The Equation of Life

Robert L. Jackson, Stilwell
Copyright® 2021

Abstract:

This study will first define the “equation of life” via the principle of least action. Then the paper will show how this “equation of life” can be used to derive smaller equations, involving transcription and translation, for [computer] modeling and simulation of a cell. The conclusion will provide a terse description of its uses in the realm of Systems Biology.

1. Introduction

In the past, scientists have tried to derive models which attempt to adequately represent life. A couple of groups in academia, the JJ Tyson lab at Virginia Tech and the Molecular Networks Dynamics at Budapest University, or notable examples of scientists who have made efforts to simulate life [1,2]. Using the same underlying mechanisms [3], they have been able to produce workable eukaryotic cell models of cycling cells that are dependent upon certain variables like protein concentrations.

To aid in the endeavor of modeling cells, it might be appropriate to define the scientific laws which dictate processes like transcription and translation. When scientific laws are appropriately described and stated, they can be used to predict natural phenomena [4]. Ideally, mathematics is used to best summarize scientific laws for certain processes in particular fields, such as Physics, Chemistry, etc. However, the first public attempt to use Lagrangian (and Hamiltonian) mechanics was performed by Ventre and associates [6]. They were able to model gene expression among a group cells. There would be many possible uses for an “equation of life” when comes to modeling various cell environments, such as the single.

If an individual can derive an "equation of life," (s)he will be better able to predict the concentration of transcripts and proteins which dictate simulations or models of cell life. This paper will first describe the "equation of life" as a principle of least action. Then the following sections will show this equation can be used to derive known and unknown expressions of transcription and translation via functional differentiation. The conclusion of this study will focus on the possible usage of the equation in terms of modeling and simulations.

1. Deriving the "equation of life"

Lagrangian mechanics, which was established by Joseph-Louis Lagrange, is a formalism of classical mechanics, based upon stationary actions [5]. It considers the position of a set of masses for two given instances, times $t_1$ and $t_2$:

\[
S[q] = \int dt \mathcal{L}(q(t), \frac{\partial q(t)}{\partial t}, t),
\]

where $S$ is the principle of least action, $L$ is the Lagrangian, and $q$ is some quantity/item. For this paper, spatial dimensions will be considered in the principle of least action, thus:

\[
S[q] = \int dtdx^3 \mathcal{L}(q(t, \vec{r}), \frac{\partial q(t, \vec{r})}{\partial t}, t, \vec{r}).
\]

Before proceeding, it is wise to discuss the types of transformation processes that will be utilized in this study. There are four different transformation processes described in this paper. Covariant index is the lower-case index which represents a forward transform from one space to another space while the contravariant index is the upper-case index which represents a backward transform from one space to another space [7,8]. For
example, the rate of synthesis for a transcript via gene and/or coupling constant $s_j$ has two transforms: the forward transform is $j$-index while the backward transform is the $i$-index. The first process involves the mapping from the gene space to the transcript space while the latter process represents the mapping from transcript space to back to the gene space. Also, transformations can either be isomorphic or heteromorphic [9,10,11]. The former process involves a one-to-one mapping while the latter process involves a differing number of mapping. Consider splice variants: one gene $g_i$ may be responsible for the synthesis of multiple splice variants $t_j$.

To derive the "equation of life" in terms of principle actions, one must define the Lagrangian terms within the ultimate expression. The Lagrangian for the "equation of life" is defined as follows:

\[ L = L_{\text{transcription}} + L_{\text{translation}}, \]

where $L_{\text{transcription}}$ and $L_{\text{translation}}$ are the Lagrangian terms for transcription and translation, respectively. The Lagrangian of transcription can be expressed as:

\[ L_{\text{transcription}} = \gamma_i \tau^j \sigma_j + \delta_j \tau_j \tau^j + \partial \tau_j \partial \tau^j, \]

