactosidase protein is relatively more stable, with a half-life of about 20 hours [13].

I used the approximation of 1 amino acid = 3 nucleotides and hypothesized that the number of polysomes is around 10-20.

```python
# Define parameters
length = 1024  # amino acids
max_speed_pol = 80  # nt/sec
max_speed_rib = 20  # aa/sec
'''We used the approximation of 1 amino acid = 3 nucleotides
and hypothesized that the number of polysomes is around 10-20'''
rib_per_mrna = 15  # average number of ribosomes per mRNA
mrna_lifetime = 3 * 60  # (3 min converted to sec)
cell_cycle = 30 * 60  # passive degradation. (30 min converted to sec)
```

7.1 mRNA model

A basic mathematical model describing mRNA levels in *E. coli* can be written as:

$$\dot{m} = k - \gamma m(t) \qquad (1)$$

where:

- $m(t)$ represents the number of mRNA molecules in the cell at that time.

- $\dot{m} = \frac{dm}{dt}$ represents the derivate over time of the m(t).
- k is the transcription rate which is the inverse of the time transcription. The time time transcription as the length of the mRNA molecule over the RNAP speed.
- γ is the mRNA degradation rate, which determines how quickly mRNA is degraded. It is defined as the $ln2$ over the mRNA lifetime.

At steady state m_{ss} ($\dot{m} = 0$), the equation simplifies to:

$$m_{ss} = \frac{k}{\gamma}$$

indicating that the steady-state mRNA level depends on the balance between synthesis and degradation.

The solution of the equation (Appendix A) is:

$$m(t) = \frac{k}{\gamma}(1 - e^{-\gamma t}) + m_0 e^{-\gamma t} \qquad (2)$$

Where m_0 is the initial concentration of mRNA.

Studying the function, we observe that for $t \to \infty$, the solution approaches the steady state, as expected. In the limit of $t \to 0$, the function grows asymptotically toward a straight line (as demonstrated in Appendix B). Furthermore, $1 - e^{-\gamma t}$ is a negative exponential function that increases towards 1.

This simple model captures the core dynamics of mRNA but omits several important biological factors: 1) The transcription rate k is not constant but depends on promoter strength, transcription factors, and environmental conditions; 2) Some mRNA molecules have structural elements that affect their degradation rates.
Incorporating these factors would lead to a more realistic but complex model of gene expression in *E. coli*.

7.2 Protein levels model

A second mathematical model describing the time evolution of protein levels can be written as:

$$\dot{p} = \alpha m(t) - \beta p(t) \qquad (3)$$

where:

- $p(t)$ represents the protein concentration at a time.

- $\dot{p} = \frac{dp}{dt}$ represents the derivate over the time of $p(t)$.
- α is the translation rate defined as the inverse of the time translation. The time translation is defined as the length of the protein over the ribosome speed.
- β is the protein degradation rate, which is generally much smaller than γ in *E. coli*, which is defined as the $ln2$ over the cell cycle time.
- $m(t)$ is the solution of the mRNA equation (2), meaning that protein production depends on the available mRNA at any given time.

At steady state p_{ss} ($\dot{p} = 0$):

