



UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”

Mestrando: Rafael Takahiro Nakajima

**Transcrição cooperativa de genes ribossomais  
em *Escherichia coli* usando um modelo  
estocástico e dependente de sequência**

Botucatu

2015

Mestrando: Rafael Takahiro Nakajima

**Transcrição cooperativa de genes ribossomais em  
*Escherichia coli* usando um modelo estocástico e  
dependente de sequência**

Dissertação apresentada ao Instituto de  
Biotecnologia de Botucatu, Câmpus de  
Botucatu, UNESP, para exame de defesa de  
Mestrado no Programa de Pós-Graduação  
em Ciências Biológicas (Genética)

Universidade Estadual Paulista – UNESP

Departamento de Física e Biofísica

Programa de Pós-Graduação: Ciências Biológicas (Genética)

Botucatu

2015

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP  
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Nakajima, Rafael Takahiro.

Transcrição cooperativa de genes ribossomais em  
Escherichia coli usando um modelo estocástico e dependente  
de sequência / Rafael Takahiro Nakajima. - Botucatu, 2015

Dissertação (mestrado) - Universidade Estadual Paulista  
"Júlio de Mesquita Filho", Instituto de Biociências de  
Botucatu

Orientador: Ney Lemke

Capes: 20202008

1. Escherichia coli. 2. Reação em cadeia de polimerase.  
3. Análise estocástica. 4. Sequenciamento de nucleotídeo.  
5. Marcadores genéticos.

Palavras-chave: Atividade RNAP; Motor molecular; Promotor  
de rrn; Sítios de pausa; rrn operons.

# Agradecimentos

Meus sinceros agradecimentos:

- Ao meu orientador Prof. Dr. Ney Lemke pela oportunidade, confiança, incentivo e aprendizagem ao longo desses dois anos de trabalho.
- Aos membros do Laboratório de Bioinformática e Biofísica Computacional, em especial ao Pedro pelos momentos de discussão, paciência e ajuda ao longo do trabalho.
- Aos meus familiares pelo apoio e incentivo.
- Aos amigos Marcos e Eder pelos almoços regrados a boas discussões...
- À Patrícia pelo incentivo e companheirismo mesmo estando um pouco longe.
- Aos colegas do departamento.
- Aos três “amigos” que me estimularam e deram início à minha formação (científica): Prof. Cesar, Diogo e Guilherme.
- Ao CNPq e à Fundação de Amparo à Pesquisa do Estado de São Paulo – processo 2013/06683-2.

# Resumo

Transcrição é o processo catalizado por um complexo enzimático, RNA polimerase (RNAP), responsável pela síntese de RNA mensageiro a partir de uma sequência de DNA. Diferentes estudos experimentais foram realizados para investigar esse processo, como técnicas bioquímicas, de pinça ótica ou magnética, microscopia de força atômica e fluorescência de molécula única. Com os estudos bioquímicos, por exemplo, sabe-se que várias RNAPs podem transcrever uma sequência simultaneamente. O número de diferentes moléculas depende da necessidade da célula, número de RNAP livres, regeneração do promotor e fatores de transcrição. Um dos sistemas mais investigados para o estudo desse processo é a síntese dos genes ribossomais em *Escherichia coli*. Os genes ribossomais são fundamentais na fisiologia dos organismos, são expressos abundantemente e existem evidências da aceleração da transcrição devido ao comportamento colaborativo entre as RNAPs. Neste trabalho, propomos simular a transcrição múltipla dos genes ribossomais em *E. coli* com o modelo estocástico e dependente de sequências desenvolvido em nosso laboratório. As reações químicas foram simuladas utilizando o algoritmo de Gillespie. Essa metodologia apresenta uma boa relação entre custo computacional e realismo biológico e inclui alguns parâmetros não utilizados em estudos teóricos prévios. O modelo considera o alongamento em *back* e *forward tracking*, identificando os sítios de pausas e colisões entre RNAPs, determinando o tempo de permanência e predizendo a ocorrência de transcrição abortiva ou a aceleração da transcrição devido ao fenômeno colaborativo entre múltiplas RNAPs. A sequência do operon ribossomal *rrnB* da *E. coli* foi simulada variando o número de RNAP ( $R$ ), a força de interação entre RNAPs ( $F$ ) e a concentração de nucleosídeo trifosfato ( $[NTP]$ ). Nossos resultados mostraram-se promissores quando utilizamos uma  $F = 15$  pN,  $R = 50$  e  $[NTP] = 100\mu\text{M}$ . O modelo proposto apresenta resultados melhores se comparados a outros estudos teóricos e possui a capacidade de predizer com uma boa acurácia a densidade de RNAP ao longo da sequência de DNA. Porém, alguns fatores importantes não foram incluídos como moléculas inibidoras, fatores de transcrição e a formação de *hairpins* no RNA nascente, variáveis que podem justificar a discrepância entre nossos resultados e os experimentais.

**Palavras-chaves:** operons *rrn*, atividade RNAP, promotor de *rrn*, motor molecular, sítios de pausa, transcrição múltipla.

# Abstract

The process that produces messenger RNAs from DNA sequences is called transcription, and these reactions are catalyzed by the RNA polymerase enzyme. Many different experimental techniques have been applied to investigate this process including biochemical techniques, optical and magnetic tweezers, atomic force microscopy and single molecule fluorescence. These biochemical process studies showed that many RNAP molecules operate simultaneously on a single DNA strand. The number of different molecules depends on cellular demands, concentration of free RNAPs, promoter strength and the presence of transcription factors. *Escherichia coli* ribosomal genes are a popular experimental model to investigate the transcription process. These genes are essential to cell physiology, and they are strongly expressed. There are evidences that some cellular mechanisms collaborate to accelerate their transcription. In this work we investigate the RNAP collaborative transcription in *E. coli* ribosomal genes using a stochastic and sequence dependent model proposed by our group. The chemical reactions were simulated using a model based on the Gillespie algorithm. This methodology is a good compromise between computational cost and biological realism and includes some ingredients that were missing in previous theoretical studies. The model considers back and forward tracking elongation and it identifies pauses by determining the dwell time on specific sites. The model also predicts abortive transcription and transcription acceleration due to collaborative RNAP interaction. The *E. coli rrnB* ribosomal operon sequence was simulated by varying (i) the number of RNAP ( $R$ ) on the DNA strand, (ii) the interaction force between two colliding RNAPs ( $F$ ) and (iii) the concentration of nucleoside triphosphate ( $[NTP]$ ). Our results are promising for  $F = 15$  pN,  $R = 50$  and  $[NTP] = 100\mu\text{M}$  but they also clearly indicate that some key ingredients are still missing in the model, like inhibiting transcription factors and the formation of hairpins in the nascent RNA. The proposed model show improved results when compared with other theoretical studies and could predict with reasonable accuracy the RNAP density profile along the DNA strand.

**Palavras-chaves:** *rrn* operons, RNAP activity, *rrn* promoter, molecular motor, pauses sites, multiple transcription.

# Lista de ilustrações

|  |    |
|--|----|
| Figura 1 – A RNA polimerase . . . . .  | 10 |
| Figura 2 – Processo de transcrição . . . . .   | 12 |
| Figura 3 – Processo de <i>scrunching</i> . . . . .   | 12 |
| Figura 4 – Processos alternativos . . . . .  | 13 |
| Figura 5 – Processo de terminação dependente de Rho . . . . .  | 15 |
| Figura 6 – Micrografia da transcrição dos genes ribossomais da <i>E. coli</i> . . . . .  | 16 |
| Figura 7 – Esquema do cromossomo da bactéria <i>Escherichia coli</i> . . . . .   | 17 |
| Figura 8 – Estrutura básica dos operons ribossomais <i>rrn</i> de <i>Escherichia coli</i> . . . . .  | 18 |
| Figura 9 – Gráfico representativo do caminho da reação . . . . .   | 25 |
| Figura 10 – <i>Heat map</i> das simulações utilizando o operon ribossomal <i>rrnB</i> completo   | 34 |
| Figura 11 – Gráfico representativo da velocidade média das RNAPs . . . . .   | 35 |
| Figura 12 – Representação gráfica da densidade relativa de RNAP da simulação do operon <i>rrnB</i> . . . . .                               | 36 |
| Figura 13 – Representação gráfica da densidade relativa de RNAP e CDF da simulação do operon <i>rrnB</i> . . . . .                         | 37 |
| Figura 14 – Histograma da área sob a curva da CDF das simulações da fase de alongamento . . . . .  | 38 |
| Figura 15 – Representação gráfica da frequência do RMSE e da função de densidade acumulada das simulações da fase de alongamento . . . . . | 39 |
| Figura 16 – Representação gráfica da diferença do tempo de pausas das simulações com a sequência completa e particionada . . . . .         | 39 |

# Lista de tabelas

|   |    |
|---|----|
| Tabela 1 – Valores experimentais para $k_{max}^{NTP}$ e $K_d^{NTP}$ . . . . .                           | 28 |
| Tabela 2 – Parâmetros para simulação da fase de alongamento . . . . .                                   | 30 |
| Tabela 3 – Parâmetros para as novas simulações do operon <i>rrnB</i> . . . . .                          | 31 |
| Tabela 4 – Constantes calculadas para simulações da fase de iniciação . . . . .                         | 32 |
| Tabela 5 – Resultados referentes as simulações do processo de alongamento . . . . .                     | 33 |
| Tabela 6 – Valores do RMSE para as simulações utilizando sequência completa e<br>particionada . . . . . | 37 |



# Sumário

|            |  |           |
|------------|--|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b>  | <b>9</b>  |
| <b>1.1</b> | <b>A RNA Polimerase dependente de DNA</b>  | <b>9</b>  |
| <b>1.2</b> | <b>O processo de transcrição em eubactérias</b>                                      | <b>11</b> |
| 1.2.1      | Processo de transcrição durante um tráfego intenso de RNAPs                          | 14        |
| 1.2.1.1    | DNA ribossomal (DNAr)  | 15        |
| 1.2.1.2    | Atividade do promotor <i>rrn</i> na presença de uma solução saturada de RNAPs livres | 16        |
| 1.2.1.3    | Modelos experimentais e teóricos para o alongamento dos genes ribossomais            | 17        |
| <b>1.3</b> | <b>Proposta</b>  | <b>20</b> |
| <b>2</b>   | <b>OBJETIVOS</b>   | <b>21</b> |
| <b>3</b>   | <b>MÉTODOS</b>   | <b>22</b> |
| <b>3.1</b> | <b>Simulação Estocástica</b>   | <b>22</b> |
| 3.1.1      | Reações químicas   | 22        |
| 3.1.2      | O conceito de estado   | 22        |
| 3.1.3      | Mudança de estado  | 23        |
| 3.1.4      | Algoritmo de Gillespie   | 23        |
| 3.1.5      | Cinética enzimática  | 24        |
| 3.1.6      | Cinética de Michaelis-Menten   | 25        |
| 3.1.7      | Reação enzimática com inibidor   | 26        |
| <b>3.2</b> | <b>O modelo de Costa et al. (CAL)</b>  | <b>27</b> |
| <b>3.3</b> | <b>Modelo para ligação da RNAP na sequência promotora</b>                            | <b>28</b> |
| <b>3.4</b> | <b>Simulações da fase de alongamento da transcrição</b>                              | <b>29</b> |
| 3.4.1      | Transcrição do DNA ribossomal considerando diferentes parâmetros                     | 29        |
| <b>4</b>   | <b>RESULTADOS E DISCUSSÃO</b>  | <b>32</b> |
| <b>4.1</b> | <b>Resultados e análise das simulação da Fase de Iniciação</b>                       | <b>32</b> |
| <b>4.2</b> | <b>Resultados e análise das simulações da Fase de Alongamento</b>                    | <b>32</b> |
| <b>5</b>   | <b>CONCLUSÃO E PERSPECTIVA</b>   | <b>40</b> |
| <b>5.1</b> | <b>Fase da iniciação da Transcrição</b>  | <b>40</b> |
| <b>5.2</b> | <b>Fase do alongamento da Transcrição</b>  | <b>40</b> |
|            | <b>Referências</b>   | <b>42</b> |

# 1 Introdução

O avanço das ferramentas, materiais e métodos disponíveis possibilitou no início do século passado a emergência de uma nova ciência, conhecida como *Biologia Molecular*. Esta ciência foca-se no estudo da estrutura, função e interação das biomoléculas que participam dos fenômenos relacionados com a transmissão e leitura da informação hereditária, e da formação e comportamento do produto por ela codificado. Em 1958, Francis Crick apresentou um quadro de como ocorre a transferência de informação entre os biopolímeros presentes no interior das células, que ficaria conhecido como *Dogma Central da Biologia Molecular*. No Dogma, três moléculas (DNA, RNA e proteínas) interagem através de três diferentes maneiras: no processo de *replicação*, a informação contida numa molécula de DNA pode ser copiada para uma nova molécula de DNA; durante o processo de *transcrição*, essa informação pode ser transferida para uma molécula de RNA; e finalmente, a informação presente no RNA pode ser lida e convertida em proteína durante o processo de *tradução*. Porém em 1970, Crick viu-se obrigado a atualizar seu Dogma, pois estudos mostraram que em condições especiais pode-se observar a transferência da informação contida em moléculas de RNA para DNA (transcrição reversa), a replicação do RNA e até mesmo a tradução direta da informação contida no DNA (CRICK, 1970). E atualmente, com o surgimento de novas classes de RNAs com um importante papel na regulação da transmissão da informação e função tem causado uma grande discussão sobre o termo *Dogma Central da Biologia Molecular*.

O presente trabalho aborda o processo de transcrição do DNA em RNA nas eubactérias, realizado pela enzima *RNA polimerase* (RNAP). Essa introdução apresenta em maiores detalhes a RNAP presente nessa reação e, em seguida, as etapas da transcrição e algumas de suas particularidades. A seção seguinte apresenta alguns modelos teóricos voltados para a previsão do comportamento da RNAP durante a reação. Finalmente a proposta desse trabalho está apresentada na última seção.

## 1.1 A RNA Polimerase dependente de DNA

A RNA polimerase dependente de DNA (Figura 1) é uma enzima que exerce papel fundamental na célula, sendo a responsável pela síntese de RNA a partir de um molde presente na fita de DNA, processo conhecido como *transcrição*. A RNAP é constituída por subunidades, das quais cinco fazem parte do *core*, homologas entre eubactérias, archeas e eucariotos (SWEETSER; NONET; YOUNG, 1987; CRAMER, 2002; WERNER, 2008). Adicionalmente nos eucariotos e archeas as RNAPs apresentam uma maior complexidade em sua composição (de 12 a 15 subunidades) e ativação. Nas eubactérias, apresetam

uma menor complexidade com apenas 6 subunidades no qual, o *core* dessa RNAP são representadas por cinco subunidades denominadas  $\alpha$  (duas subunidades),  $\beta$ ,  $\beta'$ , e  $\omega$ : as subunidades  $\alpha$  são responsáveis pela montagem correta da enzima, subunidades  $\beta$  e  $\beta'$  são os centros catalíticos e a subunidade  $\omega$  auxilia a subunidade  $\alpha$ . A molécula possui também uma subunidade específica que contém um sítio responsável pelo reconhecimento da RNAP a uma determinada sequência presente no DNA, denominada de *sequência promotora* ou simplesmente *Promotor*. Dentre essas subunidades específicas, o fator  $\sigma^{70}$  é o mais estudado por reconhecer promotores dos genes *housekeeping* durante a fase exponencial de crescimento (BURMANN; RÖSCH, 2011; MURAKAMI; DARST, 2003). Na *Escherichia coli* (*E. coli*), além do  $\sigma^{70}$  apresenta mais seis diferentes espécies de fator  $\sigma$  ( $\sigma^N$ ,  $\sigma^S$ ,  $\sigma^H$ ,  $\sigma^F$ ,  $\sigma^E$  e  $\sigma^{FecI}$ ) (ISHIHAMA, 2000). Portanto a estrutura final da RNAP possui um canal principal interno capaz de acomodar de 8 a 9 pares de bases de um híbrido RNA/DNA, um canal de entrada para os nucleotídeos (NTPs) e um canal de saída para a fita de RNA nascente. O sítio ativo, responsável por transferir um novo nucleotídeo (nt) para a extremidade 3' da fita de RNA nascente, está localizado na junção do canal principal com o canal de entrada e envolve um par simétrico de íons  $Mg^{2+}$ : um está permanentemente ligado ao canal enquanto o outro é recrutado *ad hoc* durante cada catálise (SOSUNOV et al., 2003).