where $\gamma_i$ is the antisense DNA sequence of the $i$-th gene, $\sigma_j$ is the rate of synthesis for $j$-th transcript via the $i$-th gene, $\tau_j$ is the $j$-th mRNA transcript, $\tau^j$ is a small antisense RNA molecule (i.e., TSSaRNA, microRNA, etc.) of the $j$-th transcript, $\delta_j$ is the rate of degradation of the $j$-th mRNA transcript, and $\kappa_j$ is the coefficient of [extracellular] diffusion of small secreted RNA $\tau_j$. Note: $\sigma_j$ and $\delta_j$ also serve as coupling constants of various RNA species $\theta$ with antisense DNA sequence of gene $\gamma_i$ and transcript $\tau_j$, respectively. On the other hand, the Lagrangian of translation can be expressed as:

\[ L_{\text{translation}} = \tau_j \rho^k \sigma_k + \delta_k \rho_k \rho^k + \partial \rho_k \partial \rho^k + L_{\text{oligo}}, \]

where $\sigma_k$ is the rate of synthesis for $k$-th protein via the $j$-th transcript, $\rho_k$ is the $k$-th protein, $\rho^k$ is the ribosomal protein complexes, disordered peptide sequences of the proteasome, etc. of the $k$-th protein, $\delta_k$ is the rate of degradation of the $k$-th transcript, and $\kappa_k$ is the coefficient of [extracellular] diffusion of small secreted RNA $\tau_j$. Assume a particular protein $\rho_k$ exists also as a homo-oligomerization, then:

\[ L_{\text{oligo}} = \sum_{a=1}^{n-1} (r_{2a} \rho_{a+1} - r_{2a-1} \rho_{a+1}) \left( \sum_{a=1}^{n} \rho^k - \rho^{k+1} \right) + \sum_{a=1}^{n-1} \kappa_{k+1} \partial \rho_{k+1} \partial \rho^{k+1}, \]

where $n$ is the number of homo-oligomerization, $r_{odd}$ and $r_{even}$ is a rate of association and/or dissociation. On the other hand, a particular protein $\rho_k$ could exist as a hetero-oligomer, thus:

\[ L_{\text{oligo}} = (r_2 \rho \{1,\ldots,n\} - r_1 (\prod_{a=1}^{n} \rho_a)) \left( \sum_{a=1}^{n} \rho^a - \rho^{\{1,\ldots,n\}} \right) + \kappa \{1,\ldots,n\} \partial \rho \{1,\ldots,n\} \partial \rho^{\{1,\ldots,n\}}, \]

where $\rho \{1,\ldots,n\}$ represents the total hetero-oligomer, $\kappa \{1,\ldots,n\}$ is the coefficient of diffusion for that quaternary protein structure, and $n$ is the total number tertiary protein structures in the hetero-oligomer.
\[ S[\gamma_i, \tau_j, \rho_k] = \]
\[ \int dtdx^3 \left( \gamma_i \tau^j \sigma^i_j + \tau_j \rho^k \sigma^i_k + \delta_j \tau^j \right) \]
\[ + \delta_k \rho \rho^k + \kappa_j \partial \tau_j \partial \tau^j + \partial \rho_k \partial \rho^k + L_{\text{oligo}} \].

Ultimately, the "equation of life" is defined as the total principle of least action for all genes, transcripts, and proteins. In other words, an individual must consider the sum of all genes, transcripts, and proteins, or:

\[ S[\gamma_i, \tau_j, \rho_k] = \]
\[ \int dtdx^3 \sum_i \sum_j \sum_k (\gamma_i \tau^j \sigma^i_j + \tau_j \rho^k \sigma^i_k + \delta_j \tau^j) \]
\[ + \delta_k \rho \rho^k + \kappa_j \partial \tau_j \partial \tau^j + \partial \rho_k \partial \rho^k + L_{\text{oligo}} \].

\[ \text{Mathematica} \] was used to solve the subsequent smaller equations to the basic "equation of life."