$$p_{ss} = \frac{\alpha}{\beta}\frac{k}{\gamma} = \frac{\alpha}{\beta}m_{ss}$$

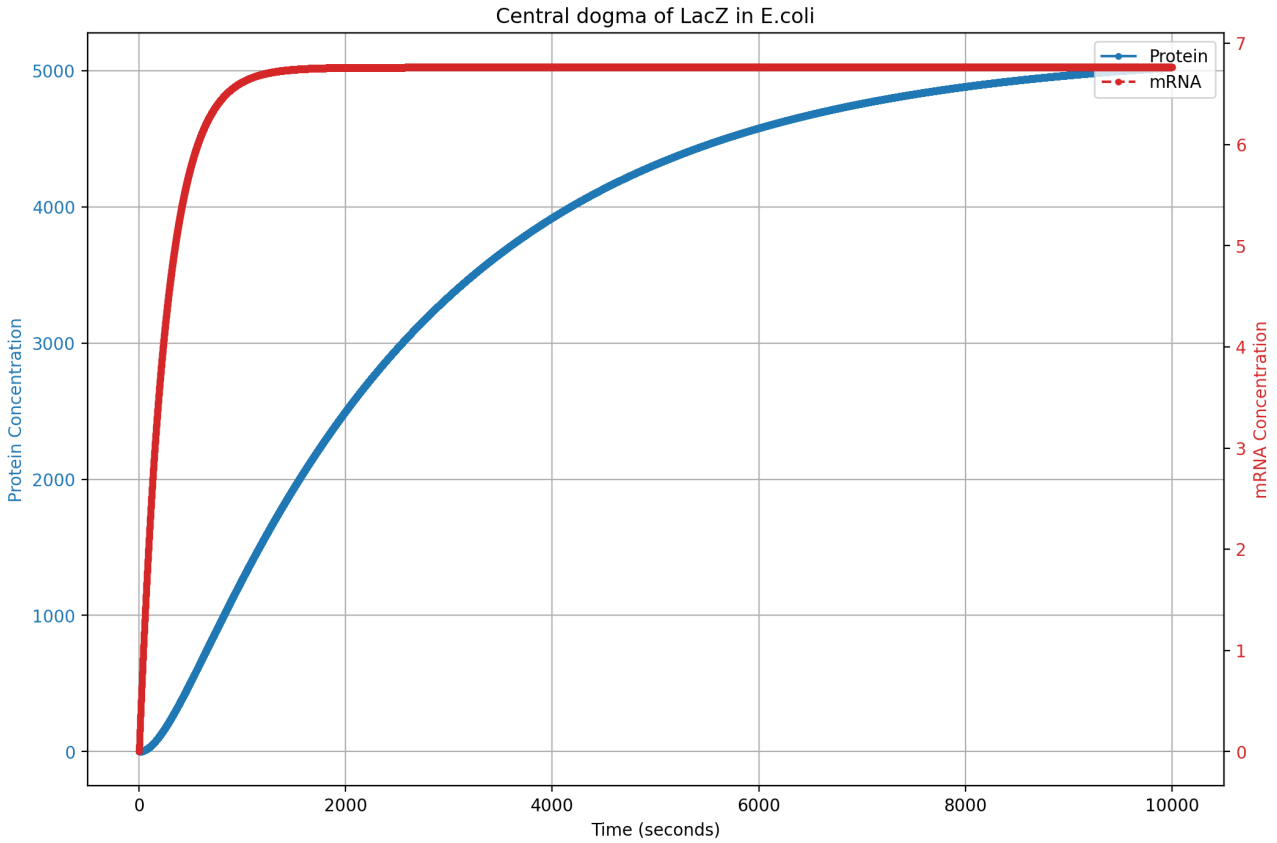
which shows that the final protein concentration depends on both transcription and translation rates, as well as mRNA and protein degradation rates.
The solution of the equation (see Appendix C) — for $m_0 = 0$ — is:

$$p(t) = \frac{\alpha \cdot k}{\gamma}\left[\frac{1}{\beta}(1 - e^{-\beta t}) - \frac{1}{\beta - \gamma}(e^{-\gamma t} - e^{-\beta t})\right] + p_0 e^{-\beta t} \qquad (4)$$

Where p_0 is the inital concentration of protein levels.
In this case, the study of the function is more complex than what was observed in the transcription model. For $t \to \infty$, the solution approaches the steady state, as expected. In the limit of $t \to 0$, the function grows asymptotically toward a parabola (as demonstrated in Appendix B).

This simple model provides a first-order approximation of gene expression but omits several biological complexities: 1) α can vary due to promoter activity, ribosome availability, external signals, no messenger RNA actions and other. 2) Actually, there is non-istananteous mRNA protein production, introducing delays. 3) The rate of degradation is influenced by the active degradation as well as the passive ones, i.e., $\beta_{real} = \beta_{active} + \beta_{passive}$

7.3 Plot the model
Below is the graph obtained from the Python code reported in “Magnati_Python_Prjct”:
It is worth noting that the y-axes have different scales to correctly display both functions, which represent mRNA concentrations (in red) and protein levels (in blue) as a function of time (measured in seconds).