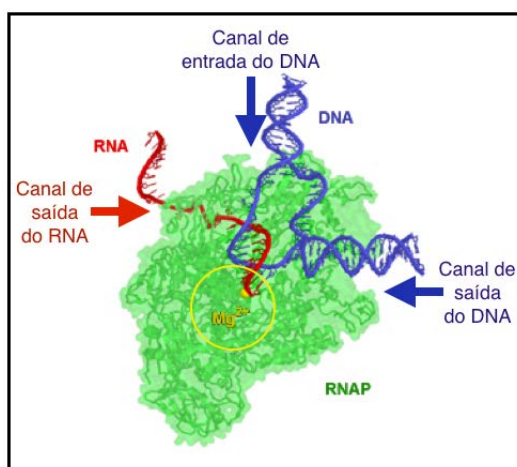


Figura 1 – Representação da RNAP, podendo-se observar o híbrido RNA/DNA presente no canal principal durante a transcrição, além do canal de entrada, onde o íon  $Mg^{2+}$  se liga, e o canal de saída da fita de RNA.

Essa enzima é considerada um motor molecular (SCHLIWA, 2003), ou seja, é capaz de converter energia química em movimento, como a miosina e a cinesina. A RNAP movimenta-se sobre a dupla hélice do DNA podendo se deslocar em ambas as direções; seu deslocamento ao longo da dupla fita é determinada pela ação dos agentes promotores. Quanto à velocidade de deslocamento, a RNAP pode chegar a ser 10 vezes mais lenta que os demais motores, mas enquanto os últimos possuem capacidade de carga de cerca de 5 pN, a RNAP é capaz de lidar com valores de três a cinco vezes maiores (YIN et al., 1995;

WANG et al., 1998).

## 1.2 O processo de transcrição em eubactérias

Os seres vivos são capazes de produzirem elementos encarregados de regular e realizar suas funções necessárias para sua sobrevivência. A princípio, essa informação se apresenta inserida na sequência de DNA e é copiada para uma molécula intermediária que pode ser traduzida. Esta molécula é o RNA, e é produzida através da *transcrição* (Figura 2). Portanto, a transcrição é um processo em que é sintetizado um RNA complementar à do filamento-molde de DNA, catalisado por um complexo enzimático cuja a principal enzima, a RNAP, incorpora um nucleotídeo de cada vez na porção 3' do RNA nascente, através de uma ligação fosfo-diéster entre o grupo hidroxila (OH) livre do açúcar no carbono 3' com o grupo fosfato no carbono 5' e liberando um pirofosfato (PPi) (GRIFFITHS et al., 2006). Para que se inicie, é necessária a presença de uma sequência promotora na fita de DNA. A RNAP é capaz de reconhecer esta região e ancorar na fita de DNA. A partir desse momento, podemos dividir a transcrição em três fases: **iniciação**, **alongamento** e **terminação**. Na fase de **iniciação**, o fator  $\sigma$  da RNAP reconhece a sequência promotora, mais especificamente as regiões consenso -10 (5'-TATAAT-3') e -35 (5'-TTGACA-3'), e se liga a mesma, formando o *complexo fechado*, em seguida, a RNAP sofre uma mudança conformacional (*complexo aberto*), abrindo a dupla hélice expondo o molde da fita que será transcrito (*bolha de transcrição*). Para abandonar essa região ocorre a *iniciação abortiva*, onde pequenas fitas de RNA (2-15 nucleotídeos) são produzidas e liberadas (MCCLURE; CECH; JOHNSTON, 1978). Experimentos mostraram que durante cada ciclo desse processo, a RNAP “puxa” 1 par de base por ligação fosfo-diéster formada, acomodando o DNA “dobrado” dentro de sua estrutura. Na liberação da molécula de RNA abortada, a RNAP expulsa o DNA acumulado, que retorna à conformação original (KAPANIDIS et al., 2006). Esse processo, representado na Figura 3, é conhecido como *scrunching*.

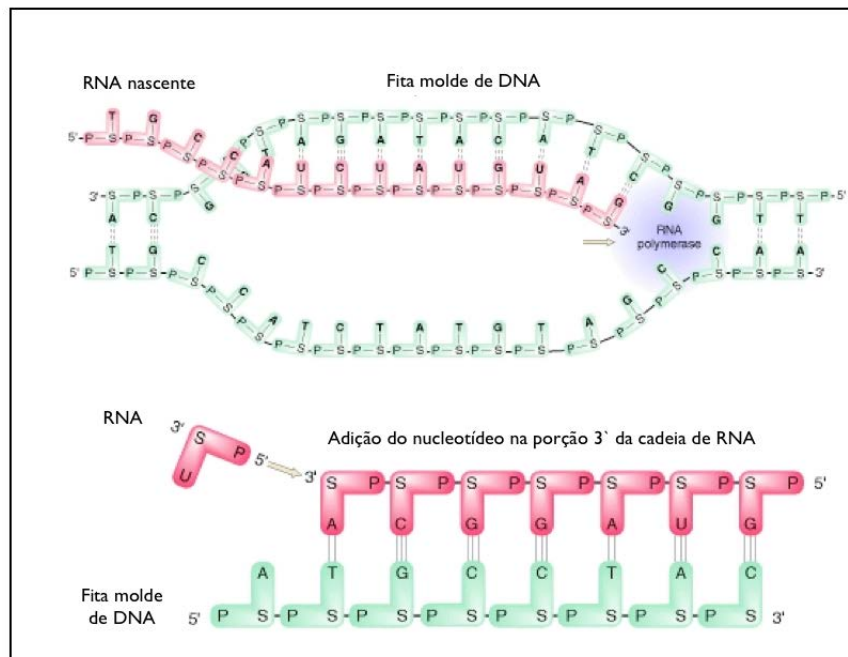


Figura 2 – Processo de transcrição. Em verde: a fita de DNA; em rosa: molécula de RNA sendo sintetizada; em roxo: a RNA polimerase. A seta indica o sentido do processo. (Fonte: Carbonaro, T.M. © 2011 com modificações.)

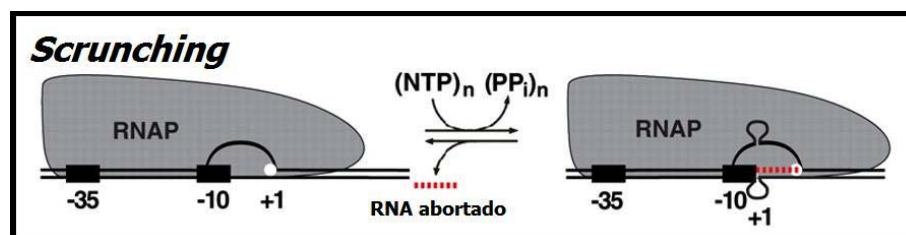


Figura 3 – Processo de *scrunching*. A RNAP “puxa” 1 par de base por ligação fosfo-diéster formada, acomodando o DNA “dobrado” dentro de sua estrutura. Na liberação da molécula de RNA abortada (pontilhado vermelho), a RNAP expulsa o DNA acumulado e retorna à conformação original. (KAPANIDIS et al., 2006).

Quando a RNAP finalmente abandona a região promotora, a transcrição dos primeiros 8 a 12 nucleotídeos leva à liberação do fator  $\sigma$  e a formação do *complexo de alongamento transcricional* (CAT), constituído pela polimerase, pelo DNA e pela fita de RNA nascente, o que indica o final da etapa de iniciação. Enquanto a região promotora manter-se livre, novas polimerases poderão ancorar nesta fita de DNA.

Na fase de **alongamento**, o CAT desliza ao longo da fita de DNA, incluindo, a cada reação, um novo nucleotídeo na fita de RNA nascente, complementar ao presente na fita molde. A RNAP se desloca na fita molde de DNA no sentido  $3' \rightarrow 5'$  e o produto será uma fita de RNA  $5' \rightarrow 3'$ , exatamente igual à fita complementar (GRIFFITHS et al., 2006). Como várias polimerases podem se ligar a um único molde de DNA, várias moléculas de RNA de um mesmo gene podem ser produzidas rapidamente. Técnicas para determinar a velocidade do alongamento, tais como “run-on assays” em *Drosophila* (O'BRIEN;

LIS, 1993), reação em cadeia da polimerase por transcriptase reversa (RT-PCR) em humanos (TENNYSON; KLAMUT; WORTON, 1995) e hibridização *in situ* fluorescência (FISH) (FEMINO et al., 2003) chegaram a valores de taxa média de polimerização que variaram de 1,1 à 2,5 quilobases por minuto, enquanto dados obtidos *in vivo* indicaram 4,3 quilobases por minuto (DARZACQ et al., 2007; DENNIS et al., 2009).

Nessa fase podem ocorrer processos alternativos, que fora denominado de pausas transcricionais (Figura 4). As pausas transcricionais são frequentes e podem não somente diminuir a taxa de produção de RNA, como também permitir que fatores regulatórios atuem no CAT, modificando a transcrição subsequente (ARTSIMOVITCH; LANDICK, 2002; RING; YARNELL, 1996). Pausas breves, com duração menor que 25 segundos e presentes em média a cada 100 pares de bases são conhecidas como “pausas ubíquas” (*ubiquitous pauses*) e representam cerca de 95% das pausas detectadas (NEUMAN et al., 2003). Pausas de maior duração, ditas estáveis, são conhecidas pelo seu papel regulatório e geralmente estão ligadas à formação de *hairpins* (estrutura em grampo). Além disso, essas pausas podem ser resultado do processo conhecido como *backtracking*, durante o qual a RNAP recua alguns pares de bases, permitindo a correção de bases incorporadas equivocadamente, através da clivagem do fragmento que contém o erro na extremidade 3' da fita de RNA. A probabilidade do CAT entrar em um desses processos alternativos depende de interações entre sequências específicas do complexo com o DNA ou com o RNA nascente, ou ainda devido a fatores de transcrição regulatórios.

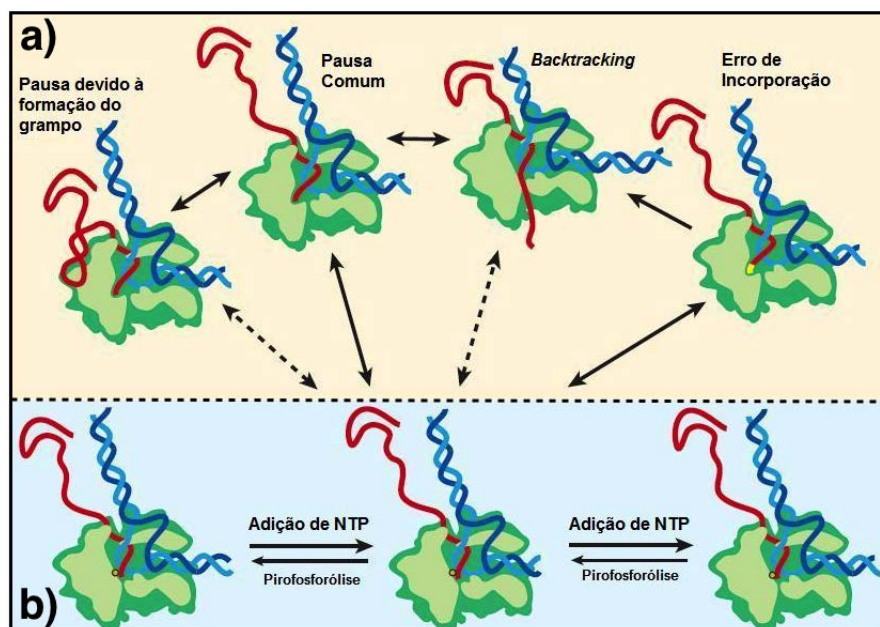


Figura 4 – Processos alternativos que podem ocorrer durante o alongamento. Setas contínuas representam processos com maior probabilidade de ocorrência em relação às tracejadas. Região azul: alongamento propriamente dito. Região alaranjada: processos alternativos (HERBERT et al., 2006).

Finalmente, se não se soltou prematuramente ou se não ficou preso a fita, o CAT

segue com o alongamento até atingir o sítio de terminação da fita molde. Durante a **terminação** ocorre o desenrolamento do híbrido RNA-DNA, a liberação da molécula de RNA pela RNAP e o colapso da bolha de transcrição (GREIVE; HIPPEL, 2005). Nas eubactérias, o processo de terminação pode ocorrer numa região conhecida como *sítio de terminação intrínseca* ou influenciada pelo fator proteico Rho. No primeiro processo, essa região de terminação intrínseca é transcrita e resulta numa estrutura em forma de grampo, seguida por uma sequência de aproximadamente 7 uridinas (cauda “Poli U”). Essa sequência de uridinas constituirá um híbrido DNA-RNA particularmente fraco com as desoxiadenosinas complementares (MARTIN; TINOCO, 1980), induzindo a pausa do CAT nessa região. O CAT fica impedido de recuar devido a interações entre o grampo formado e a polimerase, e, conseqüentemente, o tempo de parada do complexo nessa região aumenta. Isso permite o completo dobramento do grampo, e alterações estruturais na RNAP que desestabilizam a bolha de transcrição (HERBERT; GREENLEAF; BLOCK, 2008). Na terminação dependente do fator proteico Rho é necessário o reconhecimento da região de terminação pela RNAP e a ligação dos monômeros da proteína Rho na fita do RNA até a formação de um hexâmero. Este hexâmero deslizará pela fita de RNA até a região final do transcrito, indicando o término da transcrição (Figura 5). Neste processo, ao contrário do primeiro, o transcrito não apresenta cauda “Poli U” (NUDLER; GOTTESMAN, 2002).