1. Deriving the transcript and protein equations for non-dividing cells

Non-dividing cells are simply known as cells which either are arrested or leave some form of cell division (i.e., mitosis, meiosis) [12,13]. An example of cells that are arrested in cell division, or quiescent, are stem cells while mature/adult cells exemplify cells that leave (a series) of cell division[s] [14,15]. Both quiescent and mature/adult cell types leave the cell cycle indefinitely and enter the G₀ phase (figure 1). The environment inside these entities is relatively stable, thus one should not see the periodic appearance of proteins, such as cyclins, that are critical for dividing cells. It is expected that the expression of genes in non-dividing cells is indefinite for quiescent states and permanent for mature adult conditions.

![Figure 1: The different phases of the cell cycle.](image)

The diagram above shows the different phases of the cell cycle. Initially, cells prep early for division, or enter G₁ phase, by doubling their DNA/chromosome...
content that will occur in the $S$ phase. Then cells prepare for division, or enter $G_2$ phase, by segregating their DNA/chromosome content that will occur in the $M$ phase. If a cell wants to arrest or leave the cell cycle, it must enter the $G_0$ phase.

To ascertain transcription in quiescent or mature cells, one must generate the equation for transcript $\tau_j$ in a non-dividing cell. It is assumed that non-dividing cell possesses a first order transcript propagator within its Hamiltonian, or:

$$\mathcal{H}(\gamma_i, \tau_j, \rho_k) = \Pi_\tau \tau^j - \mathcal{L}(\gamma_i, \tau_j, \rho_k),$$

where:

$$\Pi_\tau = \frac{\partial \tau_j}{\partial t}.$$  

The above suggests the new principle of least action, in terms of the Hamiltonian $\mathcal{H}$, becomes:

$$S[\gamma_i, \tau_j, \rho_k] =$$

$$\int dt dx \sum_i \sum_j \sum_k \left( \frac{\partial \tau_j}{\partial t} \tau^j + \gamma_i \tau^j \sigma^i_j + \tau_j \rho^k \sigma^j_k + \delta_j \gamma^j \right),$$

$$\kappa_j = 0,$$

Reducing the above equation to a particular transcript $\tau_j$ produces the following 1st order ordinary differential equation:

$$\frac{\partial \tau_j}{\partial t} - \gamma_i \sigma^i_j + \delta_j \tau_j = 0.$$ 

One of the solutions of the prior equation for a particular transcript $\tau_j$ using the generating function technique, or GFT, [16] is:

$$\tau_j(t) = \frac{a_{10} e^{-t \delta_j}}{A} + \frac{\gamma_j \sigma^i_j}{\delta_j}. $$

If one lets $A$ equal $-1$ and $a_{10}$ equal $\gamma_j \sigma^i_j$, then the solution becomes:

$$\tau_j(t) = \gamma_j \sigma^i_j - \frac{\gamma_j e^{-t \delta_j} \sigma^i_j}{\delta_j}.$$ 

Assuming the initial concentration of transcript $\tau_j$ is 0.0, then the plot of the above solution is simply:
With respect to the starting time point $0.0$, there is an appreciable lag in the concentration of transcript $t_j$ hitting its steady-state level.

Unlike products of transcription, proteins may be secreted from cells. The process of cellular secretion generally requires special organelles called porosomes located at the cellular membrane [16]. Contents within secretory vesicles are released into the environment upon fusing of the vesicles with porosomes. In terms of the "equation of life," the rate of diffusivity for a particular protein $\rho_k$, or $\kappa_k$, is dependent on secretory vesicle-porosome fusion events. If the protein remains in the cell, then $\kappa_k$ is approximately $0.0$.

Next, an individual must derive the equation for protein monomer and oligomer inside a cell. The protein propagator for both dividing and non-dividing cells is as follows:

$$\mathcal{H}(\gamma_i, \tau_j, \rho_k) = \Pi \rho^k - \mathcal{L}(\gamma_i, \tau_j, \rho_k),$$

where

$$\Pi_p = \frac{\partial \rho_k}{\partial t}.$$

Therefore, the principle of least action, in terms of the Hamiltonian, for protein $\rho_k$ becomes:

$$S[\gamma_i, \tau_j, \rho_k] =$$

$$\int dtd\mathbf{x} \sum_i \sum_j \sum_k \left( \frac{\partial \rho_k}{\partial t} \rho^k + \gamma_i \tau^j \sigma^j_i + \tau_j \rho^k \sigma^j_k + \delta_j \tau^j_i \right)$$

$$+ \delta_k \rho_k \rho^k + \kappa_j \partial \tau_j \partial \tau^j + \kappa_k \partial \rho_k \partial \rho^k + \mathcal{L}_{\text{oligo}}.$$