To assess whether the plot holds biological significance, let's perform a sanity check on the protein-to-mRNA ratio at steady state. By dividing $\frac{p_{ss}}{m_{ss}}$, we obtain simply $\frac{\alpha}{\beta}$. With the set values, we get approximately 760, which is consistent with the previously described protein-to-mRNA ratio of around 1000.
A second sanity check can be applied to the "Proteins produced per mRNA before degradation," which can be determined by the product between the translation rate and the single mRNA lifetime (i.e., by using the ratio of $\frac{\alpha}{\gamma}$). This yields a value of approximately 76, which is consistent with the values previously discussed of about 10-100 protein per single mRNA molecules.

Finally, an interactive plot was created to allow users to easily adjust key parameters, including protein length, the maximum speeds of RNA polymerase and ribosomes, the number of ribosomes per mRNA, the mRNA degradation rate, and the cell cycle duration, all through interactive sliders (See “Magnati_Python_Prjct” file). By manipulating these sliders, users can visually explore how changes in each parameter affect gene expression dynamics, providing an intuitive and hands-on approach to studying the central dogma and gene regulation in *E. coli*.

8. Conclusion

By focusing on a minimal set of parameters, this model captures the essential dynamics of gene expression while avoiding unnecessary complexity. Despite its simplicity, it provides valuable insights into the interplay between transcription, translation, and degradation in a rapidly growing bacterial system.

Basic sanity checks indicate that the model behaves as expected under biologically relevant conditions.

Appendix A: Differential Equations and the Method of Separable Variables

Definition: A differential equation is an equation that involves an unknown function and its derivatives with respect to one or more independent variables. Formally, a differential equation has the general form:

$$F\left(t, y, \frac{dy}{dt}, \frac{d^2y}{dt^2}, \dots\right) = 0$$

We aim to derive the result obtained from equation (1). This equation is a first-order differential equation, which can be easily reduced to an initial value problem (also known as a Cauchy problem) of the form:

$$\dot{m} = f(t, m)$$
$$m(t_0) = m_0$$

where f is a real function ($f : \mathbb{R}^2 \to \mathbb{R}$) defined in a region of the plane $\mathbb{R}^2$.

Notation: We will adopt the common notational simplification of omitting the explicit dependence on *t*, writing *m=m(t)*.

A common method for solving this equation is the **method of separable variables**, which we outline below.

Definition: A differential equation is said to have separable variables if it can be written in the form:

$$\dot{m} = g(t)h(m) \qquad \text{(A1)}$$

where *g* and *h* are distinct and continuous functions of the variables *t* and *m*, respectively.

Let's define:

$$g(t) = 1,$$
$$h(m) = k - \gamma m$$

Property: To simplify the treatment, let's interpret the derivative $\dot{m} = \frac{dm}{dt}$ as a "quotient" following Leibniz notation. By dividing equation (A1) by $h(m)$ and multiplying by dt, we obtain:

$$\frac{dm}{h(m)} = g(t)dt \quad \Rightarrow \quad \frac{dm}{k - \gamma m} = dt$$

Let's now integrate both sides, choosing integration limits from an initial condition $t_0 = 0$ to an arbitrary t

$$\int_{m_0}^{m} \frac{dm}{k - \gamma m} = \int_0^t dt$$

Solving this integral, we obtain:

$$m = \frac{k}{\gamma}\left(1 - e^{-\gamma t}\right) + m_0 e^{-\gamma t}$$

Finally, for $t_0 = 0$, we can verify that:

$$m(0) = m_0$$

which satisfies the initial condition set in the Cauchy problem and completes the proof (Q.E.D.).

Appendix B: Taylor Series Expansion

Definition: The Taylor series expansion of a function *f* around a point $x_0 \in \mathbb{R}$ represents the function as the sum of a polynomial and a higher-order infinitesimal term, whose order exceeds that of the polynomial.

