Research article

Polymer physics-based mathematical models for the correlation of DNA and mRNA in a eukaryotic cell

Archit Chaturvedi*,

Canyon Crest Academy, San Diego, 92130, USA

* **Correspondence:** Email: architchaturvediresearch@gmail.com.

Abstract: DNA and mRNA are essential aspects of cells. They are responsible for much of the genomic activity that takes place in a cell, and are significant macromolecules for research in cell and molecular biology. DNA and mRNA are polymers, molecules that are composed of repeating subunits known as monomers. In the past, a number of theoretical models that elucidate the physical properties of polymers have been proposed to the scientific community. These models include the Freely-Jointed Chain, Freely-Rotating Chain, Worm-Like Chain, and Gaussian Chain Models. In this paper, I make use of such theoretical models in polymer physics, and derive a number of theoretical models that correlate DNA, its respective pre-mRNA strand, and the corresponding post-mRNA strand in a eukaryotic cell. Furthermore, graphical representations of some of the mathematical models derived in the paper are also rendered. Based on this, the theoretical models formulated in this paper can be applied to research in the fields of mathematical biology, biophysics, biochemistry, and cell/molecular biology.

Keywords: theoretical biophysics; mathematical biology; polymer physics; mathematical biophysics; mathematical modeling

1. Introduction

DNA and RNA are essential concepts in molecular biology and genomics. DNA, which stands for deoxyribonucleic acid, serves as the genetic basis for all living organisms [1–3]. DNA is a polymer, and is made up of monomers known as nucleotides. Each nucleotide consists of a nitrogenous base, deoxyribose sugar, and a phosphate group. The sequence of nitrogenous bases in the DNA strand determine the genetic code in organisms. From DNA, a process called transcription produces a molecule known as messenger ribonucleic acid (mRNA) from a template DNA strand [3, 4]. mRNA is very similar in structure to DNA, except that it is single stranded and is usually shorter than the template DNA strand. The strand of mRNA produced immediately after transcription is known as the pre-mRNA strand. After the pre-mRNA strand is transcribed, it goes through a series of post-transcriptional

modifications [5]. These post-modificational transcriptions are mRNA splicing, the addition of a 5' guanine cap, and the addition of a 3' poly-A tail. The pre-mRNA strand, after it has gone through the post-transcriptional modifications, is referred to as the post-mRNA strand. This post-mRNA strand then enters a ribosome, where it is translated to a polypeptide, which ultimately allows for the expression of the original DNA strand in the organism's phenotype. Several research articles have proposed distinct mathematical models for DNA and mRNA. For example, in [6], the author uses concepts from M-theory and string theory in order to derive mathematical models for DNA with regards to interactions between free and bound electrons. Furthermore, in [7], the author mathematically models DNA replication as a data communication model. And in [8], the authors discuss the use of stochastic modeling and computer simulations to model eukaryotic DNA replication. Therefore, in this article, a distinct approach is taken, and

DNA replication is modelled through the use of theoretical models in the field of polymer physics. As discussed, DNA, and therefore RNA, are both polymers. There are a number of theoretical models that have been developed to elucidate the physical properties of polymers [9, 10]. The simplest model for a polymer is known as the *Freely Jointed Chain* (FJC) model [11, 12]. The FJC model describes a polymer as a simple chain, where each monomer has a constant length throughout the overall chain. Using the FJC model, and fixing the angle at which each monomer is bonded in the overall chain, gives the *Freely Rotating Chain* (FRC) model [13]. The FRC model provides a more advanced model for polymers than the FJC model. If the bond angle in the FRC model is θ, then taking the limit $\theta \to 0$ gives rise to the *Worm-Like Chain* (WLC) model [14]. The WLC model makes use of the length of the overall polymer chain, as well as a value of the polymer chain known as the persistence length. The WLC model serves as one of the most advanced models in polymer physics. Finally, the *Gaussian Chain Model* (GCM) renders a stochastic model for polymers [15]. It gives the probability distribution for the conformation of a polymer chain, based on several properties of the chain. The GCM allows for the probabilistic analysis of a polymer. Comprehensively, the FJC, FRC, WLC, and GCM serve as fundamental models in polymer physics that theoretically delineate a number of physical properties of a polymer [16, 17]. This paper makes use of all of these models, and derives theoretical models that correlate the physical properties of a DNA strand, its corresponding pre-mRNA strand, and the resulting post-mRNA strand with each other.