By performing the functional differentiation of this action with respect to contravariant protein $\rho^k$, specifically, (s)he will generate the following expression:

$$\sum_j \sum_k \left( \frac{\partial \rho_k}{\partial t} - \tau_j \sigma^j_k + \delta_k \rho_k - \Delta \rho_k \kappa_k + \frac{\delta \mathcal{L}_{\text{oligo}}}{\delta \rho^k} \right) = 0.$$

Let:

$$\kappa_k = 0.$$
and

\[ \mathcal{L}_{\text{oligo}} = 0 \]

thus, the ordinary differential equation for a particular protein \( \rho_k \) in the form of a monomer is also a 1\textsuperscript{st} order expression:

\[ \frac{\partial \rho_k}{\partial t} - \tau_j \sigma^j_k + \delta_k \rho_k = 0. \]

Using the Runge-Kutta [iterative] method to solve for protein \( \rho_k \), one would obtain the following plot:

![Graph showing the plot](image)

Note: there is an even larger lag in protein \( \rho_k \) hitting its steady-state levels with regards to the starting time point.

Next, one must determine the Lagrangian density for the homodimer of a particular protein \( r_k \). By working with the expression (2.6) and setting \( n = 2 \), one obtains the following Lagrangian:

\[ \mathcal{L}_{\text{oligo}} = \left( r_2 \rho_k^2 - r_1 \rho_k^2 \right) \left( \rho^k - \rho^{k^2} \right) + \kappa_{k^2} \partial \rho_k^2 \partial \rho^{k^2}. \]

If one plug (3.16) into (3.12), sets \( \kappa_{k^2} \) equal to 0.0, and allows \( \rho_{k^2} \) to be the homodimer of protein \( r_k \), then \( (s)he \) will generate two 1\textsuperscript{st} order nonlinear differential equations after performing functional differentiation with respect to both contravariants \( \rho_k \) and \( \rho_{k^2} \):

\[ \frac{\partial \rho_k}{\partial t} - \tau_j \sigma^j_k - r_2 \rho_{k^2} + \delta_k \rho_k + r_1 \rho_k^2 = 0, \]

and

\[ \frac{\partial \rho_{k^2}}{\partial t} + r_2 \rho_{k^2} - r_1 \rho_k^2 = 0. \]

It is important to state that one must also apply a unique protein propagator, hence establish a separate Hamiltonian and principle of least action for the homodimer \( \rho_{k^2} \). The individual generates the following plot if \( (s)he \) uses the Runge-Kutta method to solve for both the monomer and homodimer for a particular protein \( \rho_k \):
Note the appreciable lags in transcript $\tau_j$, monomeric protein $\rho_k$, and homodimeric protein $\rho_{k^2}$ when approaching their relative steady-state levels.

1. Deriving the transcript and protein equations for dividing cells

The hallmark feature of dividing is the oscillatory behavior of intracellular concentration of a transcript and the extracellular concentration of a secreted protein. To produce an oscillating solution, one must use a second-order transcript propagator; thus, the Hamiltonian of for transcript $\tau_j$ is defined as:

$$\mathcal{H}(\gamma_i, \tau_j, \rho_k) = \Pi^2_{\tau} - \mathcal{L}(\gamma_i, \tau_j, \rho_k),$$

where:

$$\Pi_{\tau} = \frac{\partial \tau_j}{\partial t}.$$  \hspace{1cm} (4.2)