Theorem: *"Taylor's Formula with Peano Remainder"*. Let $n \geq 0$ and f be n-times differentiable at x_0. Then, the Taylor series expansion of f at x_0 is given by:

$$\begin{aligned}
f(x) &= Tf_{0,x_0}(x) + o(1) \\
f(x) &= Tf_{1,x_0}(x) + o(x - x_0) \\
f(x) &= Tf_{2,x_0}(x) + o((x - x_0)^2) \\
&\vdots \\
f(x) &= Tf_{n,x_0}(x) + o((x - x_0)^n), \quad \text{as } x \to x_0
\end{aligned}$$

where the Taylor polynomial of order n is given by:

$$\begin{aligned}
Tf_{n,x_0}(x) &= \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x - x_0)^i \\
&= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n
\end{aligned}$$

Remark: A Taylor series expansion centered at $x_0 = 0$ is also known as a *Maclaurin series*.

Properties: For $f(x) = e^x$, the Maclaurin series expansion of order n is given by:

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + o(x^n)$$
$$= \sum_{i=0}^{n} \frac{x^i}{i!} + o(x^n)$$

This result is particularly useful in the analysis of equations (2) and (4).

Let $m_0 = 0$, then for equation (2), we have:

$$m(t) = kt + o(t), \quad t \to t_0$$

By exploiting the first-order Maclaurin expansion, the result indicates that in the initial stages, the mRNA count grows asymptotically along a straight line passing through the origin with a slope coefficient equal to *k*.

A similar argument holds for equation (4) for $p_0 = 0$. It is easy to show that the first-order Maclaurin expansion of the function vanishes, suggesting the use of higher-order infinitesimals. Limiting ourselves to the second-order expansion, we have:

$$p = \frac{\alpha k}{2} t^2, \quad t \to t_0$$

This result indicates that near *t=0*, the function grows asymptotically as a concave parabola with the minimum value of p occurs at $t = 0$.

Appendix C: Non-Homogeneous Linear Differential Equations

Definition: A linear differential equation is an equation of the form:

$$Y' + A(t)y = B(t)$$

Remark: Note the difference from what is stated in Appendix A: in this case, the function

$$f(t, y) = -A(t)y + B(t)$$

is a first-degree polynomial in *y*, with coefficients depending on *t*.

Remath 2: If $B(t) = 0$, then the problem reduces to case (A1), and the equation is said to be homogeneous. If$B(t) \neq 0$, the equation is classified as non-homogeneous.

Property: The general solution of a non-homogeneous linear equation is given by:

$$p(t) = p_h + p_a$$

where:

- p_h is the solution of the homogeneous equation,
- p_a is the particular solution of the associated equation.

We aim to prove the result obtained in equation (3), given that:

$$\dot{p} + \beta p = \alpha m$$
$$p(t_0) = p_0$$

***Notation*:** As before, we adopt the common notational simplification of omitting the explicit dependence on t, writing $m = m(t)$ and $p = p(t)$.

Recalling that, for $m_0 = 0$, we have:

$$m(t) = \frac{k}{\gamma}(1 - e^{-\gamma t})$$

To solve for p_h, we set $\alpha m = 0$. Using Leibniz notation, as explained in Appendix A, we obtain:

$$\frac{dp_h}{p_h} = -\beta dt \quad \Rightarrow \quad p_h = p_0 e^{-\beta t}$$

To solve the p_h, we define another function μ(t), known as the integrating factor, such that:

$$\mu(t) = e^{\beta t}$$

Multiplying both sides of the equation by μ(t), we get:

$$p_a' \mu(t) + \beta p_a \mu(t) = \alpha m \mu(t)$$

which simplifies to:

$$\dot{p_a} e^{\beta t} + \beta p_a e^{\beta t} = \alpha m e^{\beta t}$$

Remark: The left-hand side can be rewritten as:

$$\frac{d}{dt}\left(p e^{\beta t}\right) = \frac{\alpha k}{\gamma} e^{\beta t} - \frac{\alpha k}{\gamma} e^{(\beta - \gamma)t}$$

using the product rule for differentiation:

$$\frac{d}{dx}(fg) = f'g + fg'$$

Let's now integrate both sides, choosing integration limits from an initial condition to an arbitrary *t.* Solving the integral, we obtain:

$$p_a = \frac{\alpha k}{\gamma}\left[\frac{1}{\beta}(1 - e^{-\beta t}) - \frac{1}{\beta - \gamma}(e^{-\gamma t} - e^{-\beta t})\right]$$