2. Theoretical models for DNA

Here, I propose theoretical models for DNA as a single strand, making them relevant to describe the template DNA strand from which the pre-mRNA strand is transcribed. Suppose that a DNA strand has N total bonds between its nucleotides. This results in the DNA strand having $N + 1$ total nucleotides. Suppose that each nucleotide in the DNA strand has an approximate length of b. Based on this, the extension of the DNA strand can be denoted as that of the FJC:

$$R_{DNA} = b\sqrt{N} \tag{2.1}$$

From this, the entropic spring constant of the DNA strand is given to be:

$$k_{DNA} = \frac{3k_B T}{Nb^2} \tag{2.2}$$

And the radius of gyration of the DNA strand can be given through the *Debye Result* for the strand:

$$G_{DNA} = b\sqrt{\frac{N}{6}} \tag{2.3}$$

Now, as done in the freely-rotating chain, suppose that the nucleotides are bound to each other at a fixed angle θ. From this, the extension of the DNA strand can also be written as:

$$R_{DNA} = b\sqrt{N\left(\frac{1 + cos(\theta)}{1 - cos(\theta)}\right)} \tag{2.4}$$

Letting L_{DNA} denote the length of the DNA strand also gives:

$$L_{DNA} = N \cdot b \tag{2.5}$$

Denoting l_p as the persistence length of the DNA strand:

$$l_p = \frac{-b}{\ln(cos(\theta))}, \tag{2.6}$$

gives the extension of the DNA strand based on the worm-like chain model:

$$R_{DNA} = \sqrt{2l_p L_{DNA} - 2l_p^2(1 - e^{-\frac{L_{DNA}}{l_p}})} \tag{2.7}$$

From the Gaussian chain, the probability distribution for the conformation of the DNA strand is also written as:

$$P(R_{DNA}) = \left(\frac{3}{2\pi Nb^2}\right)^{\frac{3}{2}} \exp\left(\frac{-3R_{DNA}^2}{2Nb^2}\right), \tag{2.8}$$

where R_{DNA} denotes the extension of the DNA strand, and can be written through any of the formulations in equations 2.1, 2.4, and 2.7. This summarizes the theoretical models of DNA based on the FJC, FRC, WLC, and the GCM. Based on this, the mathematical models for mRNA are derived and proposed.

3. Theoretical models for pre-mRNA: Difference treatment

Here, I provide theoretical models that correlate a DNA strand and its corresponding pre-mRNA strand based on a "difference treatment." Furthermore, I also provide some visualizations of my mathematical models, rendered in section 8. The transcribed mRNA strand by RNA polymerase during the process of transcription only works with a certain number of nucleotides from the overall DNA strand. Suppose that from the template DNA strand, t nucleotides are not used in the pre-mRNA transcript. This results in there being a total of $N + 1 - t$ nucleotides in the pre-mRNA strand. Therefore, there are a total of $N - t$ bonds in the pre-mRNA strand. Assuming that the length of each nucleotide remains approximately b, the extension of the pre-mRNA strand based on equation 2.1 is:

$$R_{pmRNA} = b\sqrt{N - t} \qquad (3.1)$$

From equation 2.2, the entropic spring constant of the pre-mRNA strand can also be found as:

$$k_{pmRNA} = \frac{3k_BT}{b^2(N - t)} = \frac{3k_BT}{b^2N - tb^2} \qquad (3.2)$$

Furthermore, the radius of gyration for the pre-mRNA strand can be found from equation 3 as:

$$G_{pmRNA} = b\sqrt{\frac{N - t}{6}} \qquad (3.3)$$

Assuming that the bond angles between the nucleotides in the pre-mRNA molecule remain the same as that of the template DNA strand, at a value of θ, gives another formula for the extension of the pre-mRNA chain:

$$R_{pmRNA} = b\sqrt{(N - t)\left(\frac{1 + cos(\theta)}{1 - cos(\theta)}\right)} \qquad (3.4)$$

