The Equation of Life

Robert L. Jackson, Stilwell

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 which 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. There would be many possible uses for an "equation of life" when comes to modeling cell environments.

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 possible usage of the equation in terms of modelling and simulations.

2. *Deriving the "equation of life"*

Lagrangian mechanics, which was established by Joseph-Louis Lagrange, is a formulism 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:

(2.1) $$\mathcal{S}(q) = \int dt\, \mathcal{L}\Big(q(t),\, \dot{q}(t),\, t\Big),$$

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

(2.2) $$\mathcal{S}(q) = \int dt dx^3\, \mathcal{L}\Big(q\Big(t,\vec{x}\Big),\, \dot{q}\Big(t,\vec{x}\Big),\, t,\, \vec{x}\Big).$$

Before proceeding, it is wise to discuss the types of transformation processes that will be utilized in this study. There are 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 [6,7]. For example, the rate of synthesis for a transcript via gene and/or coupling constant σ_j^i 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 *homomorphic* or *heteromorphic* [8,9,10]. The former process involves a one-to-one mapping while the latter process involves a differing number of mapping. Consider splice variants: one gene γ_i may be responsible for the synthesis of multiple splice variants τ_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:

(2.3) $$\mathcal{L} = \mathcal{L}_{\text{transcription}} + \mathcal{L}_{\text{translation}},$$

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

(2.4) $$\mathcal{L}_{\text{transcription}} = \gamma_i\, \tau^j\, \sigma_j{}^i - \delta_j\, \tau_j\, \tau^j,$$

where γ_i is the antisense DNA sequence of the *i*-th gene, σ_j^i is the rate of synthesis for *j*-th transcript via the *i*-th gene, τ_j is the *j*-th mRNA transcript, τ^j is the transcription start site associated RNA (TSSaRNA), microRNA, etc. of the *j*-th transcript, and δ_j is the rate of degradation of the *j*-th mRNA transcript. Note: σ_i^j and δ_j also serve as coupling constants of various RNA species τ^j with antisense DNA sequence of gene γ_i and transcript τ_j, respectively. On the other hand, the Lagrangian of translation can be expressed as:

(2.5) $$\mathcal{L}_{\text{translation}} = \tau_j\, \rho^k\, \sigma_k{}^j - \delta_k\, \rho_k\, \rho^k + \kappa_k\, \partial\rho_k\, \partial\rho^k - \mathcal{L}_{\text{oligo}},$$

where σ_k^j is the rate of synthesis for *k*-th protein via the *j*-th transcript, ρ_k is the *k*-th protein, ρ^k is the ribosomal protein complexes, disordered peptide sequences of the proteasome, etc. of the *k*-th protein, δ_k is the rate of degradation of the *k*-th transcript, and κ_k is the rate of extracellular diffusion of protein ρ_k. Note: σ_k^j and δ_k also serve as coupling constants of various protein ρ^k with transcript τ_j and protein ρ_k, respectively. $\mathcal{L}_{oligo}$ is the Lagrangian of hetero- and homo-oligomerization: it is dependent upon a set of protein ρ_k and utilizes principles of "mass conversation" among the proteins being studied. Therefore, the total principle of least action for a particular gene γ_i and subsequent transcripts τ_j and proteins ρ_k becomes:

$$S(\gamma_i, \tau_j, \rho_k) = \int dt dx^3 \left(\gamma_i \tau^j \sigma_j{}^i + \tau_j \rho^k \sigma_k{}^j - \delta_j \tau_j \tau^j - \delta_k \rho_k \rho^k + \kappa_k \partial\rho_k \partial\rho^k - \mathcal{L}_{\text{oligo}}\right) \tag{2.6}$$

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

$$S(\{\gamma_i\}, \{\tau_j\}, \{\rho_k\}) = \int dt dx^3 \sum_i \sum_j \sum_k \left(\gamma_i \tau^j \sigma_j{}^i + \tau_j \rho^k \sigma_k{}^j - \delta_j \tau_j \tau^j - \delta_k \rho_k \rho^k + \kappa_k \partial\rho_k \partial\rho^k - \mathcal{L}_{\text{oligo}}\right) \tag{2.7}$$

Mathematica was used to solve the subsequent smaller equations to the basic "equation of life."

3. *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) [11,12]. Example of cells which are arrested in cell division, or quiescent, are stem cells while mature/adult cells exemplify cells which leave (a series) of cell division[s] [13,14]. Both quiescent and mature/adult cell types leave the cell cycle [indefinitely] and enter the G_0 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.