### 1.2.1 Processo de transcrição durante um tráfego intenso de RNAPs

Durante o processo de transcrição novas RNAPs podem identificar a região e transcrever a mesma sequência, fenômeno denominado de *transcrição múltipla*. Em eucariotos esse fenômeno ocorre quando fatores de transcrição mantêm-se ligados a região promotora; nas eubactérias, isso ocorre quando o promotor está livre para novas interações e há presença de RNAPs livres no sistema. Geralmente, este processo é um modo de regular a síntese de RNA e está intimamente ligado a atividade do promotor. Um bom exemplo é a transcrição dos genes ribossomais em *E. coli*: durante seu crescimento exponencial, até 13.000 moléculas de RNAP transcrevem ativamente cerca de apenas 90 unidades transcricionais (GRIGOROVA et al., 2006). Genes ribossomais são conhecidos por serem altamente transcritos e famosos pela sua organização em forma de “árvore de Natal” quando vistas através micrografia eletrônica (EM) (Figura 6). Nestas micrografias é possível ver até centenas de “galhos”: cada um é na verdade uma molécula de RNA recém transcrita em diferentes regiões, confirmando a alta concentração de RNAPs nesses genes (FOE, 1977; HAMMING et al., 1981). Vários artigos confirmam que o comportamento transcricional é alterado nesses casos. Epshtein e Nudler (2003) mostraram que quando uma sequência possui um sítio de pausa muito intenso, a transcrição completa dessa sequência será eficiente apenas na presença de múltiplas moléculas de RNAP. Outro trabalho do mesmo laboratório (EPSHTEIN et al., 2003) mostra que a presença de

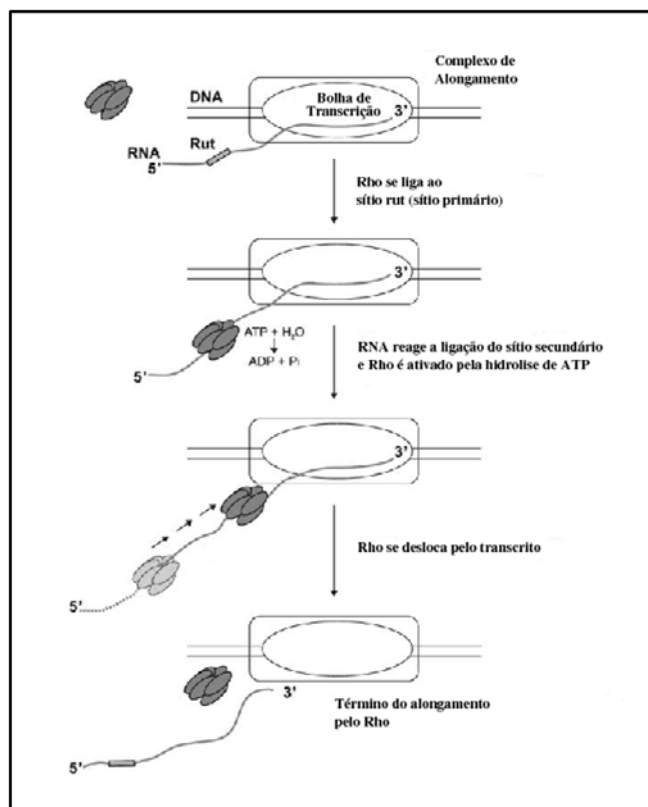


Figura 5 – Processo de terminação dependente de Rho. A proteína Rho se liga ao sítio primário (rut) na fita de RNA formando um hexâmero. Após se ligar a um segundo sítio, a proteína Rho inicia o deslocamento pela fita de RNA a partir da hidrólise de ATP até o fim do transcrito (NUDLER; GOTTESMAN, 2002).

obstáculos proteicos ligados à fita de DNA, como presentes *in vivo*, também é dependente da transcrição múltipla: as moléculas de RNAP removem esses obstáculos presentes na fita no momento em que se chocam com eles, mas a intensidade desse choque dependerá do número de enzimas participantes. Além disso, o modelo para tráfego intenso de RNAPs apresentado por Klumpp (2011) mostrou que a ocorrência de *backtracking* fica restrita pois as enzimas acabam funcionando como obstáculos para o fenômeno. Rajala et al. (2010) concluíram através de seu modelo que a presença de múltiplas enzimas e de pausas na sequência molde acarreta em pequenos “surtos” de produção de RNA. A diversidade de estudos utilizando os genes ribossomais de *E. coli* no processo de transcrição múltipla nos leva a utilizar este modelo biológico em nossas simulações.

#### 1.2.1.1 DNA ribossomal (DNAr)

A família multigênica que codifica o RNA ribossomal (RNAr) é chamada de *DNA ribossomal*, e está geralmente organizada *in tandem* na fita de DNA. Em eucariotos, é dividida em diferentes unidades transcricionais: DNAr 18S, 5.8S e 25S/28S com regiões internas transcritas (ITS I e II) e gene DNAr 5S, com sequências não transcritas (NTS) flanqueando-a (LONG; DAWID, 1980; EICKBUSH; EICKBUSH, 2007). Nas eubactérias,



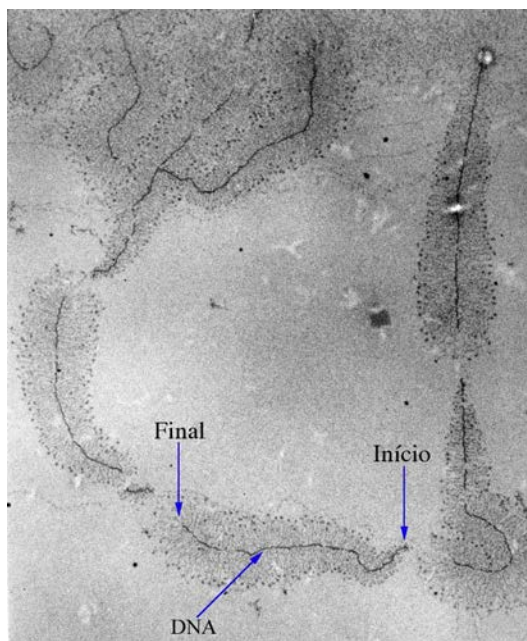


Figura 6 – Micrografia da transcrição dos genes ribossomais da *E. coli*, note a conformação de “árvore de Natal”. Fonte: Wikimedia Commons.

encontram-se apenas três RNAs ribossomais: 16S, 23S e 5S, com um RNA transportador entre os DNAr 16S e 23S (CONDON; SQUIRES; SQUIRES, 1995; PAUL et al., 2004). Os produtos destes genes constituem as principais estruturas do ribossomo: os genes RNAr 23S e 5S (eubactérias), 28S, 5.8S e 5S (eucariotos) constituem parte da subunidade maior do ribossomo, enquanto o gene RNAr 16S (eubactérias) e 18S (eucariotos) a subunidade menor do ribossomo. Na bactéria *E. coli* estão presentes sete diferentes operons que codificam esses genes ribossomais, denominados *rrnA*, *rrnB*, *rrnC*, *rrnD*, *rrnE*, *rrnG* e *rrnH* (Figura 7) que, apesar de possuírem diferentes codificações, são semelhantes na organização de sua estrutura. Estes operons apresentam dois promotores principais (P1 e P2) separados por aproximadamente 100 pares de bases (pb), além de sequências reguladoras (Fis, UP elemento) flanqueando o promotor P1, uma sequência do complexo do sistema de antiterminação do *rrn* boxB-boxA-boxC após promotor P2 (Figura 8), um RNA transportador (RNAt) e dois sítios de terminação (Figura 8) (CONDON; SQUIRES; SQUIRES, 1995; PAUL et al., 2004; KLUMPP; HWA, 2008).

#### 1.2.1.2 Atividade do promotor *rrn* na presença de uma solução saturada de RNAPs livres

Define-se a *atividade de um promotor* como a taxa de ocorrência da iniciação da transcrição por unidade de tempo (1/s). Para promotores constitutivos, a atividade depende de sua especificidade a respostas na variação da concentração de RNAP disponível. Já em promotores regulados, a atividade depende do efeito dos fatores de regulação. Os promotores P1 são geralmente regulados positivamente pelo fator Fis e negativamente pela guanosina tetrafosfato (ou pentafosfato) (ppGpp). Os promotores P2 apresentam-se

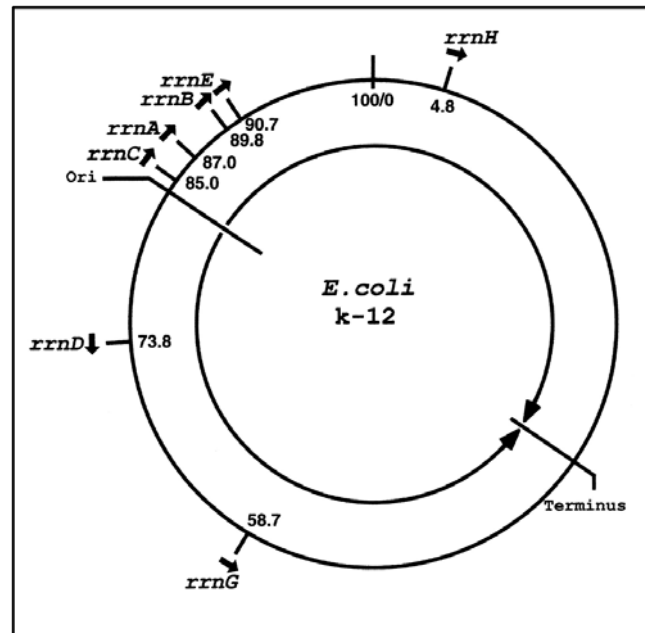


Figura 7 – Esquema do cromossomo da bactéria *Escherichia coli* K-12, com a localização em minutos dos sete operons de RNAr, *rrnA*, *rrnB*, *rrnC*, *rrnD*, *rrnE*, *rrnG* e *rrnH*, além do sítio de origem da replicação, *oriC*. As setas representam a direção da transcrição (NOMURA, 1999).

como promotores constitutivos, ou seja, apresentam uma transcrição basal (DENNIS et al., 2009; EHRENBERG; DENNIS; BREMER, 2010). Durante o crescimento rápido em meio rico, a *E. coli* possui um aumento na atividade do promotor P1. Para que isto ocorra é necessário que o promotor esteja disponível para sucessivas ligações com novas RNAPs. Esta atividade está intimamente relacionada à concentração de RNAPs livre: em condições normais, o promotor P1 não atinge sua atividade máxima devido a esse fator. Entretanto, um valor próximo a esse valor máximo foi observado em um experimento no qual foi realizada a deleção de 4 operons de RNAr. A atividade nesse caso foi 60% maior do que a observado na sequência selvagem (EHRENBERG; DENNIS; BREMER, 2010). Portanto, em *E. coli*, a alta atividade do promotor P1 durante o rápido crescimento em meio rico é devido à saturação da ligação do fator Fis ao seu alvo, à saturação da ligação da RNAP ao respectivo promotor e aos baixos níveis da molécula ppGpp. Quando isto ocorre, a região *downstream* do promotor P2 não é acessível devido à sua alta taxa de transcrição. Entretanto, no crescimento da bactéria em condições não ótimas, o promotor P1 sofre *down-regulation* devido à baixa concentração de Fis e alta concentração de ppGpp. Consequentemente, a transcrição de todo RNAr é realizada a partir do promotor constitutivo P2 (CONDON; SQUIRES; SQUIRES, 1995; EHRENBERG; DENNIS; BREMER, 2010).

### 1.2.1.3 Modelos experimentais e teóricos para o alongamento dos genes ribossomais

Para explicar o movimento da RNAP na fita de DNA, diversos modelos e experimentos foram propostos. O modelo da “catraca browniana”, onde flutuações

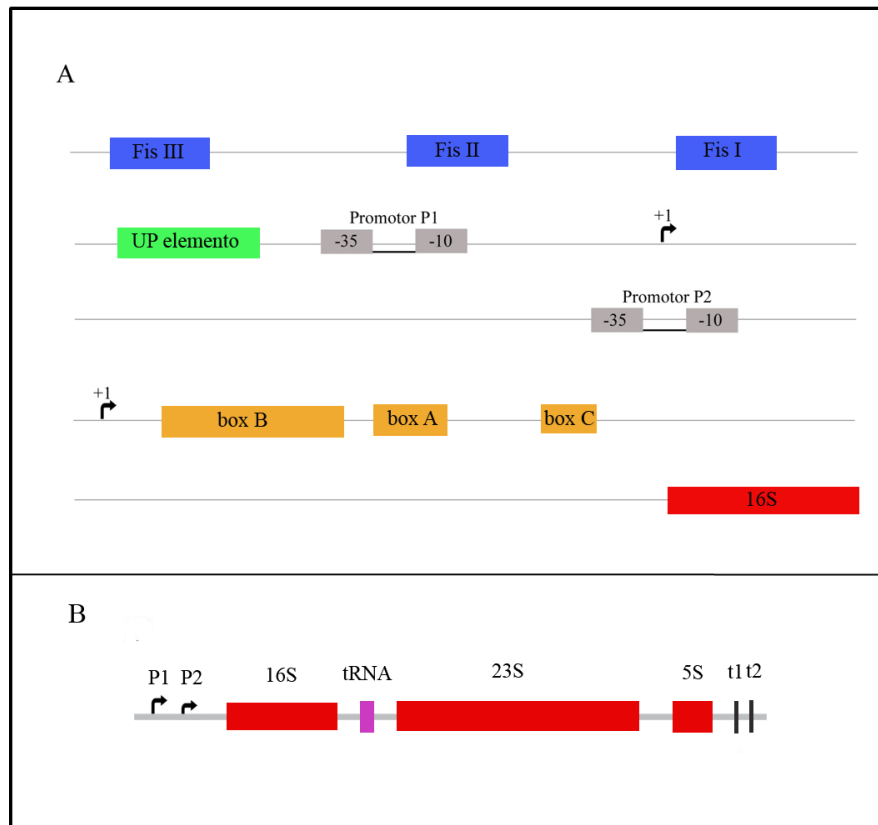


Figura 8 – Estrutura básica dos operons *rrn* de *Escherichia coli*. **A.** Caixas cinzas representam os sítios -35 e -10 dos promotores P1 e P2, responsáveis pela ligação da RNAP; Em azul estão os sítios de ligação para os fatores de regulação Fis III, II, I; a caixa verde representa um fator de ativação da região *upstream*, UP elemento; Em laranja estão as sequências do sistema de antiterminação boxB-boxA-boxC; Em vermelho o DNAr 16S. **B.** Sequência dos operons *rrn*. Setas são os promotores P1 e P2; em vermelho estão os genes ribossomais 16S, 23S e 5S; em roxo, temos o RNA transportador (tRNA) entre o DNAr 16S e 23S; no fim, da sequência em preto o sistema de terminação t1 e t2 (CONDON; SQUIRES; SQUIRES, 1995) com modificações.

térmicas entre os estados pré e pós-translocados desta enzima são mecanicamente retificados pela hidrólise durante a ligação do NTP, resultando em um movimento unidirecional, é o mais aceito (HERBERT; GREENLEAF; BLOCK, 2008). Entretanto, existem complicações para a simulação do processo de transcrição durante tráfego intenso de RNAPs.

Klumpp e Hwa (2008) propuseram um modelo estocástico para essa situação. Nesse modelo a RNAP é um objeto de tamanho  $L$  confinado a se mover ao longo de um eixo onde estão representados as bases da fita molde de DNA. Este modelo considera sítios de pausa e de terminação, além da ação do complexo de antiterminação (AT) na transcrição do RNAr durante o rápido crescimento da bactéria *E. coli*. As características do modelo foram baseadas em estudos quantitativos da transcrição *in vivo* e na observação da dinâmica individual da RNAP *in vitro*. O processo foi analisado passo a passo, incluindo interação RNAP-RNAP, efeitos de pausa, *backtracking*, supressão de pausa pelo complexo

AT, supressão pelo complexo AT incompleto e terminação dependente de Rho. Foram realizadas simulações de Monte Carlo e os parâmetros utilizados no modelo foram baseados em dados experimentais. Entretanto, esse modelo considera que a taxa de alongamento é constante durante toda a transcrição e estima de forma sumária os parâmetros utilizados nas simulações.