Since b remains constant between the template DNA strand and pre-mRNA strand, and so does θ, therefore, the persistence length of the pre-mRNA strand is equal to that the DNA strand. However, since the number of bonds is distinct between the DNA and pre-mRNA strands, therefore, the length of the pre-mRNA strand is written as:

$$L_{pmRNA} = b(N - t) \qquad (3.5)$$

Therefore, the extension of the pre-mRNA strand based on equation 2.7 is:

$$R_{pmRNA} = \sqrt{2l_pL_{pmRNA} - 2l_p^2(1 - e^{-\frac{L_{pmRNA}}{l_p}})} \qquad (3.6)$$

$$= \sqrt{2l_pb(N - t) - 2l_p^2(1 - e^{-\frac{b(N-t)}{l_p}})} \qquad (3.7)$$

Then, the probability distribution for the conformation of the pre-mRNA strand is also given from equation 2.8, and is written as:

$$P(R_{pmRNA}) = \left(\frac{3}{2\pi b^2(N - t)}\right)^{\frac{3}{2}} \exp\left(\frac{-3R_{pmRNA}^2}{2b^2(N - t)}\right) \qquad (3.8)$$

4. Theoretical models for pre-mRNA: Fractional treatment

Here, I derive further mathematical models to correlate a DNA strand and its corresponding pre-mRNA strand in a cell. However, this time, I make use of a "fractional treatment." Visualizations of some of the theoretical models in this section are also rendered in section 8 of the paper. Suppose here that f nucleotides from the original template DNA strand are used in the pre-mRNA strand. Therefore, the fraction of bonds in the pre-mRNA strand and DNA strand is given by:

$$\frac{bonds_{DNA}}{bonds_{pmRNA}} = \frac{f - 1}{N} \qquad (4.1)$$

As before, the extension of the pre-mRNA strand is written as:

$$R_{pmRNA} = b\sqrt{f - 1}, \qquad (4.2)$$

and so from this, the extensions of the pre-mRNA strand and DNA strand are correlated as:

$$\frac{R_{pmRNA}}{R_{DNA}} = \frac{b\sqrt{f - 1}}{b\sqrt{N}} = \sqrt{\frac{f - 1}{N}} \qquad (4.3)$$

The FRC model's correlation is also equal to the value in the equation above, and therefore, there is no need to restate it. From this, a fractional correlation between the entropic spring constants of the pre-mRNA strand and DNA strand can also be formulated as:

$$\frac{k_{pmRNA}}{k_{DNA}} = \frac{3k_BT}{b^2(f - 1)} \cdot \frac{Nb^2}{3k_BT} = \frac{N}{f - 1} \qquad (4.4)$$

The radius of gyration can also be correlated between the pre-mRNA strand and DNA strand, the result being the same as that of equation 4.3:

$$\frac{G_{pmRNA}}{G_{DNA}} = \sqrt{\frac{f-1}{N}} \qquad (4.5)$$

Then, a fractional correlation between the extensions of the pre-mRNA strand and DNA strand can be derived using the WLC model, and is given as:

$$\frac{R_{pmRNA}}{R_{DNA}} = \sqrt{\frac{2l_p b(f-1) - 2l_p^2(1 - e^{-\frac{b(f-1)}{l_p}})}{2l_p bN - 2l_p^2(1 - e^{-\frac{bN}{l_p}})}}$$

$$= \sqrt{\frac{b(f-1) - l_p(1 - e^{-\frac{b(f-1)}{l_p}})}{bN - l_p(1 - e^{-\frac{bN}{l_p}})}} \qquad (4.6)$$

Finally, the fractional correlation of the probability distributions of the pre-mRNA and DNA strands based on the GCM model is given by:

$$\frac{P(R_{pmRNA})}{P(R_{DNA})} = \frac{\left(\frac{3}{2\pi b^2(f-1)}\right)^{\frac{3}{2}} \exp\left(\frac{-3R_{pmRNA}^2}{2b^2(f-1)}\right)}{\left(\frac{3}{2\pi Nb^2}\right)^{\frac{3}{2}} \exp\left(\frac{-3R_{DNA}^2}{2Nb^2}\right)}$$

$$= \left(\frac{N}{f-1}\right)^{\frac{3}{2}} \exp\left(\frac{-3R_{pmRNA}^2}{2b^2(f-1)} + \frac{3R_{DNA}^2}{2Nb^2}\right)$$

$$= \left(\frac{N}{f-1}\right)^{\frac{3}{2}} \exp\left(\frac{3}{2b^2}\left(\frac{R_{DNA}^2}{N} - \frac{R_{pmRNA}^2}{f-1}\right)\right) \qquad (4.7)$$