Summing the homogeneous and particular solutions $p = p_a + p_h$, we obtain the solution of equation (3):

$$p(t) = \frac{\alpha \cdot k}{\gamma}\left[\frac{1}{\beta}(1 - e^{-\beta t}) - \frac{1}{\beta - \gamma}(e^{-\gamma t} - e^{-\beta t})\right] + p_0 e^{-\beta t}$$

Finally, for $t_0 = 0$, we verify that:

$$p(0) = p_0$$

which satisfies the initial condition set in the Cauchy problem and completes the proof (Q.E.D.).

9. References

1. Hobman JL, Penn CW, Pallen MJ. Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully? Mol Microbiol. 2007 May;64(4):881-5. doi: 10.1111/j.1365-2958.2007.05710.x. PMID: 17501914.
2. CRICK FH. On protein synthesis. Symp Soc Exp Biol. 1958;12:138-63. PMID: 13580867.
3. Irastortza-Olaziregi M, Amster-Choder O. Coupled Transcription-Translation in Prokaryotes: An Old Couple With New Surprises. Front Microbiol. 2021 Jan 21;11:624830. doi: 10.3389/fmicb.2020.624830. PMID: 33552035; PMCID: PMC7858274.
4. Miller, O. L., Hamkalo, B. A. & Thomas, C. A. Visualization of bacterial genes in action. Science 169, 392–395 (1970).
5. Milo, R., & Phillips, R. (2015). Cell biology by the numbers. CRC Press.
6. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. Mol Microbiol. 2012 Jul;85(1):21-38. doi: 10.1111/j.1365-2958.2012.08081.x. Epub 2012 May 24. PMID: 22624875; PMCID: PMC3383343.
7. Noll H. The discovery of polyribosomes. Bioessays. 2008 Nov;30(11-12):1220-34. doi: 10.1002/bies.20846. Erratum in: Bioessays. 2009 Jan;31(1):125. PMID: 18937376.
8. Zouridis H, Hatzimanikatis V. A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. Biophys J. 2007 Feb 1;92(3):717-30. doi: 10.1529/biophysj.106.087825. Epub 2006 Nov 10. PMID: 17098800; PMCID: PMC1779991.
9. Jackson MW, Silva-Herzog E, Plano GV. The ATP-dependent ClpXP and Lon proteases regulate expression of the Yersinia pestis type III secretion system via regulated proteolysis of YmoA, a small histone-like protein. Mol Microbiol. 2004 Dec;54(5):1364-78. doi: 10.1111/j.1365-2958.2004.04353.x. PMID: 15554975.
10. Ciechanover A, Schwartz AL. The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death. Proc Natl Acad Sci U S A. 1998 Mar 17;95(6):2727-30. doi: 10.1073/pnas.95.6.2727. PMID: 9501156; PMCID: PMC34259.
11. Miller, J.H. (1972). The Lac Operon: A Short History of a Genetic Paradigm. Trends in Genetics, 18(7), 221-229.
12. Fowler AV, Zabin I. The amino acid sequence of beta-galactosidase of Escherichia coli. Proc Natl Acad Sci U S A. 1977 Apr;74(4):1507-10. doi: 10.1073/pnas.74.4.1507. PMID: 323855; PMCID: PMC430818.
13. Lazzarini, R.A., & Goulian, M. (1978). "Half-life of β-galactosidase in *Escherichia coli*." *Journal of Bacteriology*, 134(2), 601-605.