Condon et al. (1993) afim de verificar a regulação da síntese dos genes ribossomais na bactéria *E. coli* realizaram a transcrição *in vivo* de todos os operons *rrn* intactos e com deleção-inativação de até 4 dos 7 operons ribossomais. Sem a deleção-inativação de nenhum operon, a taxa de alongamento da transcrição foi mensurada através do acúmulo de RNAt triptofano presente no fim do operon. O volar estimado de velocidade média de incorporação nessas condições foi de 90 nucleotídeos por segundo (nt/s). Entretanto, com a inativação de 4 dos 7 operons ribossomais, ocorre um aumento de 60 % na taxa de iniciação e 50 % na velocidade de alongamento (183 nt/s). Através da microscopia eletrônica (EM), identificaram, devido a conformação clássica em “árvore de Natal”, o operon *rrn* sendo transcrito e constataram que na transcrição do operon intacto há um aumento da taxa de iniciação e alongamento em reflexo ao aumento do número de polimerases por operon. Os autores concluíram que a deleção de alguns operons aumenta a síntese dos genes ribossomais para suprir a necessidade de ribossomos da célula, devido ao aumento da taxa de iniciação, da velocidade de alongamento e do número de moléculas de RNAP que transcrevem o operon simultaneamente.

Quan et al. (2005) através da transcrição *in vivo* e da EM estudaram o comportamento das moléculas de RNAP na presença e ausência de proteínas reguladoras, NusA e NusB, na transcrição do operon ribossomal. Analisaram o número de moléculas de RNAP nos genes 16S e 23S, mediram a distância entre duas polimerases adjacentes e determinaram a distribuição das moléculas de RNAP em todo operon. Para isso, a partir da observação da EM, eles contaram o número relativo de RNAPs com RNA nascente ligado e obtiveram uma média a partir da transcrição de 11 operons. Posteriormente, aproximaram o tamanho do operon *rrn* para 6000 nt e dividiram-no em 20 seções de 300 nt, determinando a densidade de moléculas de RNAP para cada seção. Em relação a esses resultados os autores concluíram que as proteínas reguladores são necessárias para maior eficiência da expressão do operon *rrn*, apresentando uma maior densidade de moléculas de RNAP nos genes 16S e 23S devido a sítios de pausa e ocasionando uma redução na velocidade de transcrição nesses trechos.

Dennis et al. (2009), baseados nos experimentos de Quan et al. (2005), realizaram a transcrição *in vivo* do operon *rrn* e a EM em condições ótimas de crescimento, além de simulações de Monte Carlo baseada no algoritmo de Gillespie (GILLESPIE, 1977) para o processo de transcrição. Analisaram a velocidade e densidade de moléculas de RNAP ao longo do operon. Em relação a velocidade de alongamento, houve uma variação ao longo

da sequência: no início do operon, a RNAP transcreveu aproximadamente 230 nt/s; a velocidade decaiu para aproximadamente 60 nt/s na região dos genes RNAr 16S e 23S; na região intergênica (entre 16S e 23S) estimou-se 100 nt/s e, no fim do operon atingiu mais de 400 nt/s. Em relação a densidade de moléculas de RNAP, eles reproduziram o experimento de Quan et al. (2005) e realizaram análises teóricas. Com estes estudos verificaram uma distribuição não uniforme das moléculas de RNAP, encontrando um acúmulo de RNAPs no início dos genes ribossomais 16S e 23S, provavelmente resultante pela formação de uma estrutura secundária entre estes genes e a molécula de RNAP.

Fange et al. (2014) propuseram um modelo termodinâmico dependente da sequência para transcrição do operon ribossomal considerando a fase de iniciação da transcrição, translocação e as situações normais de alongamento e em *back* e *forward tracking*. Realizaram várias simulações implementando cada fator citado acima ao modelo, verificando o comportamento das RNAPs e comparando com os dados de Quan et al. (2005). Entretanto, obtiveram um resultado mais significativo a partir da implementação de uma penalização de 3 RT na energia livre do *back tracking* para cada passo subsequente ao primeiro. Os autores concluíram que a alta taxa de iniciação influencia no número de RNAPs e na velocidade de transcrição, e a presença da alta densidade nos genes ribossomais 16S e 23S deve-se a um longo sítio de pausa, resultado da provável formação de um *hairpin* pelo RNA nascente. Porém, devido a falta de estudos sobre este fenômeno (dobramento do RNA nascente) não é possível incluí-lo em um modelo termodinâmico.

### 1.3 Proposta

Diante das informações apresentadas e dos resultados obtidos pelos modelos citados, propomos utilizar o modelo para transcrição múltipla dependente de sequência de Costa, Acencio e Lemke (2013) (CAL) atualizado para o estudo do comportamento das RNAPs durante a transcrição múltipla da sequência presente no operon ribossomal de *E. coli*. Esse modelo baseia-se em simulações estocásticas, dependentes dos valores de energia livres encontrados para diferentes conformações do complexo de alongamento transcricional, e apresenta boa aderência nas sequências onde foi testado, principalmente nos resultados das simulações de transcrição múltipla.

## 2 Objetivos

- Determinar e implementar de um modelo *in silico* para as reações que representam a ligação da RNAP em promotores, utilizando dados da literatura como referência;
- Integrar o modelo estocástico dependente de sequência para a transcrição múltipla e o modelo implementado do item anterior;
- Utilizar o modelo integrado, prever o comportamento das RNAPs e do promotor em sequências de DNA ribossomal de *E. coli*;
- Confrontar os resultados obtidos *in silico* com os experimentais.

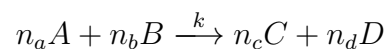
## 3 Métodos

Utilizamos uma abordagem semelhante à proposta por Klumpp e Hwa (2008) e Costa, Acencio e Lemke (2013): representamos o ancoramento da RNAP no promotor e a fase de alongamento da transcrição através de reações químicas. A ligação da RNAP no promotor depende, portanto, da taxa de ocorrência dessas reações na presença ou na ausência de fatores de transcrição e pequenos ligantes (BROWNING; BUSBY, 2004; BALLEZA et al., 2009); a incorporação de cada nucleotídeo depende da sequência a ser transcrita e de cada RNAP na fita de DNA em situações normais de alongamento e em *back* e *forward tracking*. Identificamos os sítios de pausas e as colisões entre as RNAPs, determinando o tempo de permanência e predizendo ocorrência de transcrição abortiva ou a aceleração da transcrição devido ao fenômeno cooperativo entre múltiplas RNAPs. A sequência do operon ribossomal *rrnB* (Genbank U00096.3) foi baixada do Ecogene3.0 (ZHOU; RUDD, 2013).

### 3.1 Simulação Estocástica

#### 3.1.1 Reações químicas

Reações químicas são a linguagem canônica para modelos biológicos. Expressam processos químicos complexos tanto qualitativamente como quantitativamente. Especificá-las corretamente é fundamental para o sucesso do modelo. Genericamente,



representa que  $n_a$  moléculas do tipo  $A$  reagem com  $n_b$  do tipo  $B$  para produzirem  $n_c$  moléculas do tipo  $C$  e  $n_d$  de  $D$ . A letra  $k$  acima da seta, que indica o sentido da reação, representa a taxa de ocorrência da mesma, ou seja, qual a probabilidade que a colisão entre  $n_a$  moléculas de  $A$  com  $n_b$  de  $B$  ocorra com energia suficiente para que reajam entre si e gerem seus produtos.

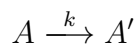
#### 3.1.2 O conceito de estado

O *estado* de um determinado sistema mostra condição do mesmo num instante definido, e contém informação suficiente para prever seu comportamento no futuro. Intuitivamente, trata-se de um conjunto de variáveis que precisam ser acompanhadas num determinado modelo. Na modelagem estocástica, o estado depende de uma lista do

número atual de moléculas de cada tipo, chamada de *configuração do sistema*, podendo ser a própria configuração, ou a distribuição de probabilidades atual da configuração.

### 3.1.3 Mudança de estado

A reação



é lida como duas espécies químicas,  $A$  e  $A'$ , sendo  $A$  transformado em  $A'$  a uma taxa  $k$ . Ou seja, representa como e quão rápido ocorre a alteração entre estados do sistema. No modelo estocástico, essa mudança de estado ocorre da seguinte forma: uma única molécula de  $A$  é convertida numa molécula de  $A'$ ; o número total de elementos de  $A$  decresce em uma unidade e o de  $A'$  aumenta em uma. A probabilidade desse evento ocorrer num intervalo de tempo diferencial  $dt$  é dada pelo produto entre a taxa  $k$ , o número de moléculas de  $A$ , e pelo tempo  $dt$ .

### 3.1.4 Algoritmo de Gillespie

Nesse algoritmo, para cada período de tempo o sistema está em um único estado exato. Uma transição consiste em executar uma reação  $r$ , sendo  $n$  o número máximo de possíveis transições a partir de um estado. Utilizando-se geradores de números aleatórios com a distribuição de probabilidade correta, o algoritmo determina qual reação será executada.

Para um sistema em um determinado estado, o método direto retorna qual será a próxima reação a ocorrer e quando ela irá ocorrer. Estas questões podem ser respondidas probabilisticamente especificando a função densidade de probabilidade  $P(r, t)$  para a próxima reação  $r$  que ocorre no tempo  $t$  (GIBSON; BRUCK, 2000). Pode ser mostrado que

$$P(r, t)dt = a_r \text{Exp}\left(-t \sum_{j=1}^n a_j\right)dt,$$

onde  $a_r$  é proporcional a probabilidade de uma certa reação  $r$  ocorrer. A seguir o Algoritmo 1 apresenta os passos para calcular qual será a próxima reação e quando ela irá ocorrer usando a função densidade de probabilidades das reações. O método da primeira reação, apresentado no Algoritmo 2, calcula um tempo estimado  $t_i$  para todas as reações possíveis e seleciona qual é o menor dentre eles, atribuindo a  $r$  essa reação e a  $t$  o seu tempo.



**Algoritmo 1: Método direto**

1. Inicializa, determine o tempo ( $T$ ) e número de moléculas de cada composto inicial.
2. Calcule a função densidade de probabilidade  $a_i$ , para todas as reações.
3. Escolha  $r$  de acordo com a distribuição:

$$P_r(r) = \frac{a_r}{\sum_{j=1}^n a_j}.$$

4. Escolha  $t$  de acordo com uma distribuição exponencial:

$$P(t) = \left( \sum_{j=1}^n a_j \right) \text{Exp} \left( -t \sum_{j=1}^n a_j \right) dt.$$

5. Altere o número de moléculas para refletir a execução da reação  $r$  e o tempo para  $T = T + t$ .
6. Volte para o passo 2.

**Algoritmo 2: Método da primeira reação**

1. Inicialize, determine o tempo ( $T$ ) e número de moléculas de cada composto inicial.
2. Calcule a função densidade de probabilidade  $a_i$  para todas as reações.
3. Calcule o tempo estimado  $t_i$  para cada reação de acordo com uma distribuição exponencial com parâmetro  $a_i$ :

$$t_i = \frac{1}{a_i} \log \left( \frac{1}{\text{rand}} \right).$$

4. Atribua a  $r$  a reação com o menor  $t_i$ .
5. Atribua a  $t$  o valor de  $t_r$ .

**3.1.5 Cinética enzimática**

A cinética enzimática estuda a probabilidade de uma reação química catalizada por uma enzima ocorrer. Essas enzimas possuem a capacidade de se ligar ao substrato e aumentar a velocidade da reação. Grande parte das reações que ocorrem em ambiente celular possui uma enzima responsável para tornar sua velocidade compatível com a vida. Basicamente, o conjunto de enzimas disponíveis numa célula determina seu metabolismo. Cerca de 3700 enzimas já foram caracterizadas (BAIROCH, 2000). Esta reação pode ser regulada por moléculas inibidoras que diminuem a atividade enzimática, ou por moléculas

ativadoras capazes de aumentá-la. Além dessas a temperatura, pressão, pH e concentração dos substratos e enzimas podem afetar a velocidade de ocorrência da reação.

Toda reação química é determinada pelas leis da Termodinâmica. Tratamos uma reação como *espontânea* quando não há necessidade de absorção de energia para sua ocorrência ou como *não espontânea*, quando sua ocorrência depende da absorção de energia na forma de calor, luz, por exemplo. Essa energia é chamada de energia de ativação. A catalização de uma reação leva à diminuição de sua energia de ativação, como podemos ver na Figura 9. Como a energia necessária para a ocorrência da reação diminui, a catalização leva a sua aceleração. Algumas enzimas chegam a acelerar milhões de vezes uma reação (RADZICKA; WOLFENDEN, 1995)

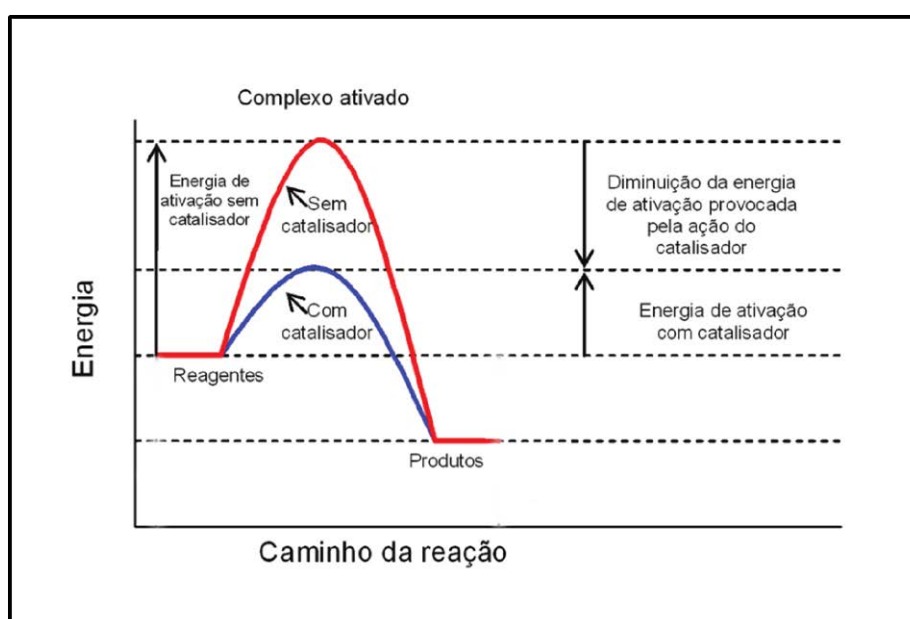
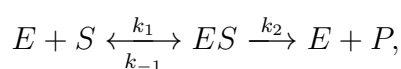


Figura 9 – Gráfico representativo do caminho da reação. A energia de ativação representa a barreira energética que necessita ser ultrapassada para sua ocorrência. Note a alteração de seu valor na presença de um catalisador.