5. Theoretical models from pre-mRNA to post-mRNA: Difference treatment

Using the same "difference" treatment as that of section 3, theoretical models for the post-mRNA strand after post-transcriptional modifications from the pre-mRNA strand can be formulated. Suppose that ω nucleotides are spliced from the pre-mRNA strand. Then, suppose that a nucleotides are added in the poly-A tail to the pre-mRNA strand, and ρ nucleotides are added via the 5' cap to the pre-mRNA strand to finally give the post-mRNA strand. This results in the total bonds in the post-mRNA strand being:

$$bonds_{fmRNA} = N - t - \omega + a + \rho \qquad (5.1)$$

Due to this, the extension of the post-mRNA strand based on the FJC model is:

$$R_{fmRNA} = b\sqrt{N - t - \omega + a + \rho} \qquad (5.2)$$

Then, the extension based on the FRC model is given to be:

$$R_{fmRNA} = b\sqrt{(N - t - \omega + a + \rho)\left(\frac{1 + cos(\theta)}{1 - cos(\theta)}\right)} \qquad (5.3)$$

The entropic spring constant of the post-mRNA strand can be written as:

$$k_{fmRNA} = \frac{3k_B T}{b^2(N - t - \omega + a + \rho)} \qquad (5.4)$$

The radius of gyration is again given by the Debye Result:

$$R_{fmRNA} = b\sqrt{\frac{N - t - \omega + a + \rho}{6}} \qquad (5.5)$$

The extension of the strand based on the WLC model is given as:

$$R_{fmRNA} = \sqrt{2l_p b(N - t - \omega + a + \rho) - 2l_p^2(1 - e^{-\frac{b(N-t-\omega+a+\rho)}{l_p}})} \qquad (5.6)$$

And finally, the probability distribution for the conformation of the post-mRNA strand is formulated:

$$P(fmRNA) = \left(\frac{3}{2\pi b^2(N-t-\omega+a+\rho)}\right)^{\frac{3}{2}} \exp\left(\frac{-3R_{fmRNA}^2}{2b^2(N-t-\omega+a+\rho)}\right)$$

6. Theoretical models from pre-mRNA to post-mRNA: Fractional treatment

Using the treatment discussed in section 4, fractional models correlating the pre-mRNA and post-mRNA strands can be derived. Suppose that after the post-transcriptional modifications to the pre-mRNA strand, μ nucleotides remain. Therefore, the post-mRNA strand has $\mu - 1$ bonds. From this, the fractional correlation between the extensions of the pre-mRNA strand and post-mRNA strand is given as:

$$\frac{R_{fmRNA}}{R_{pmRNA}} = \frac{b\sqrt{\mu-1}}{b\sqrt{f-1}} = \sqrt{\frac{\mu-1}{f-1}} \qquad (6.1)$$

As discussed, the fractional correlations between the pre-mRNA strand and the post-mRNA strand for the FRC model is exactly the same as that of equation 6.1. Furthermore,

the fractional correlation of the entropic spring constant between the pre-mRNA and post-mRNA strand is:

$$\frac{k_{fmRNA}}{k_{pmRNA}} = \frac{f-1}{\mu-1} \qquad (6.2)$$

From section 4, it can be seen that the fractional correlation for the radius of gyration between two strand is simply the same as that of the extension. Therefore, the fractional correlation for the radius of gyration between the pre-mRNA and post-mRNA strand is equal to equation 6.1, and therefore does not need to be rewritten. Then, the fractional correlation for the extension of the pre-mRNA strand and post-mRNA strand with regards to the WLC model can be written as:

$$\frac{R_{fmRNA}}{f_{pmRNA}} = \sqrt{\frac{b(\mu-1) - l_p(1 - e^{-\frac{b(\mu-1)}{l_p}})}{b(f-1) - l_p(1 - e^{-\frac{b(f-1)}{l_p}})}} \qquad (6.3)$$