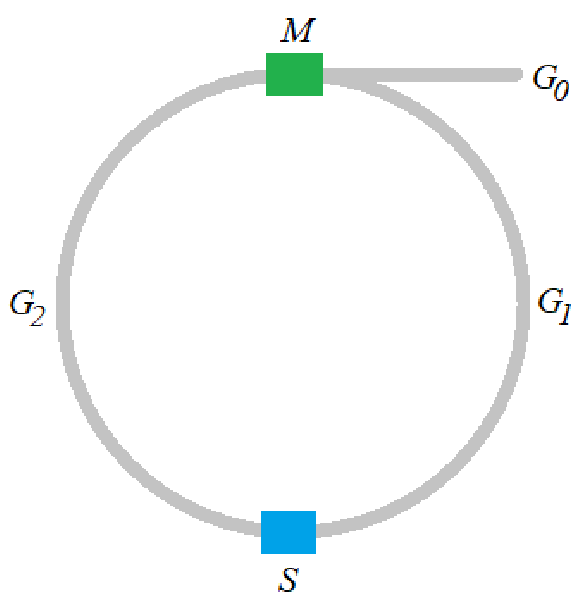


Figure 1: The different phases of the cell cycle. The diagram above shows the different phases of the cell cycle. Initially, cells prep early for division, or enter G_1 phase, by doubling their DNA/chromosome content that will occur in the *S* phase. Then cells prep late for division, or enter G_2 phase, 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 τ_j in a non-dividing cell. It is assumed that non-dividing cell possesses a first order transcript propagator within its Hamiltonian, or:

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

where:

(3.2) $$\Pi_\tau = \dot{\tau}_j.$$

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

(3.3) $$\mathcal{S}(\{\gamma_i\}, \{\tau_j\}, \{\rho_k\}) = \int d\text{t}d\text{x}^3 \sum_i \sum_j \sum_k \left(\dot{\tau}_j \, \tau^j - \gamma_i \, \tau^j \, \sigma_j{}^i - \tau_j \, \rho^k \, \sigma_k{}^j + \delta_j \, \tau_j \, \tau^j + \delta_k \, \rho_k \, \rho^k - \kappa_k \, \partial\rho_k \, \partial\rho^k + \mathcal{L}_{\text{oligo}}\right).$$

The functional differentiation of this equation with respect to the contravariant transcript t^j, specifically, produces:

(3.4) $$\sum_i \sum_j \left(\dot{\tau}_j - \gamma_i \, \sigma_j{}^i + \delta_j \, \tau_j\right) = 0.$$

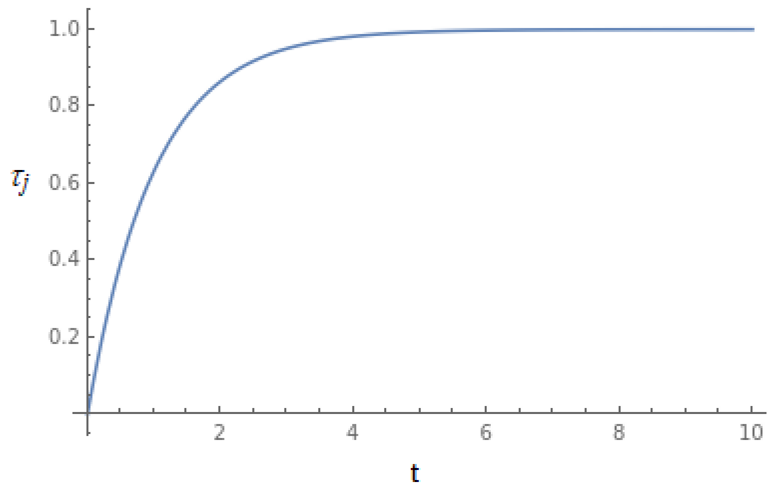
One of the solutions of the prior equation for a particular transcript τ_j using the generating function technique, or GFT, [15] is:

(3.5) $$\tau_j(t) = \frac{a_{10} \, e^{-t\,\delta_j}}{\text{A}} + \frac{\gamma_i \, \sigma_j{}^i}{\delta_j}.$$

If one lets *A* equal *–1* and a_{10} equal $\gamma_i \, \sigma_j^i/\delta_j$, then the solution becomes:

(3.6) $$\tau_j(t) = \frac{\gamma_i \, \sigma_j{}^i}{\delta_j} - \frac{\gamma_i \, e^{-t\,\delta_j} \, \sigma_j{}^i}{\delta_j}.$$

Assuming the initial concentration of transcript τ_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 τ_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 ρ_k, or κ_k, is dependent on secretory vesicle-porosome fusion events. If the protein remains in the cell, then κ_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:

(3.7)
$$\mathcal{H}(\gamma_i, \tau_j, \rho_k) = \Pi_\rho \, \rho^k - \mathcal{L}(\gamma_i, \tau_j, \rho_k),$$

where

(3.8)
$$\Pi_\rho = \dot{\rho}_k,$$

Therefore, the principle of least action, in terms of the Hamiltonian, for protein ρ_k becomes:

(3.9)
$$\mathcal{S}(\{\gamma_i\}, \{\tau_j\}, \{\rho_k\}) = \int dt dx^3 \sum_i \sum_j \sum_k \left(\dot{\rho}_k \rho^k - \gamma_i \tau^j \sigma_j{}^i - \tau_j \rho^k \sigma_k{}^j + \delta_j \tau_j \tau^j + \delta_k \rho_k \rho^k - \kappa_k \partial\rho_k \partial\rho^k + \mathcal{L}_{\text{oligo}} \right).$$

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

(3.10)
$$\sum_j \sum_k \left(\dot{\rho}_k - \tau_j \sigma_k{}^j + \delta_k \rho_k - \Delta\rho_k \kappa_k + \frac{\delta \mathcal{L}_{\text{oligo}}}{\delta \rho^k} \right) = 0.$$

Let:

(3.11)
$$\kappa_k = 0$$

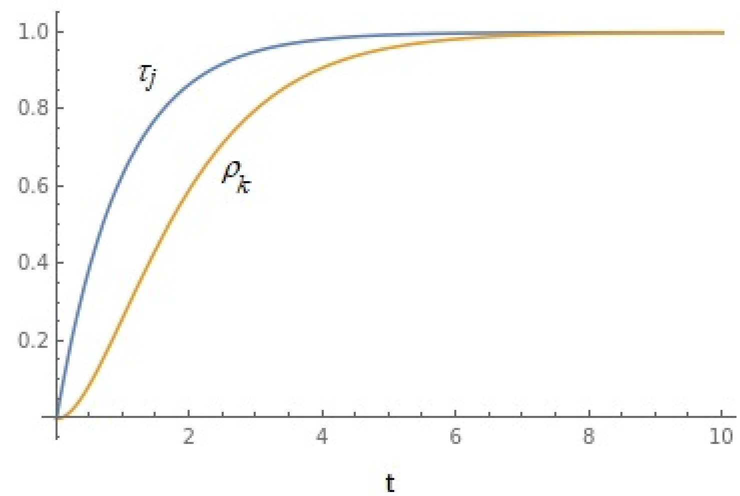
and

(3.12) $$\mathcal{L}_{\text{oligo}} = 0,$$

thus, a particular protein ρ_k in the form of a monomer can defined as:

(3.13) $$\dot{\rho_k} - \tau_j \, \sigma_k{}^j + \delta_k \, \rho_k = 0.$$

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



.

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

Assume a particular protein ρ_k exists also as homodimer, then the Lagrangian of the homo-dimerization of protein ρ_k is:

(3.14) $$\mathcal{L}_{\text{oligo}}(\rho_k, \text{d}\rho_k) = k_1 \, \rho_k^2 \left(\rho^k - \text{d}\rho^k\right) + k_2 \, \text{d}\rho_k \, \text{d}\rho^k,$$

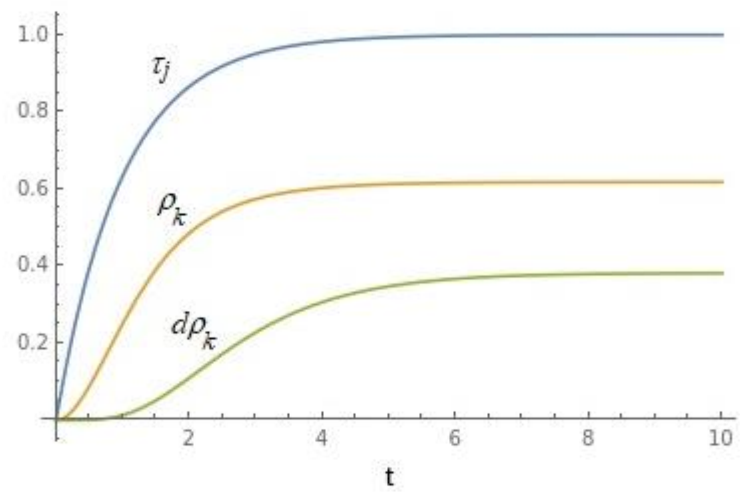
where $d\rho_k$ is the homodimer of protein ρ_k while k_1 and k_2 are rates of association and dissociation, respectively. If one must plug (3.14) into (3.9), (s)he will generate two equations after performing functional differentiation with respect to both contravariants ρ_k and dρ_k:

(3.15) $$\dot{\rho_k} - \tau_j \, \sigma_k{}^j + \delta_k \, \rho_k + k_1 \, \rho_k^2 = 0$$

and

(3.16) $$\text{d}\dot{\rho}_k + k_2 \, \text{d}\rho_k - k_1 \, \rho_k^2 = 0.$$

An individual would generate the following plot if (s)he uses the Runge-Kutta method to solve for both the monomer and homodimer for a particular protein ρ_k:



.

Besides exhibiting lags in transcript τ_j, monomeric protein ρ_k, and homodimeric protein $d\rho_k$ when approaching their relative steady-state levels, the plot shows a depreciation in the magnitude of the steady-state levels for both the monomeric protein ρ_k and homodimeric protein $d\rho_k$.

4. 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 τ_k is defined as:

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

where:

(4.2) $$\Pi_\tau = \dot{\tau}_j.$$

The principle of least action regarding the Hamiltonian of transcript τ_j is as follows:

(4.3) $$\mathcal{S}(\{\gamma_i\}, \{\tau_j\}, \{\rho_k\}) = \int dt dx^3 \sum_i \sum_j \sum_k \left(\dot{\tau}_j \dot{\tau}^j - \gamma_i \tau^j \sigma_j{}^i - \tau_j \rho^k \sigma_k{}^j + \delta_j \tau_j \tau^j + \delta_k \rho_k \rho^k - \kappa_k \partial\rho_k \partial\rho^k + \mathcal{L}_{\text{oligo}} \right).$$

Functional differentiation of the principle of least action apropos the contravariant transcript τ^j, specifically, produces:

(4.4) $$\sum_i \sum_j \left(\ddot{\tau}_j - \gamma_i \sigma_j{}^i + \delta_j \tau_j \right) = 0.$$

If one is trying to derive the solution for a particular transcript τ_j, (s)he must use the next equation:

(4.5) $$\ddot{\tau}_j - \gamma_i \sigma_j + \delta_j \tau_j = 0$$.

The general solution to a particular transcript using GFT is:

(4.6) $$\tau_j(t) = \frac{1}{2}\left((a_{10} - a_{20})\,\mathrm{A}\, e^{i t \sqrt{\delta_j}} + \frac{(a_{10} + a_{20})\, e^{-i t \sqrt{\delta_j}}}{\mathrm{A}} + \frac{2 \gamma_i \sigma_j{}^i}{\delta_j}\right)$$.

Assuming *A* is equal to *-1*, a_{20} is equal to *0*, and a_{10} is equal to $\gamma_i\ \sigma_j^i/\delta_j$, then the solution becomes:

(4.7) $$\tau_j(t) = -\frac{\gamma_i \sigma_j{}^i \left(\cos\left(t \sqrt{\delta_j}\right) - 1\right)}{\delta_j}$$.

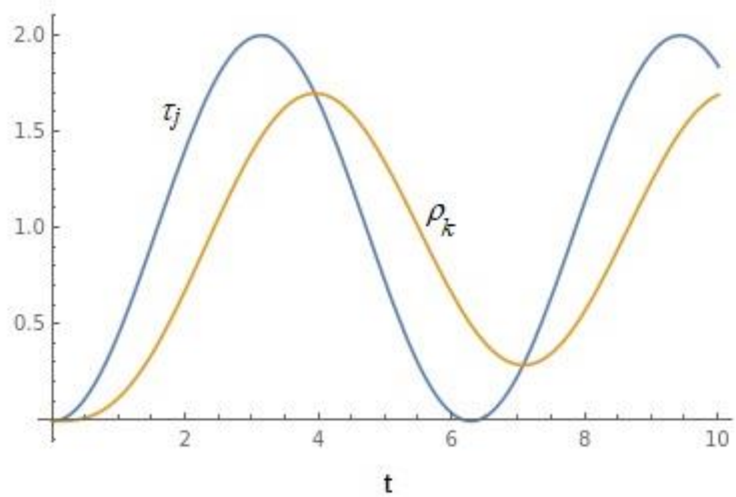
Next, one must solve for the concentration of the secreted monomeric protein ρ_k. Using the same Hamiltonian (3.7), (s)he will be left with the principle of least action (3.10). Since an individual is dealing with a monomer protein, expression (3.12) is true, and the following equation is left after performing functional differentiation with respect to the contravariant protein ρ^k, specifically:

(4.8) $$\sum_j \sum_k \left(\dot{\rho}_k - \tau_j \sigma_k{}^j + \delta_k \rho_k - \Delta\rho_k \kappa_k\right) = 0$$.

Note: the delta symbol is a Laplacian operator, thus the protein ρ_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 ρ_k, one must solve:

(4.9) $$\dot{\rho}_k - \tau_j \sigma_k{}^j + \delta_k \rho_k - \Delta\rho_k \kappa_k = 0$$.

Implementing the Runge-Kutta method to solve (4.9) produces the following plot assuming the initial [derivative] values for transcript τ_j and protein ρ_k are *0.0*:



.

Note: |**r**| from the cell is set at *10.0*. Both the transcript τ_j and secreted monomeric protein ρ_k oscillate from the starting time point *0.0*. Also, the peaks and troughs of secreted monomeric protein ρ_k lag just behind the peaks and troughs of transcript τ_j.

5. 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 simulate 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

[1] Tyson, J.J. (2012) “Irreversible Transitions, Bistability and Checkpoint Controls in the Eukaryotic Cell Cycle: A Systems-level Understanding” *Handbook of Systems Biology* (edit by A. J. Marian Walhout, Marc Vidal and Job Dekker), @Elsevier, San Diego, CA. "The JJ Tyson Lab". Virginia Tech. Retrieved 2011-07-20.

[2] Attila Csikasz-Nagy, Dorjsuren Battogtokh, Kathy Chen, Bela Novak & John J. Tyson Analysis of a generic model of eukaryotic cell cycle regulation published in Biophysical Journal (2006) 90:4361-79. "The Molecular Network Dynamics Research Group". Budapest University of Technology and Economics.

[3] Csikasz, N. (2006) “Analysis of a generic model of eukaryotic cell-cycle regulation” *Biophys J.* Jun 15; 90(12):4361-79.

[4] “Law of nature” *Oxford English Dictionary* (Online ed.). Oxford University Press

[5] Lagrange, J. L. (1811). *Mécanique analytique*

[6] Arfken, G. "Noncartesian Tensors, Covariant Differentiation." §3.8 in *Mathematical Methods for Physicists, 3rd ed.* Orlando, FL: Academic Press, pp. 158-164, 1985.

[7] Morse, P. M. and Feshbach, H. *Methods of Theoretical Physics, Part I.* New York: McGraw-Hill, pp. 44-46, 1953.

[8] Ritter, E. (1892). "Die eindeutigen automorphen Formen vom Geschlecht Null, eine Revision und Erweiterung der Poincaré'schen Sätze" [The unique automorphic forms of genus zero, a revision and extension of Poincaré's theorem]. *Mathematische Annalen* (in German). 41: 1–82. doi:10.1007/BF01443449

[9] 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

[10] Ellerman, D. (2007). "Adjoint functors and heteromorphisms" arXiv:0704.2207v1 Apr. 17.

[11] Griffiths AJ (2012). Introduction to genetic analysis (10th ed.). New York: W.H. Freeman and Co.

[12] "10.2 The Cell Cycle - Biology 2e | OpenStax". openstax.org.

[13] Hüttmann, A (2001). "Functional heterogeneity within rhodamine123lo Hoechst33342lo/sp primitive hemopoietic stem cells revealed by pyronin Y". Experimental Hematology. 29 (9): 1109–1116.

[14] Hayflick L, Moorhead PS (December 1961). "The serial cultivation of human diploid cell strains". Experimental Cell Research. 25 (3): 585–621.

[15] Jackson, R. (2019) "A possible theory of partial differential equations" vixra.org/1910.0064

[16] Lee JS, Jeremic A, Shin L, Cho WJ, Chen X, Jena BP (July 2012). "Neuronal porosome proteome: Molecular dynamics and architecture". Journal of Proteomics. 75 (13): 3952–62.