### 3.1.6 Cinética de Michaelis-Menten

A cinética de Michaelis-Menten é uma equação matemática descrita em 1913 pelo bioquímico Leonor Michaelis e pelo físico Maud Menten baseado na reação de hidrólise da sacarose em glicose e frutose pela enzima invertase (MENTEN; MICHAELIS, 1913). Nesta equação, a enzima livre E liga-se ao substrato S e forma um complexo ES, convertido num produto P mais a enzima. Podemos representar esta reação da seguinte forma:



onde  $k_1$ ,  $k_{-1}$  e  $k_2$  são as taxas de ocorrência para cada reação e as setas indicam o sentido da reação, podendo ocorrer uma reação reversível da ligação enzima-substrato.

Se assumirmos que a concentração enzimática é muito menor que a concentração de substrato, que a variação da concentração do complexo ES pode ser tomada como zero e que a concentração total da enzima não varia com o tempo, a taxa de formação do produto, ou velocidade global da reação, é dada por

$$v = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]}.$$

Logo, a taxa dessa reação aumenta assintoticamente com o aumento da concentração do substrato S até o valor máximo de  $V_{max}$ , obtido quando todo o substrato estiver ligado a uma enzima. A constante de Michaelis-Menten,  $K_m$ , é dada pelo inverso da afinidade do substrato com a enzima, ou seja,

$$K_m = \frac{k_{-1} + k_2}{k_1}.$$

Seu valor representa a concentração de substrato necessária para que a reação ocorra a uma velocidade igual a metade da velocidade máxima (LENINGER, 2006). Os valores de  $K_m$  e de  $V_{max}$  podem ser determinados a partir de uma série de ensaios enzimáticos, realizados variando-se a concentração dos substratos.

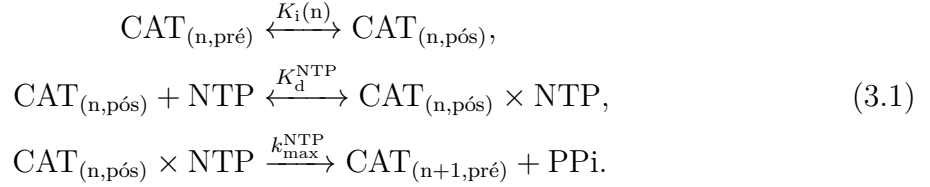
### 3.1.7 Reação enzimática com inibidor

Inibidores enzimáticos são moléculas que reduzem ou eliminam a atividade das enzimas. Na presença destes inibidores caracterizam dois tipos de reações: as reversíveis e as irreversíveis (LENINGER, 2006). Nas reações irreversíveis o inibidor apresenta uma constante de dissociação tão baixa que inativa a enzima permanentemente. Os inibidores reversíveis podem ser classificados em 4 tipos:

1. **Competitivo:** Neste caso, o inibidor e o substrato não podem se ligar simultaneamente na mesma enzima e competem pelo seu sítio ativo. Este tipo de inibidor não altera o valor de  $V_{max}$ , mas aumenta o valor de  $K_m$  aparente,  $K_m^{app}$ .
2. **Incompetitivo:** Quando o inibidor se liga apenas ao complexo enzima-substrato. O valor de  $V_{max}$  diminui, assim como o  $K_m^{app}$ .
3. **Misto:** O inibidor misto pode se ligar tanto a enzima como ao complexo enzima-substrato. Com isso, o valor de  $V_{max}$  diminui e o valor de  $K_m$  aparente aumenta.
4. **Não competitivo:** A inibição não competitiva é uma forma de inibição mista. Neste caso, se o inibidor estiver ligado à enzima, sua atividade decai, mas sem afetar a taxa de ligação do substrato a mesma. Como resultado, o valor de  $V_{max}$  diminui enquanto o  $K_m^{app}$  se mantém.

## 3.2 O modelo de Costa et al. (CAL)

O modelo implementado por Costa, Acencio e Lemke (2013) baseia-se no trabalho de Bai, Shundrovsky e Wang (2004). Cada passo do alongamento foi representado como uma reação em três etapas:



O modelo inclui o deslocamento do CAT (complexo de alongamento transcricional), a ligação de um novo nucleotídeo (NTP) e sua catálise química. A variável  $n$  é o comprimento do RNA e os estados *pré* e *pós* para CAT representam a posição do RNA nascente em relação ao sítio ativo da enzima.  $K_d^{\text{NTP}}$  e  $k_{\text{max}}^{\text{NTP}}$  possuem valores experimentais estabelecidos e dependem do nucleotídeo que será incorporado (Tabela 1).  $K_i(n)$  é dado por

$$K_i(n) = \exp[(\Delta G_{(n,\text{pós})} - \Delta G_{(n,\text{pré})} - Fd)/k_B T], \tag{3.2}$$

onde  $\Delta G_{(n,m)}$  é a energia livre de Gibbs para determinada conformação do CAT, foi baseado no método dos vizinhos mais próximos apresentado por SantaLucia, Allawi e Seneviratne (1996),  $F$  representa uma força externa aplicada na RNAP e  $d$  é a distância entre dois pares de base de nucleotídeos na fita de DNA (aproximadamente 0,34 nm) (BAI; FULBRIGHT; WANG, 2007).  $m$  representa a posição do sítio ativo da RNAP (*pré*=0 e *pós*=1). Para obtenção da taxa de incorporação de cada ribonucleotídeo, utilizou-se da cinética de Michaelis-Menten na presença de inibidor competitivo (GUAJARDO; SOUSA, 1997):

$$k_{\text{main}}(n) = \frac{k_{\text{max}}^{\text{NTP}} [\text{NTP}]}{K_d^{\text{NTP}} \{1 + K_i(n)\} + [\text{NTP}]} \tag{3.3}$$

e a taxa de *backtracking/forwardtracking* será dada por

$$k_{n,m \rightarrow m \pm 1} = k_0 \exp[(\Delta G^\ddagger - \Delta G_{(n,m)} + Fd(F))/k_B T], \tag{3.4}$$

onde  $k_0$  é um pré fator constante ( $1,0 \times 10^9 \text{ s}^{-1}$ ) e  $\Delta G^\ddagger$  é a barreira energética ( $41, 2k_B T$ ). Portanto o algoritmo base para simulação da transcrição considera que a RNAP pode tomar dois caminhos distintos: enzima pode estar em alongamento normal, ou em *backtracking*. Quando em alongamento normal, calcula-se o valor de  $k_{\text{main}}$  utilizando a Equação (3.3) considerando-se um híbrido RNA-DNA com 9 nt antes do sítio ativo da enzima ser desocupado, e uma bolha de transcrição com 15 nt. Determina-se também um valor para  $k_{n,m \rightarrow m \pm 1}$  ( $k_{\text{back}}$ ) a partir da Equação (3.4), para a mesma conformação do CAT. Para ambos os casos,  $F = 0$  pN. Se a RNAP entrou em *back tracking*, calcula-se apenas

Tabela 1 – Valores experimentais para  $k_{max}^{NTP}$  e  $K_d^{NTP}$ . Cada nucleotídeo possui seu valor e foram determinados experimentalmente por Bai, Fulbright e Wang (2007).

|                 | ATP  | UTP  | GTP   | CTP  |
|-----------------|------|------|-------|------|
| $k_{max}^{NTP}$ | 38±7 | 24±4 | 62±18 | 7±4  |
| $K_d^{NTP}$     | 50±6 | 18±1 | 36±5  | 33±6 |

um valor,  $k_{back}$ , tanto para o movimento no sentido  $5' \rightarrow 3'$ , como para o movimento no sentido oposto.

Para a transcrição múltipla, determina-se durante a simulação uma equação para posição de cada polimerase em função do tempo, podendo ou não ocasionar um choque entre duas polimerases. Quando há choques, a RNAP traseira aplicará uma força  $F$  na RNAP dianteira alterando momentaneamente os valores das taxas de ocorrência das enzimas, pois o valor de  $F$  nas Equações (3.3) e (3.4) deixará de ser nulo.

Com base no tempo e posição de cada RNAP nas simulações podemos calcular a velocidade da transcrição através da equação:

$$v_m = \frac{n_{seq}}{t_{seq}}, \quad (3.5)$$

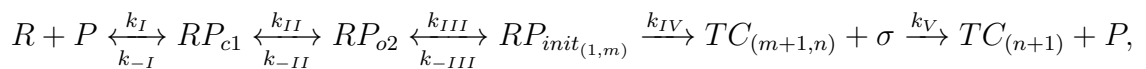
onde  $v_m$  é a velocidade média da transcrição em nucleotídeo por segundo (nt/s),  $n_{seq}$  é o tamanho da sequência em nucleotídeo e  $t_{seq}$  é o tempo em que levou para transcrever a sequência em segundos (s). Determinamos a densidade de RNAPs em um determinado trecho baseado no tempo, ou seja, quando uma RNAP estiver em uma posição  $x$  no tempo  $y$  verificamos quantas RNAPs possuem anteriormente a essa posição e com um tempo menor que  $y$ .

Além disso, o modelo permite a variação de alguns parâmetros de entrada para a realização das simulações como:

- Número de moléculas de RNAP ( $R$ );
- Força de interação entre RNAPs, dada em pN ( $F$ );
- Concentração de nucleotídeos dada em micromolar ( $\mu\text{M}$ )( $[NTP]$ ).

### 3.3 Modelo para ligação da RNAP na sequência promotora

O modelo que será implementado está baseado no modelo proposto por Dennis, Ehrenberg e Bremer (2004). O modelo é descrito pelas reações:



onde  $R$  representa a RNAP,  $P$  é o promotor,  $RP_{c1}$  representa a formação do complexo fechado RNAP-promotor,  $RP_{o2}$  é o complexo binário aberto,  $RP_{init(1,m)}$  representa o

complexo ternário iniciado incluindo a iniciação abortiva, onde o transcrito é menor que  $m$  ( $m = 10$  nucleotídeos),  $TC_{(m+1,n)} + \sigma$  representam a formação do complexo de alongamento entre a fase de iniciação e a fase de alongamento com a liberação do fator sigma,  $\sigma$ , onde  $n$  é igual a 50 nucleotídeos,  $TC_{(n+1)} + P$  é o complexo de alongamento após a regeneração do promotor e  $k_n$  são as taxas de ocorrências das reações.

Para as simulações, consideramos todos os passos da iniciação e para isso, calculamos suas constantes seguindo as equações:

$$k_I = \frac{n_{c1}}{(\tau_{RP_{c1}}[Rf])}, \quad k_{II} = \frac{1}{\tau_{RP_{o2}}}, \quad k_{III} = \frac{1}{\tau_{RP_{init}(1,m)}}, \quad k_{IV} = \frac{1}{\tau_{TC_{(m+1,n)}}} \quad (3.6)$$

$$k_V = \frac{1}{\tau_{TC_{50}}}, \quad k_{-I} = \frac{1}{\tau_{R+P}}, \quad k_{-II} = \frac{1}{\tau_{RP_{c1}}}, \quad k_{-III} = \frac{1}{\tau_{RP_{o2}}},$$

onde  $\tau_n$  é o tempo de ocorrência para cada reação,  $n_{c1}$  é o número médio de vezes em que a RNAP se liga ao promotor por iniciação e  $[Rf]$  é a concentração de RNAP livre.

## 3.4 Simulações da fase de alongamento da transcrição

As reações químicas foram simuladas a partir do algoritmo de Gillespie (GILLESPIE, 1977), que apresenta uma boa relação entre custo computacional, complexidade matemática e realismo biológico. Instalamos um pacote para o *software Mathematica*® com as funções necessárias para o algoritmo (SHAPIRO et al., 2003) e demos início as simulações.

### 3.4.1 Transcrição do DNA ribossomal considerando diferentes parâmetros

A partir dos modelos citados no item 1.5, inicialmente realizamos simulações com o modelo CAL para transcrição múltipla da região inicial do operon *rrnB* e do gene RNAr 5S, e estudamos o comportamento cinético da RNAP, alterando os parâmetros de entrada do algoritmo conforme a Tabela 2 mantendo um valor fixo para a força de interação entre RNAP de 12,5 pN e verificamos a velocidade em que ocorre a transcrição (Eq.3.5).

Tabela 2 – Parâmetros de entrada para simulação da fase de alongamento.  $[\text{NTP}]$  é a concentração de nucleotídeo trifosfato em micromolar ( $\mu\text{M}$ ),  $\mathbf{R}$  é o número de RNA polimerases na fase de alongamento,  $n$  é o número de simulações realizadas e **sequência** é a sequência de DNA transcrita. **Seq1** representa 396 nucleotídeos iniciais do operon ribossomal e **Seq5S** é a sequência do gene RNAr 5S.

| $\mathbf{R}$ | $[\text{NTP}](\mu\text{M})$ | $n$  | <b>sequência</b> |
|--------------|-----------------------------|------|------------------|
| 5            | 400                         | 1200 | Seq1             |
| 10           | 400                         | 1200 | Seq1             |
| 15           | 400                         | 1200 | Seq1             |
| 5            | 800                         | 1200 | Seq1             |
| 10           | 800                         | 1200 | Seq1             |
| 15           | 800                         | 1200 | Seq1             |
| 5            | 1600                        | 1200 | Seq1             |
| 10           | 1600                        | 1200 | Seq1             |
| 15           | 1600                        | 1200 | Seq1             |
| 5            | 400                         | 1200 | Seq5S            |
| 10           | 400                         | 1200 | Seq5S            |
| 15           | 400                         | 1200 | Seq5S            |
| 5            | 800                         | 1200 | Seq5S            |
| 10           | 800                         | 1200 | Seq5S            |
| 15           | 800                         | 1200 | Seq5S            |
| 5            | 1600                        | 1200 | Seq5S            |
| 10           | 1600                        | 1200 | Seq5S            |
| 15           | 1600                        | 1200 | Seq5S            |

Posteriormente, com base nos dados de Quan et al. (2005), Dennis et al. (2009) e Fange et al. (2014), realizamos simulações da sequência do operon *rrnB* completo e dividido em 20 seções, variando novamente os parâmetros de entrada de acordo com a Tabela 3 para verificar a acurácia e custo computacional do modelo CAL. Para quantificar a aderência entre os dados de nossas simulações e os resultados experimentais calculamos:

- Densidade relativa de RNAP;
- Erro quadrático médio (RMSE);
- Velocidade média das RNAPs;
- Função de densidade acumulada (CDF);
- Área da função de densidade acumulada;
- Relação do erro quadrático médio e área da função de densidade acumulada;
- Pausas transcricionais em função do tempo.

Primeiramente, calculamos a densidade relativa de RNAPs das simulações e por meio do  $p$ -value e do RMSE encontramos quais os melhores parâmetros a serem utilizados para obtermos uma maior correlação entre o dado teórico e o experimental. Estes parâmetros foram apresentados através de um *Heat map* ( $\mathbf{F}$ ,  $[\mathbf{NTP}]$  e RMSE). Neste caso foram considerados estatisticamente válidos resultados com  $p$ -value  $< 0,05$  e com um valor do RMSE  $< 9,2$  apresentado por Fange et al. (2014). Calculamos a velocidade média das RNAPs segundo a formula 3.5, para os primeiros 100, 2800 e 5600 nucleotídeos, afim de verificarmos o comportamento de cada uma e determinamos qual seria mais apropriada para análise posterior. A função de densidade acumulada foi calculada para todos os valores (sequência completa, particionada e experimental), pois ela é mais robusta e depende menos das flutuações estatísticas que ocorrem experimentalmente e em nossas simulações. Além disso para verificar se as nossas simulações (seq. completa) apresentaram valores próximos ao experimental, analisamos a frequência das simulações relacionadas ao RMSE, CDF e por fim uma relação do RMSE com a área da função de densidade acumulada. Determinamos também as pausas transcricionais, relacionando o tempo em que se mantinham em uma certa posição na presença ou não do *backtracking*.