And finally, the fractional correlation between the probability distributions for the conformations of the pre-mRNA strand and post-mRNA strand is given by:

$$\frac{P(R_{fmRNA})}{P(R_{pmRNA})} = \left(\frac{f-1}{\mu-1}\right)^{\frac{3}{2}} \exp\left(\frac{3}{2b^2}\left(\frac{R_{pmRNA}^2}{f-1} - \frac{R_{fmRNA}^2}{\mu-1}\right)\right) \qquad (6.4)$$

7. Conclusions

This paper formulates a number of theoretical models correlating a DNA strand, its corresponding pre-mRNA strand, and the corresponding post-mRNA strand to each other. Using the freely-jointed chain, freely-rotating chain, worm-like chain, and Gaussian chain models, mathematical models with regards to the physical properties of DNA are laid out in section 2 of the paper. Then, using the theoretical models for DNA, simple mathematical manipulations result in a number of mathematical models for the respective pre-mRNA strand, and are laid out in section 3 and 4 of the paper. The difference treatment and fractional treatment allow for two alternative means to correlate the physical properties of a DNA strand and its corresponding pre-mRNA strand. Then, further mathematical manipulations of the theoretical models for the pre-mRNA strand yield a number of additional theoretical models for the corresponding post-mRNA strand that has

undergone post-transcriptional modifications. Furthermore, the paper also provides graphical visualizations of the equations derived and rendered in this paper for certain parameters Together, the derived equations in this paper provide a theoretical and mathematical means to correlate a DNA strand and its corresponding pre-mRNA and post-mRNA strands. Through this, the mathematical models derived in the paper can be applied to research in the field of mathematical/theoretical biology, biochemistry/biophysics, as well as in cell/molecular biology.
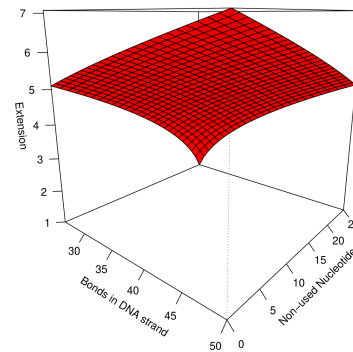
8. Figures

Figure 1. Visualization of the theoretical model for the extension of the pre-mRNA strand based on equation 3.1, for $b = 1$.
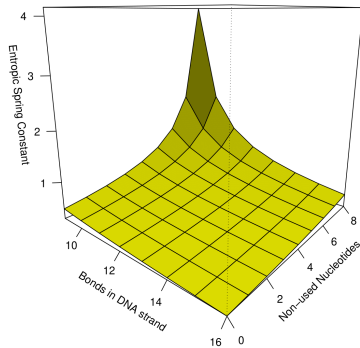
Figure 2. Visualization of the theoretical model for the entropic spring constant of the pre-mRNA strand based on equation 3.2. Here, T is set to 310.15 K, the ideal temperature of the human body, and $b = 1$ again.
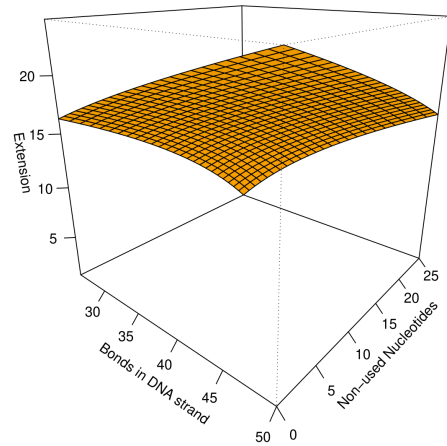

Figure 4. Visualization for the extension of the pre-mRNA strand based on equation 3.6. Here, $l_p = 6.95212$ since $\theta = 30°$, and $b = 1$.
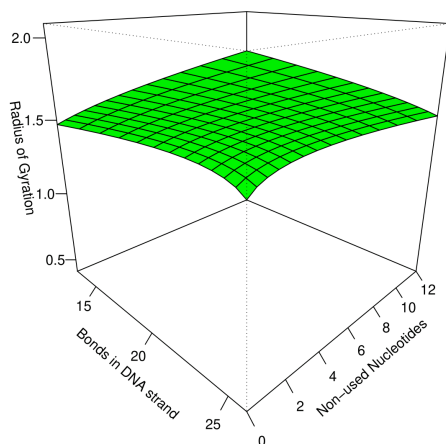