The principle of least action regarding the Hamiltonian of transcript $\tau_j$ is as follows:

$$\delta[\gamma_i, \tau_j, \rho_k] =$$

$$\int dt \, dx \sum_i \sum_j \sum_k \left( \frac{\partial \tau_j}{\partial t} \frac{\partial \tau_j}{\partial t} \right) + \gamma_i \tau_j^2 + \tau_j \rho^k_s \sigma_i^j + \delta_j \tau_j^2$$

$$+ \delta_k \rho^k + \kappa_j \partial \tau_j \partial \tau_j^2 + \kappa_k \partial \rho^k \partial \rho^k + \mathcal{L}_{\text{oligo}}.$$

After setting $\kappa_j$ equal to null, the functional differentiation of the principle of least action apropos the contravariant transcript $\tau^j$, specifically, produces:

$$\sum_i \sum_j \left( \frac{\partial^2 \tau_j}{\partial t^2} - \gamma_i \sigma_i^j + \delta_j \tau_j \right) = 0.$$  \hspace{1cm} (4.4)

If one is trying to derive the solution for a particular transcript $\tau_j$, (s)he must use the next equation:
\[
\frac{\partial^2 \tau_j}{\partial t^2} - \gamma_i \sigma_j^i + \delta_j \tau_j = 0.
\]

The general solution to a particular transcript \(\tau_j\) using GFT is:

\[
\tau_j (t) = \frac{1}{2} \left( (a_{10} - a_{20}) A e^{it\sqrt{\delta_j}} + \frac{(a_{10} + a_{20}) e^{-it\sqrt{\delta_j}} + 2\gamma_i \sigma_j^i}{\delta_j} \right).
\]

Assuming \(A\) is equal to negative unity or \(-1.0\), \(a_{20}\) is equal to \(0.0\), and \(a_{10}\) is equal to \(\gamma_i \sigma_j^i \delta_j\), then the solution becomes:

\[
\tau_j (t) = -\frac{\gamma_i \sigma_j^i (\cos(t\sqrt{\delta_j}) - 1)}{\delta_j}.
\]

Next, one must solve for the concentration of the secreted monomeric protein \(\rho_k\). Using the same Hamiltonian (3.7), (s)he will be left with the principle of least action (3.10). Since an individual is dealing with a monomer protein, expression (3.13) is true, and the following equation is left after performing functional differentiation with respect to the contravariant protein \(\rho_k\), specifically:

\[
\sum_j \sum_k \left( \frac{\partial \rho_k}{\partial t} - \tau_j \sigma_k^j + \delta_k \rho_k - \Delta \rho_k \kappa_k \right) = 0.
\]

Note: the delta symbol is a Laplacian operator, thus the protein \(\rho_k\) is dependent upon three spatial dimensions \(\{x,y,z\}\) or radius \(r\) besides time \(t\). By limiting the above expression to a particular protein \(\rho_k\), one must solve the following 2nd order partial differential equation:

\[
\frac{\partial \rho_k}{\partial t} - \tau_j \sigma_k^j + \delta_k \rho_k - \Delta \rho_k \kappa_k = 0.
\]

Implementing the Runge-Kutta method to solve (4.9) produces the following plot assuming the initial [derivative] values for transcript \(\tau_j\) and protein \(\rho_k\):
Note: $|r|$ from the cell is set at 0.0 for the top plot. Both the transcript $\tau_j$ and secreted monomeric protein $\rho_k$ oscillate from the starting time point 0.0. Also, the peaks and troughs of secreted monomeric protein $\rho_k$ lag just behind the peaks and troughs of transcript $\tau_j$.

1. Conclusion

By using some concepts in cell biology and classical mechanics, one can generate the "equation of life." The equation can be utilized in various ways to help model important elements inside and outside the cell. For instance, an individual can stimulate transcription for dividing and non-dividing cells. Also, (s)he can model intracellular and secreted protein concentrations in the same set of cells. Ultimately, one should be able to simulate more sophisticated environments, such as transduction and transfection.

References


[10] Fricke, R. (1892). "Ueber den arithmetischen Charakter der zu den Verzweigungen (2,3,7) und (2,4,7) gehörenden Dreiecksfunctionen" [On the arithmetic character of the triangle functions belonging to the branch points (2,3,7) and (2,4,7)]. Mathematische Annalen (in German). 41: 443–468. doi:10.1007/BF01443421