Tabela 3 – Parâmetros utilizados nas simulações após ajustes no modelo.  $[\mathbf{NTP}]$  é a concentração de nucleotídeos em micromolar( $\mu M$ ),  $\mathbf{F}$  é a força de interação entre duas RNA polimerases durante a colisão dada em pN e  $\mathbf{R}$  é o número de RNA polimerases na fase de alongamento.

| Parâmetros              | simulações  |
|-------------------------|-------------|
| $[\mathbf{NTP}](\mu M)$ | [100; 1400] |
| $\mathbf{F}(\text{pN})$ | [2.5; 50]   |
| $\mathbf{R}$            | [50;100]    |



## 4 Resultados e Discussão

### 4.1 Resultados e análise das simulação da Fase de Iniciação

As taxas de ocorrência foram calculadas segundo Dennis, Ehrenberg e Bremer (2004) com base no tempo em que cada uma ocorre (Tabela 4).

Tabela 4 – Taxas de ocorrência calculadas para simulações da fase de iniciação da transcrição. Constantes baseadas em dados experimentais, dada em  $s^{-1}$ .

| Constantes | valor ( $s^{-1}$ ) | Constantes | valor ( $s^{-1}$ ) |
|------------|--------------------|------------|--------------------|
| $k_I$      | 2,85               | $k_{-I}$   | 100                |
| $k_{II}$   | 500                | $k_{-II}$  | 100                |
| $k_{III}$  | 500                | $k_{-III}$ | 0,01               |
| $k_{VI}$   | 9,90               |            |                    |
| $k_V$      | 2,27               |            |                    |

As reações do processo de iniciação da transcrição foram simuladas 1200 vezes com o promotor apresentando uma média de 57 iniciações por minuto. A média foi encontrada a partir da multiplicação do resultado da simulação por 60 (uma iniciação = 0,94 segundos)  $0,94 \times 60 = 56,75$  iniciações por minuto. Este resultado está condizente com a literatura que apresenta uma média de 60 iniciações por minuto.

### 4.2 Resultados e análise das simulações da Fase de Alongamento

Para a fase de alongamento do RNA ribossomal 5S e da região inicial do operon *rrnB* foram realizadas 1200 simulações para cada caso presente na Tabela 2. Como resultado, obtivemos o tempo médio para a transcrição e a partir desses dados calculamos sua velocidade média em nucleotídeos por segundo (Equação 3.5). Os resultados das simulações para ambas as sequências foram similares, conforme a Tabela 5.

Tabela 5 – Velocidade média de incorporação de nucleotídeo na fase de alongamento da transcrição.  $[\text{NTP}]$  é a concentração de nucleotídeo trifosfato,  $\mathbf{R}$  é o número de RNAPs, **tempo médio** é o tempo médio que a RNAP levou para transcrever a sequência em segundo ( $s$ ) e  $v_m$  (**nt/s**) é a velocidade média da transcrição em nucleotídeo por segundo (número de nt/tempo médio).

| <b>seq1.1</b> ( $[\text{NTP}] = 400 \mu\text{M}$ )  |                       |                       | <b>seq5S.1</b> ( $[\text{NTP}] = 400 \mu\text{M}$ )  |                       |                       |
|---|-----------------------|-----------------------|--|-----------------------|-----------------------|
| $\mathbf{R}$  | <b>tempo médio(s)</b> | $v_m$ ( <b>nt/s</b> ) | $\mathbf{R}$   | <b>tempo médio(s)</b> | $v_m$ ( <b>nt/s</b> ) |
| 5   | 56,44                 | 7,01                  | 5  | 25,65                 | 7,52                  |
| 10  | 56,43                 | 7,01                  | 10   | 27,50                 | 7,01                  |
| 15  | 56,62                 | 6,99                  | 15   | 23,73                 | 8,13                  |
| <b>seq1.2</b> ( $[\text{NTP}] = 800 \mu\text{M}$ )  |                       |                       | <b>seq5S.2</b> ( $[\text{NTP}] = 800 \mu\text{M}$ )  |                       |                       |
| $\mathbf{R}$  | <b>tempo médio(s)</b> | $v_m$ ( <b>nt/s</b> ) | $\mathbf{R}$   | <b>tempo médio(s)</b> | $v_m$ ( <b>nt/s</b> ) |
| 5   | 34,07                 | 11,62                 | 5  | 15,90                 | 12,13                 |
| 10  | 34,24                 | 11,56                 | 10   | 14,56                 | 13,25                 |
| 15  | 31,37                 | 12,62                 | 15   | 15,50                 | 12,44                 |
| <b>seq1.3</b> ( $[\text{NTP}] = 1600 \mu\text{M}$ ) |                       |                       | <b>seq5S.3</b> ( $[\text{NTP}] = 1600 \mu\text{M}$ ) |                       |                       |
| $\mathbf{R}$  | <b>tempo médio(s)</b> | $v_m$ ( <b>nt/s</b> ) | $\mathbf{R}$   | <b>tempo médio(s)</b> | $v_m$ ( <b>nt/s</b> ) |
| 5   | 22,60                 | 17,52                 | 5  | 9,95                  | 19,30                 |
| 10  | 22,94                 | 17,25                 | 10   | 10,17                 | 18,96                 |
| 15  | 23,19                 | 17,07                 | 15   | 9,54                  | 20,21                 |

Com base nestes dados verificamos que com uma força de 12,5 pN a concentração de nucleotídeos influencia diretamente na velocidade da transcrição, podendo duplicar a taxa de incorporação de nucleotídeos, porém não houve uma diferença significativa na velocidade com o aumento no número de RNAPs.

Posteriormente, realizamos as simulações utilizando a sequência do operon ribossomal *rrnB* completa e particionada em 20 segmentos. Para essas simulações, cruzamos todos os parâmetros estipulados na Tabela 3 e através de um *Heat map* ( $\mathbf{F}$ ,  $[\text{NTP}]$  e RMSE) identificamos que com uma  $\mathbf{F}$  de aproximadamente de 15 a 25 pN e  $[\text{NTP}]$  de 100  $\mu\text{M}$  obtivemos melhores resultados e com valores estatísticos significativos ( $p\text{-value} < 0,05$ ) (Figura 10).

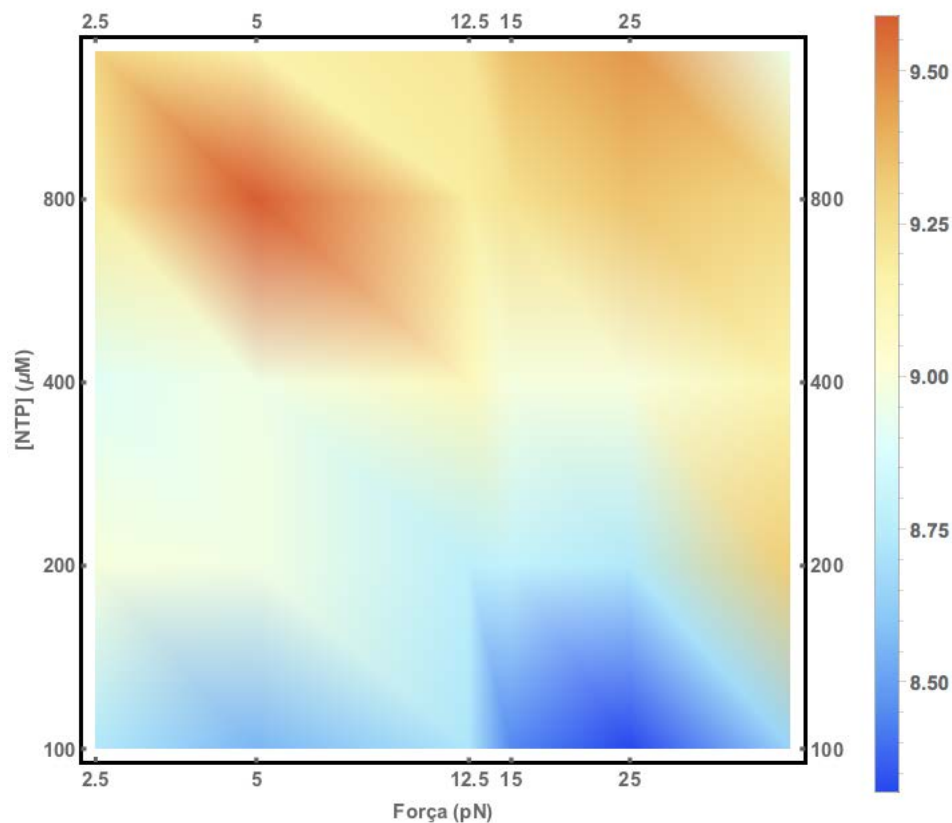


Figura 10 – *Heat map* das simulações utilizando o operon ribossomal *rrnB* completo. Podemos observar um melhor resultado para uma  $F$  de aproximadamente de 15 a 25 pN e  $[NTP]$  de  $100 \mu M$ . Eixo das ordenadas:  $[NTP]$ , eixo das abscissas: **Força de interação entre RNAP** e a escala varia entre o valor mínimo e máximo para o RMSE (8,32 a 9,59).

A partir desses valores, aumentamos o número de simulações e calculamos a velocidade média de cada RNAP e verificamos que a terceira RNAP já apresentava um comportamento característico do sistema simulado (Figura 11). Levando isso em consideração, as seguintes análises foram baseadas na terceira RNAP, ou seja, quando essa chegar ao fim da sequência paramos a simulação e verificamos as posições das RNAPs posteriores, descartando assim a primeira e a segunda RNAP. Em comparação entre os dois modos simulados, o tempo de transcrição da sequência completa é aproximadamente 146 segundos mais devagar que a transcrição para sequência particionada, ou seja, observamos uma diferença de 12% entre esses casos.

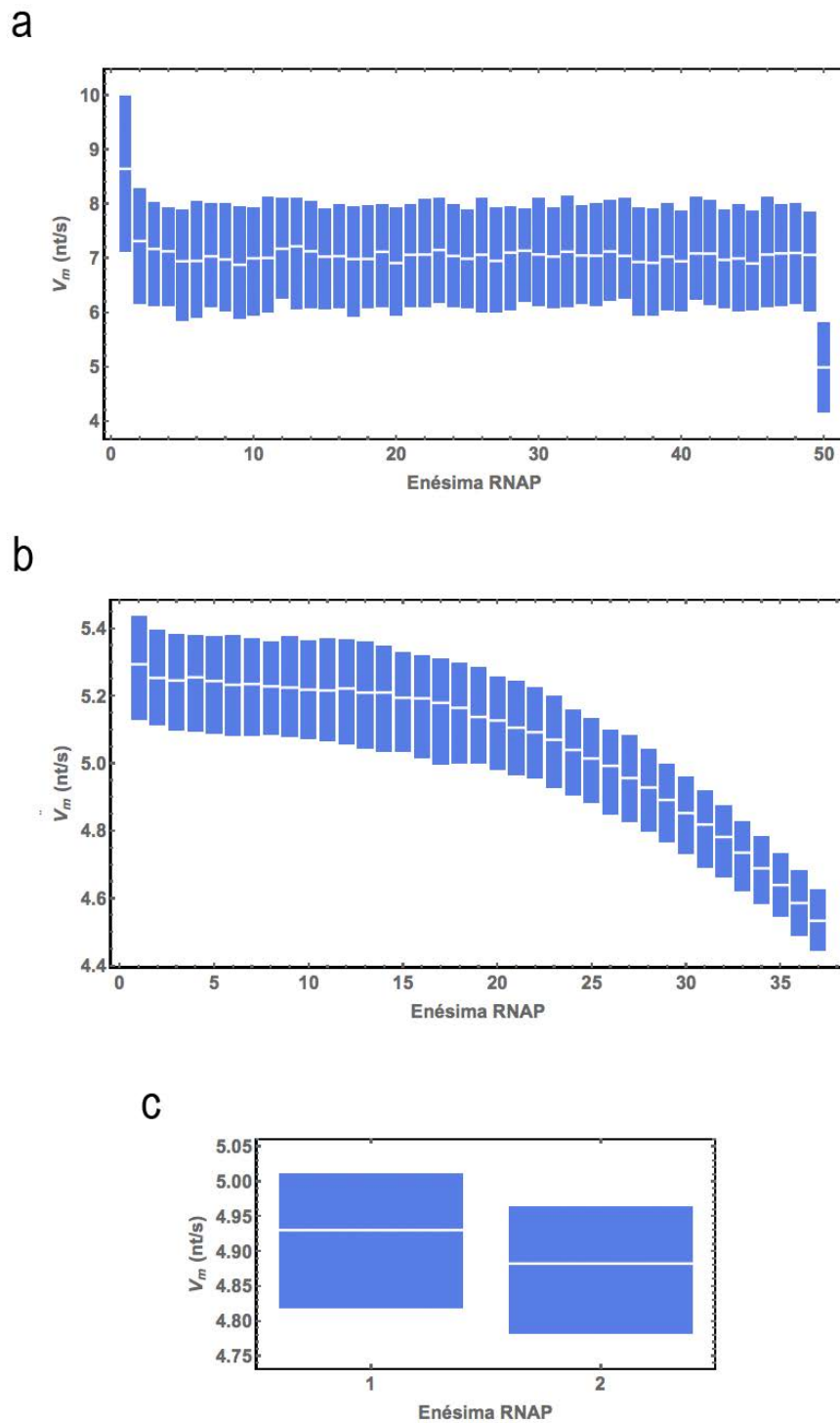


Figura 11 – Gráfico representativo da velocidade média das RNAPs em nucleotídeos por segundo. Os limites inferior e superior da barra representam o primeiro e o terceiro quartil, respectivamente. A mediana está destacada em branco. **a)** Velocidade média das RNAPs para os primeiros 100 nt. **b)** Velocidade média para os primeiros 2800 nt. **c)** Velocidade média das duas RNAPs que terminaram a transcrição da sequência completa.

Foram calculadas as densidades relativas das RNAPs para as sequências completa e particionada e verificamos que as simulações utilizando a sequência completa apresentaram um melhor resultado com um menor valor para o RMSE: 8,20 para completa e 9,62 para a particionada. Além disso, comparando a curva do CDF, os valores para a sequência completa também foram melhores, apresentando uma maior semelhança com o experimental e diferindo da distribuição constante. Entretanto, ainda não apresentando similaridade nos segmentos 12 ao 19 em relação ao experimental (Figura 12).