Figure 3. Visualization for the radius of gyration model for the pre-mRNA strand based on equation 3.3. b is set to 1 here as well.
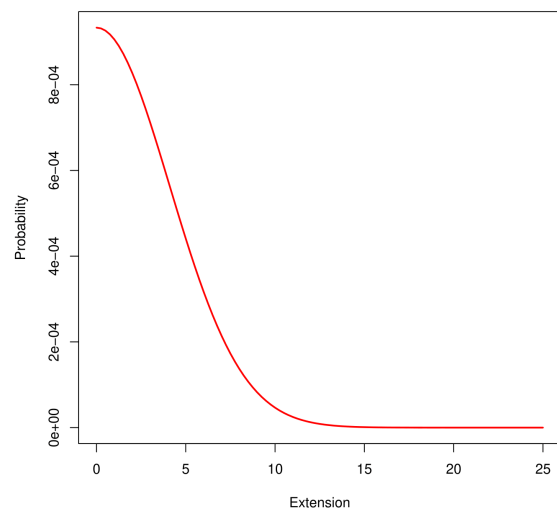
Figure 5. Plot for the Probability Distribution of the Conformation of the pre-mRNA strand based on equation 3.8. Here, $b = 1$, $N = 100$, and $t = 50$ are the parameters for the plot.
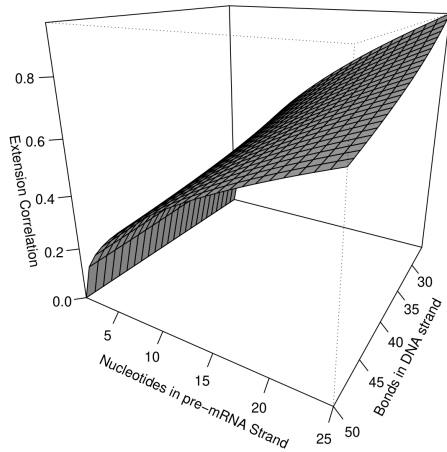
Figure 6. Three-dimensional visualization for the fractional correlation between the pre-mRNA strand and DNA strand based on equation 4.3.
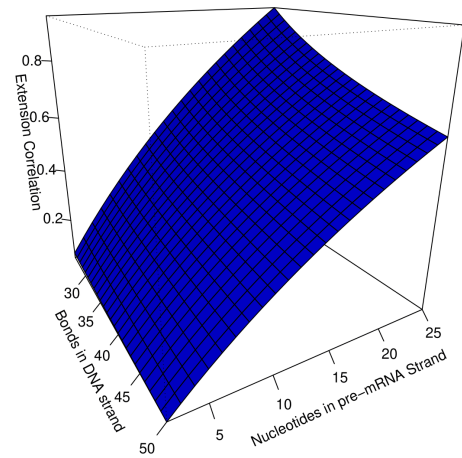

Figure 8. Three-dimensional visualization for the fractional correlation of the extension between the pre-mRNA strand and DNA strand based on equation 23. The parameters for the plot are the same as that of figure 4.
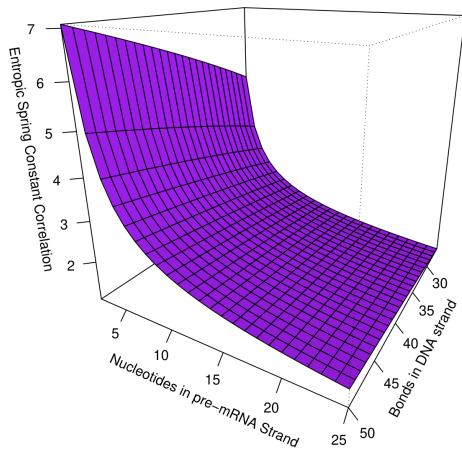