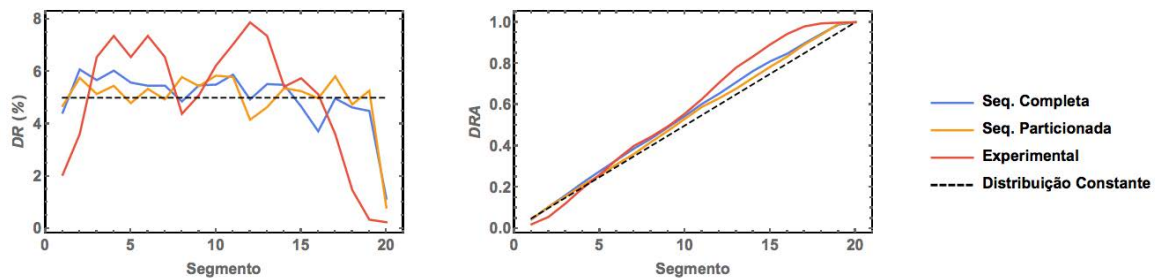


Figura 12 – Representação gráfica da densidade relativa de RNAP da simulação do operon *rrnB*. Gráfico à esquerda: flutuações da densidade relativa de RNAP em cada segmento na transcrição do operon *rrnB*. Para as simulações da sequência completa (azul) obtivemos um valor de 8,20 para o RMSE e 9,62 para a sequência particionada (laranja). Eixo das ordenadas: densidade relativa de RNAP e eixo das abscissas: segmentos da sequência simulada. Gráfico à direita: representação da função de densidade acumulada para as simulações. Eixo das ordenadas: área da densidade relativa de RNAP e eixo das abscissas: Segmentos da sequência simulada. Aderência utilizada nas simulações para ambas as sequências:  $\mathbf{F}$  de 15pN, 50  $\mathbf{R}$ ,  $[\mathbf{NTP}]$  de 100  $\mu\text{M}$ . Dado experimental determinado por Quan et al. (2005).

Dessa forma para comparar os resultados com os dados experimentais, utilizamos as simulações com a sequência completa. No entanto, apesar de nossos dados não apresentarem uma grande flutuação como demonstrado por Quan et al. (2005), principalmente nos segmentos 1, 4, 6 e 12, nossos resultados foram representativos e com melhores valores para o RMSE do que apresentado por Fange et al. (2014) (Tabela 6). Essas flutuações presentes nos dados experimentais provavelmente são ocasionadas pela alta taxa de iniciação e número de RNAPs livres no sistema, diminuindo a densidade no início, e pela formação de estruturas secundárias nas regiões de DNAr 16S (segmentos 2-7) e 23S (segmentos 9-19) interagindo com a RNAP, podendo retardá-la e promover um aumento da densidade nesses segmentos (DENNIS et al., 2009; FANGE et al., 2014). Entretanto, considerando que os valores experimentais foram obtidos basicamente pela análise da densidade de apenas 11 operons ribossomais, fizemos uma comparação de 1:1, ou seja, foi comparada cada simulação com os valores experimentais e verificamos a acurácia através do RMSE e CDF. Em relação ao RMSE obtivemos um melhor resultado com um valor igual a 6,4 (RMSE) e para a curva da CDF observamos uma aproximação ao experimental com uma pequena diferença no fim relacionada a uma maior flutuação das densidades na região (Figura 13).

Tabela 6 – Comparação do valor do RMSE entre as simulações utilizando seqüência completa, particionada, de Fange et al. (2014) e valor aleatório baseado nos dados de Quan et al. (2005).

| dados               | RMSE (% densidade) |
|---------------------|--------------------|
| seq. completa       | 8,20               |
| seq. particionada   | 9,62               |
| Quan <i>et al.</i>  | 10,41              |
| Fange <i>et al.</i> | 9,2                |

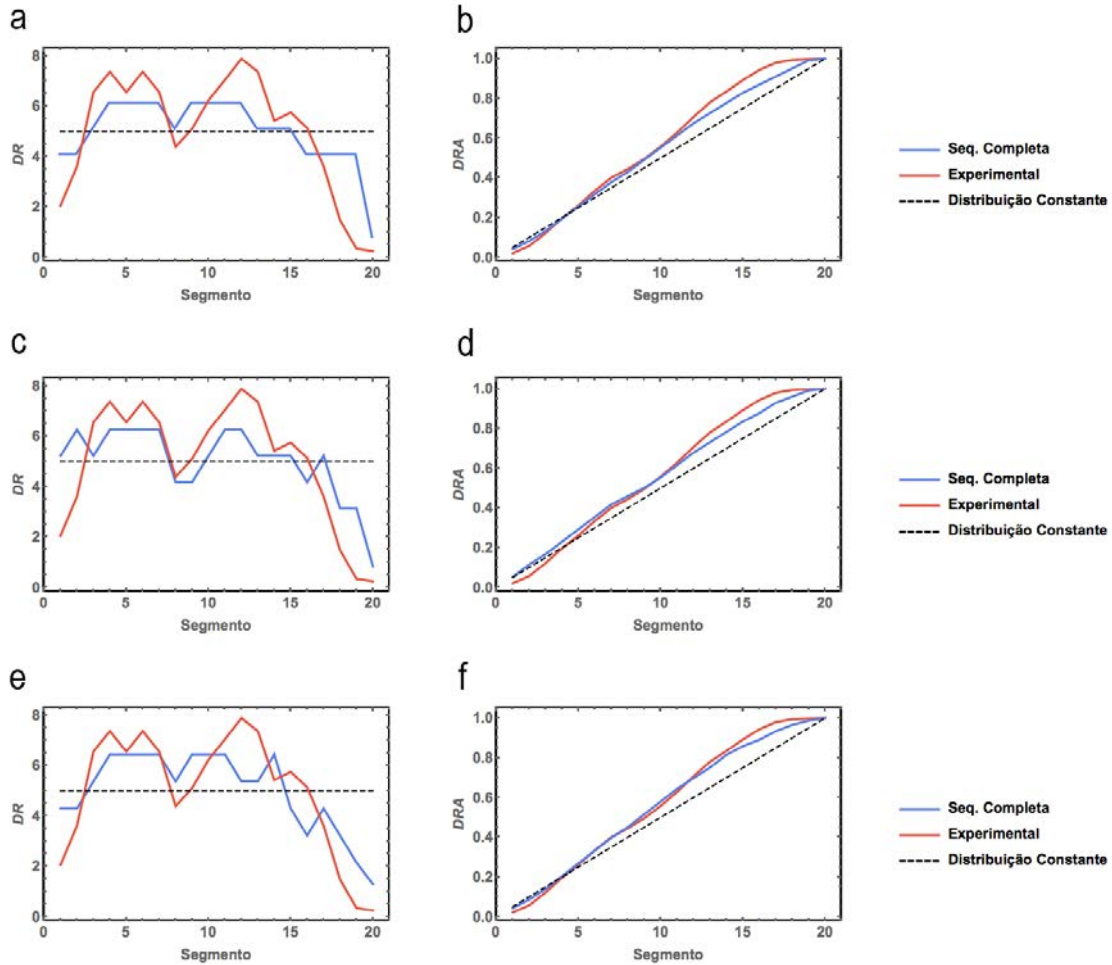


Figura 13 – Representação gráfica da densidade relativa de RNAP e CDF da simulação do operon *rrnB*. Gráficos **a**, **c** e **e** apresentam as flutuações da densidade relativa de RNAP em cada segmento na transcrição do operon *rrnB* e possuem os melhores valores para o RMSE: 6,66; 6,92; 6,4 respectivamente. Eixo das ordenadas: densidade relativa de RNAP e eixo das abscissas: segmentos da seqüência simulada. Gráficos **b**, **d** e **f** representam a função de densidade acumulada para as simulações equivalentes ao gráfico do lado. Podemos observar a semelhança das curvas teórica e experimental com uma leve diferença em uma região específica no fim da seqüência. Eixo das ordenadas: área da densidade relativa de RNAP e eixo das abscissas: segmentos da seqüência simulada. Aderência utilizada nas simulações para ambas as seqüências: **F** de 15pN, 50 **R**, [NTP] de 100  $\mu$ M. Dado experimental determinado por Quan et al. (2005).

Para assegurar que nosso resultado esteja correto verificamos que a frequência dos valores da CDF das simulações apresentaram uma distribuição normal no qual a maioria

das simulações são melhores que a distribuição aleatória, mas somente uma pequena fração é semelhante ao experimental, demonstrando a necessidade de ajustes do modelo (Figura 14). Os valores do RMSE para cada simulação também apresentaram-se próximos a uma distribuição normal, com um leve deslocamento: a maioria das simulações possui um RMSE menor que o valor aleatório de 10,4 (QUAN et al., 2005). Já na relação do RMSE com a área da função de densidade acumulada, observamos que aproximadamente 78% de nossas simulações se apresentam melhores que os resultados descritos por Fange et al. (2014) e cerca de 98% são melhores em relação a distribuição constante descrita por Quan et al. (2005) (Figura 15). Apesar de nossos resultados ainda apresentarem uma diferença em comparação aos dados experimentais, nosso modelo possui uma maior precisão que os encontrados na literatura (DENNIS et al., 2009; KLUMPP; HWA, 2008; FANGE et al., 2014).

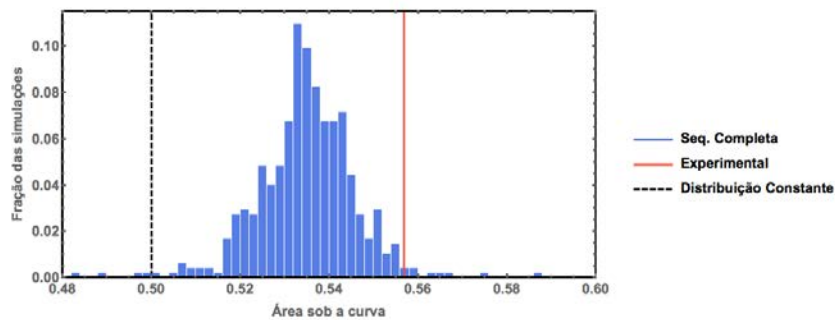


Figura 14 – Histograma da área sob a curva da CDF das simulações da fase de alongamento. Podemos observar uma pequena fração encontra-se próxima ao valor experimental, mas maior parte das simulações apresentam-se melhores que a distribuição constante (linha tracejada). Eixo das ordenadas: fração das simulações realizadas para sequência completa e eixo das abscissas: área sob a curva do CDF.

Em relação as pausas transcricionais, comparamos as simulações com a sequência completa e particionada e identificamos a presença de pausas mais intensas e outras mais moderadas em locais diferentes ao longo da sequência. A simulação da sequência completa apresentou um maior número de pausas intensas, principalmente nos intervalos entre 2000 a 3000 nt e 4500 e 5500 nt, onde se encontra o DNAr 23S. E para sequência particionada, encontramos pausas mais intensas no intervalo de 3000 a 4000 nt (Figura 16). Além disso, este comportamento de pausas mais intensas na região entre 2000 a 3000 nt no primeiro caso influencia no comportamento das últimas RNAPs presentes neste trecho, ocasionando uma diminuição da velocidade média (Figura 11b).

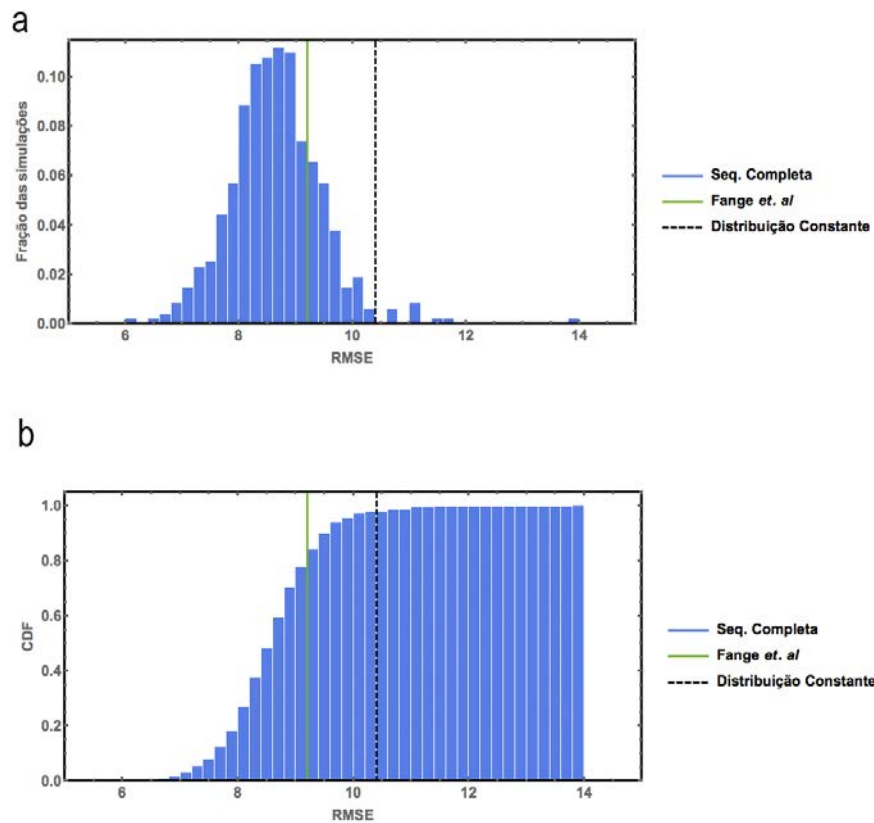


Figura 15 – Representação gráfica da frequência do RMSE (% da densidade) e da função de densidade acumulada das simulações da fase de alongamento. Em **a** podemos observar uma aproximação da distribuição normal dos dados com um leve deslocamento, na qual maioria das simulações possuem um valor para o RMSE menor que o valor aleatório de 10,4 (linha tracejada). Em **b** observamos que cerca de 78% dos resultado possuem um valor menor que o apresentado por Fange et al. (2014) e mais de 95% em relação ao aleatório (QUAN et al., 2005). A linha verde indica o valor encontrado por Fange et al. (2014) e a linha tracejada indica o valor aleatório encontrado por Quan et al. (2005).

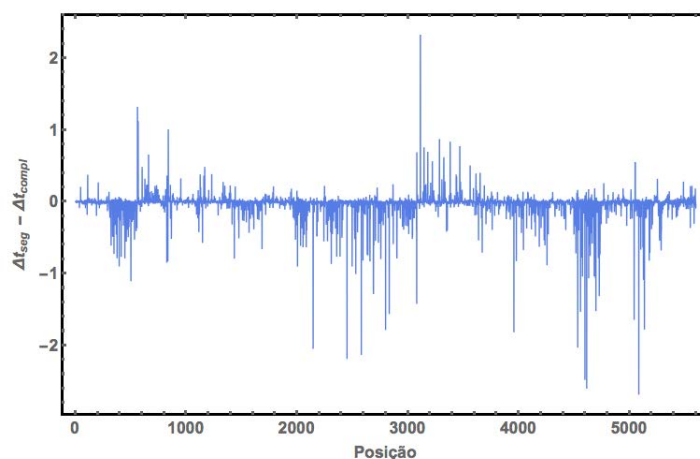


Figura 16 – Representação gráfica da diferença do tempo de pausas das simulações com a sequência completa e particionada. Parte superior do gráfico representa os sítios onde as pausas foram mais acentuadas na sequência particionada e a parte inferior onde ocorreu o oposto. Eixo das ordenadas: diferença dos tempos de pausa em segundos (s) e eixo das abscissas: posição das pausas em nucleotídeos (nt).