Figure 7. Three-dimensional visualization for the fractional correlations of the entropic spring constants between the pre-mRNA strand and DNA strand based on equation 4.4.
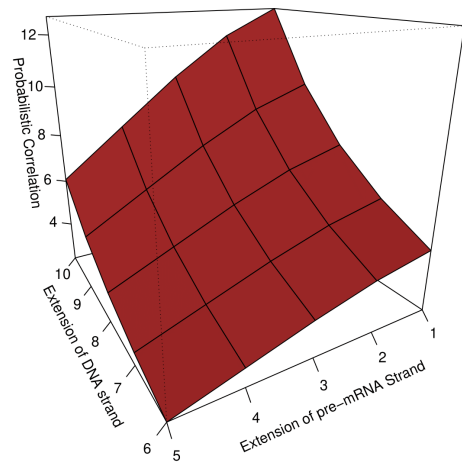

Figure 9. Three-dimensional visualization for the probability distribution between the pre-mRNA strand and DNA strand based on the fractional treatment in equation 26. As done previously in figure 5, $N = 100$, $f = 50$, and $b = 1$.

Conflict of interest

The author has no conflicts of interest to declare that are relevant to the content of this article.

References

1. J. D. Watson, F. H. C. Crick, The structure of DNA, *Cold Spring Harbor symposia on quantitative biology*, Cold Spring Harbor Laboratory Press, 1953. https://doi.org/10.1101/SQB.1953.018.01.020

2. J. D. Watson, F. H. C. Crick, Genetical implications of the structure of deoxyribonucleic acid, *JAMA*, **269** (1993), 1967–1969. https://doi.org/10.1001/jama.269.15.1967

3. J. B. Reece, L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky, R. B. Jackson, *Campbell biology*, Pearson Boston, 2014.

4. R. Langridge, P. J. Gomatos, The structure of RNA, *Science*, **141** (1963), 1024–1024. https://doi.org/10.1126/science.141.3585.1024-b

5. S. Nachtergaele, C. He, The emerging biology of rna post-transcriptional modifications, *RNA biology*, **14** (2017), 156–163. https://doi.org/10.1080/15476286.2016.1267096

6. A. Sepehri, A mathematical model for DNA, *Int. J. Geom. Methods M.*, **14** (2017), 1750152. https://doi.org/10.1142/S0219887817501523

7. A. A. Rushdi, A mathematical model of dna replication, *International Magazine on Advances in Computer Science and Telecommunications (IMACST)*, **1** (2010), 23–30.

8. O. Hyrien, A. Goldar, Mathematical modelling of eukaryotic dna replication, *Chromosome Res.*, **18** (2010), 147–161. https://doi.org/10.1007/s10577-009-9092-4

9. P. O. J. Scherer, S. F. Fischer, *Theoretical molecular biophysics*, Springer, 2010. https://doi.org/10.1007/978-3-540-85610-8

10. U. Schwarz, *Theoretical Biophysics*, heidelberg University, 2021.

11. P. H. Verdier, Relaxation behavior of the freely jointed chain, *The Journal of Chemical Physics*, **52** (1970), 5512–5517. https://doi.org/10.1063/1.1672818

12. M. Mazars, Statistical physics of the freely jointed chain, *Phys. Rev. E*, **53** (1996), 6297. https://doi.org/10.1103/PhysRevE.53.6297

13. A. Perico, S. Bisio, C. Cuniberti, Polymer dynamics in dilute solutions: the freely rotating chain, *Macromolecules*, **17** (1984), 2686–2689. https://doi.org/10.1021/ma00142a041

14. A. Marantan, L. Mahadevan, Mechanics and statistics of the worm-like chain, *Am. J. Phys.*, **86** (2018), 86–94. https://doi.org/10.1119/1.5003376

15. Y. H. Kim, R. P. Wool, A theory of healing at a polymer-polymer interface, *Macromolecules*, **16** (1983), 1115–1120. https://doi.org/10.1021/ma00241a013

16. J. Gao, J. H. Weiner, Range of validity of the entropic spring concept in polymer melt relaxation, *Macromolecules*, **25** (1992), 3462–3467. https://doi.org/10.1021/ma00039a024

17. M. Fixman, Radius of gyration of polymer chains, *The Journal of Chemical Physics*, **36** (1962), 306–310. https://doi.org/10.1063/1.1732501

AIMS Press