## 5 Conclusão e Perspectiva

### 5.1 Fase da iniciação da Transcrição

Para o processo de iniciação da transcrição, nosso modelo apresentou resultados condizentes com a literatura (60 iniciações por minuto). Além disso, podemos observar a influência de consecutivas iniciações na densidade de RNAP ao longo da transcrição, que apresentando um pequeno aumento no número de RNA polimerases, corroborando com a literatura (DENNIS et al., 2009; CONDON et al., 1993; QUAN et al., 2005; FANGE et al., 2014), entretanto não apresentando alterações significativas na velocidade da transcrição.

### 5.2 Fase do alongamento da Transcrição

Os primeiros resultados das simulações para o início do operon *rrnB* e do DNAr 5S demonstraram a influência da concentração de NTP aumentando a velocidade da transcrição, corroborando com Schneider, Gaal e Gourse (2002), entretanto não atingindo uma velocidade média de 210 nt/s para o início da sequência e mais de 400 nt/s para o DNAr 5S (DENNIS et al., 2009). Esses altos valores podem ser resultantes de fatores não considerados tais como: taxa de crescimento da bactéria, controle de estringência (deficiência de aminoácidos), fatores de ativação da região *upstream* (UP elementos) e fatores de antiterminação (*boxB*, *boxA* e *boxC*) (CONDON; SQUIRES; SQUIRES, 1995).

As simulações para o alongamento do operon *rrnB* apresentaram um comportamento similar com os dados de referência (QUAN et al., 2005), com uma variação da densidade provavelmente causada por sítios de pausa. Entretanto, apresentaram uma dificuldade em reproduzir as grandes flutuações, principalmente no começo da sequência e na seção 12. Nesse segmento, o principal ponto crítico para as simulações é prever e incluir no modelo a formação de *hairpin* pelo RNA nascente e a influencia que ele exerce sob a RNAP e conseqüentemente na transcrição. Em relação as pausas transcricionais, observamos que o tamanho da sequência simulada interfere nos sítios de pausas justamente devido a influência da interação entre as RNAPs: quanto maior a sequência, maior o número de RNAPs e conseqüentemente maior número de interações. Considerando que as pausas mais intensas possuem importante papel na regulação da expressão gênica, esse fenomeno não afetando somente a velocidade, mas também possibilita a interação de outras moléculas como fatores de transcrição e moléculas inibidoras, aumentando ou diminuindo o nível de expressão. Portanto, nosso modelo apresenta os melhores resultados existentes até agora, prevendo a essência da RNAP na transcrição, sem considerar fatores de ativação, inibidores e formação de estruturas secundárias pelo RNA nascente. Além

disso, com a nova métrica descrita nesse trabalho, relacionando o RMSE com a área da função de densidade acumulada, podemos corroborar valores teóricos com um pequeno número de experimentos, visto que ela é mais robusta e depende menos das flutuações estatísticas que ocorrem experimentalmente.

Para trabalhos futuros, deverão ser realizados estudos experimentais, principalmente *in vitro*, já que o modelo CAL é baseado em dados *in vitro*. Isso resultaria em uma maior correlação entre os resultados teóricos e experimentais na transcrição do operon *rrnB*. Um outro fator que pode ser incluído ao modelo é a influência das estruturas secundárias no comportamento da transcrição.

# Referências

- ARTSIMOVITCH, I.; LANDICK, R. The transcriptional regulator rfah stimulates rna chain synthesis after recruitment to elongation complexes by the exposed nontemplate dna strand. **Cell**, v. 109, n. 2, p. 193–203, 2002. Citado na página 13.
- BAI, L.; FULBRIGHT, R. M.; WANG, M. D. Mechanochemical kinetics of transcription elongation. **Physical Review Letters**, v. 98, n. 6, 2007. Citado 2 vezes nas páginas 27 e 28.
- BAI, L.; SHUNDROVSKY, A.; WANG, M. D. Sequence-dependent kinetic model for transcription elongation by rna polymerase. **Journal of Molecular Biology**, v. 344, n. 2, p. 335–349, 2004. Citado na página 27.
- BAIROCH, A. The enzyme database in 2000. **Nucleic Acids Research**, v. 28, n. 1, p. 304–305, 2000. Citado na página 24.
- BALLEZA, E. et al. Regulation by transcription factors in bacteria: beyond description. **Fems Microbiology Reviews**, 33, n. 1, p. 133–151, JAN 2009. Citado na página 22.
- BROWNING, D.; BUSBY, S. The regulation of bacterial transcription initiation. **Nature Reviews Microbiology**, 2, n. 1, p. 57–65, JAN 2004. Citado na página 22.
- BURMANN, B. M.; RÖSCH, P. The role of *E. coli* Nus-factors in transcription regulation and transcription:translation coupling. **Transcription**, 2, n. 3, p. 130–134, MAY-JUN 2011. Citado na página 10.
- CONDON, C. et al. Depletion of functional ribosomal RNA operons in *Escherichia coli* causes increased expression of the remaining intact copies. **The EMBO Journal**, 12, n. 11, p. 4305–4315, NOV 1993. Citado 2 vezes nas páginas 19 e 40.
- CONDON, C.; SQUIRES, C.; SQUIRES, C. Control of Ribosomal rna Transcription in *Escherichia coli*. **Microbiological Reviews**, 59, n. 4, p. 623–645, DEC 1995. Citado 4 vezes nas páginas 16, 17, 18 e 40.
- COSTA, P. R.; ACENCIO, M. L.; LEMKE, N. Cooperative RNA Polymerase Molecules Behavior on a Stochastic Sequence-Dependent Model for Transcription Elongation. **Plos One**, 8, n. 2, FEB 21 2013. Citado 3 vezes nas páginas 20, 22 e 27.
- CRAMER, P. Multisubunit RNA polymerases. **Current Opinion In Structural Biology**, 12, n. 1, p. 89–97, FEB 2002. Citado na página 9.
- CRICK, F. Central Dogma of Molecular biology. **Nature**, v. 227, n. 5258, p. 561–563, 1970. Citado na página 9.
- DARZACQ, X. et al. *In vivo* dynamics of RNA polymerase II transcription. **Nature Structural and Molecular Biology**, v. 14, n. 9, p. 796–806, 2007. Citado na página 13.

- DENNIS, P. P.; EHRENBERG, M.; BREMER, H. Control of rRNA Synthesis in *Escherichia coli*: a Systems Biology Approach. **Microbiology AND Molecular Biology Reviews**, 68, n. 4, p. 639–668, DEC 2004. Citado 2 vezes nas páginas 28 e 32.
- DENNIS, P. P. et al. Varying Rate of RNA Chain Elongation during *rrn* Transcription in *Escherichia coli*. **Journal OF Bacteriology**, 191, n. 11, p. 3740–3746, JUN 2009. Citado 7 vezes nas páginas 13, 17, 19, 30, 36, 38 e 40.
- EHRENBERG, M.; DENNIS, P. P.; BREMER, H. Maximum *rrn* promoter activity in *Escherichia coli* at saturating concentrations of free RNA polymerase. **Biochimie**, 92, n. 1, p. 12–20, JAN 2010. Citado na página 17.
- EICKBUSH, T. H.; EICKBUSH, D. G. Finely orchestrated movements: Evolution of the ribosomal RNA genes. **Genetics**, 175, n. 2, p. 477–485, FEB 2007. Citado na página 15.
- EPSHTEIN, V.; NUDLER, E. Cooperation between rna polymerase molecules in transcription elongation. **Science**, v. 300, n. 5620, p. 801–805, 2003. Citado na página 14.
- EPSHTEIN, V. et al. Transcription through the roadblocks: the role of rna polymerase cooperation. **EMBO Journal**, v. 22, n. 18, p. 4719–4727, 2003. Citado na página 14.
- FANGE, D. et al. Thermodynamic Modeling of Variations in the Rate of RNA Chain Elongation of *Escherichia coli* *rrn* Operons. **Biophysical Journal**, 106, n. 1, p. 55–64, JAN 2014. Citado 8 vezes nas páginas 20, 30, 31, 36, 37, 38, 39 e 40.
- FEMINO, A. M. et al. Visualization of single molecules of mRNA *in situ*. **Methods in Enzymology**, v. 361, p. 245–304, 2003. Citado na página 13.
- FOE, V. E. Modulation of ribosomal RNA synthesis in *Oncopeltus fasciatus*: An electron microscopic study of the relationship between changes in chromatin structure and transcriptional activity. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 42, p. 723–740, 1977. Citado na página 14.
- GIBSON, M. A.; BRUCK, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. **Journal Physical Chemistry**, v. 104, n. 9, p. 1876–1889, 2000. Citado na página 23.
- GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. **Journal Physical Chemistry**, v. 81, n. 25, p. 2340–2361, 1977. Citado 2 vezes nas páginas 19 e 29.
- GREIVE, S. J.; HIPPEL, P. H. von. Thinking quantitatively about transcriptional regulation. **Nature Reviews Molecular Cell Biology**, v. 6, n. 3, p. 221–232, 2005. Citado na página 14.
- GRIFFITHS, A. J. F. et al. **Introdução á Genética**. 8<sup>a</sup>. ed. São Paulo: Guanabara Koogan, 2006. Citado 2 vezes nas páginas 11 e 12.
- GRIGOROVA, I. L. et al. Insights into transcriptional regulation and sigma competition from an equilibrium model of rna polymerase binding to dna. **Proceedings of the National Academy of Sciences of the United States of America**, v. 130, n. 14, p. 5332–5337, 2006. Citado na página 14.

- GUAJARDO, R.; SOUSA, R. A model for the mechanism of polymerase translocation. **Journal of Molecular Biology**, v. 265, n. 1, p. 8–19, 1997. Citado na página 27.
- HAMMING, J. et al. Electron microscopic analysis of transcription of a ribosomal rna operon of *E. coli*. **Nucleic Acids Research**, v. 9, n. 6, p. 1339–1350, 1981. Citado na página 14.
- HERBERT, K. M.; GREENLEAF, W. J.; BLOCK, S. M. Single-molecule studies of RNA polymerase: Motoring along. **Annual Review of Biochemistry**, 77, p. 149–176, 2008. Citado 2 vezes nas páginas 14 e 18.
- HERBERT, K. M. et al. Sequence-resolved detecton of pausing by single rna polymerase molecules. **Cell**, v. 125, n. 6, p. 1083–1094, 2006. Citado na página 13.
- ISHIHAMA, A. Function modulation of *Escherechia coli* RNA polymerase. **Annual Review Microbiology**, 54, n. , p. 499–518, OCTUBER 2000. Citado na página 10.
- KAPANIDIS, A. N. et al. Initial transcription by rna polymerase proceeds through a dna-scrunching mechanism. **Science**, v. 314, n. 5802, p. 1144–1147, 2006. Citado 2 vezes nas páginas 11 e 12.
- KLUMPP, S. Pausing and backtracking in transcription under dense traffic conditions. **Journal of Statistical Physics**, v. 142, n. 6, p. 1252–1267, 2011. Citado na página 15.
- KLUMPP, S.; HWA, T. Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. **Proceedings of the National Academy of Sciences of the United States of America**, 105, n. 47, p. 18159–18164, NOV 25 2008. Citado 4 vezes nas páginas 16, 18, 22 e 38.
- LENINGER, A. **Princípios de Bioquímica**. 4ª. São Paulo: Artmed, 2006. Citado na página 26.
- LONG, E.; DAWID, I. Repeated genes in Eukaryotes. **Annual Review OF Biochemistry**, 49, p. 727–764, 1980. Citado na página 15.
- MARTIN, F. H.; TINOCO, I. J. Dna-rna hybrid duplexes containing oligo(da:ru) sequences are exceptionally unstable and may facilitate termination of transcription. **Nucleic Acids Research**, v. 6, n. 10, p. 2295–2299, 1980. Citado na página 14.
- MCCLURE, W. R.; CECH, C.; JOHNSTON, D. E. A steady state assay for the rna polymerase initiation reaction. **The Journal of Biological Chemistry**, v. 253, n. 24, p. 8941–8948, 1978. Citado na página 11.
- MENTEN, L.; MICHAELIS, M. I. Die kinetik der invertinwirkung. **Biochem. Z**, v. 49, p. 333–369, 1913. Citado na página 25.
- MURAKAMI, K.; DARST, S. Bacterial RNA polymerases: the whoho story. **Current Opinion in Structural Biology**, 13, n. 1, p. 31–39, FEB 2003. Citado na página 10.
- NEUMAN, K. C. et al. Ubiquitous transcriptional pausing is independent of rna polymerase backtracking. **Cell**, v. 115, n. 11, p. 437–447, 2003. Citado na página 13.

- NOMURA, M. Engineering of bacterial ribosomes: Replacement of all seven *Escherichia coli* rRNA operons by a single plasmid-encoded operon. **Proceedings of the National Academy of Sciences**, v. 96, n. 5, p. 1820–1822, 1999. Disponível em: <<http://www.pnas.org/content/96/5/1820.short>>. Citado na página 17.
- NUDLER, E.; GOTTESMAN, M. Transcription termination and anti-termination in *E. coli*. **Genes to Cells**, 7, n. 8, p. 755–768, AUG 2002. Citado 2 vezes nas páginas 14 e 15.
- O'BRIEN, T.; LIS, J. T. Rapid changes in drosophila transcription after an instantaneous heat shock. **Molecular and Cellular Biology**, v. 314, n. 6, p. 3456–3463, 1993. Citado na página 13.
- PAUL, B. et al. rRNA transcription in *Escherichia coli*. **Annual Review of Genetics**, 38, p. 749–770, 2004. Citado na página 16.
- QUAN, S. et al. Transcriptional polarity in rRNA operons of *Escherichia coli* nusA and nusB mutant strains. **Journal of bacteriology**, v. 187, n. 5, p. 1632–1638, 2005. Citado 8 vezes nas páginas 19, 20, 30, 36, 37, 38, 39 e 40.
- RADZICKA, A.; WOLFENDEN, R. A proficient enzyme. **Science**, v. 267, n. 5194, p. 90–93, 1995. Citado na página 25.
- RAJALA, T. et al. Effects of transcriptional pausing on gene expression dynamics. **PLoS Computational Biology**, v. 6, n. 3, 2010. Citado na página 15.
- RING, B. Z.; YARNELL, W. S. Function of *E. coli* RNA polymerase sigma factor sigma(70) in promoter-proximal pausing. **Cell**, v. 86, n. 3, p. 485–493, 1996. Citado na página 13.
- SANTALUCIA, J.; ALLAWI, H. T.; SENEVIRATNE, A. Improved nearest-neighbor parameters for predicting dna duplex stability. **Biochemistry**, v. 35, n. 11, p. 3555–3562, 1996. Citado na página 27.
- SCHLIWA, M. Molecular motors. **Nature**, v. 422, n. 4, p. 759–765, 2003. Citado na página 10.
- SCHNEIDER, D. A.; GAAL, T.; GOURSE, R. L. NTP-sensing by rRNA promoters in *Escherichia coli* is direct. **PNAS**, 99, n. 13, p. 8602–8607, JUNE 2002. Citado na página 40.
- SHAPIRO, B. E. et al. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. **Bioinformatics**, v. 19, n. 5, p. 677–678, 2003. Citado na página 29.
- SOSUNOV, V. et al. Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. **Embo Journal**, 22, n. 9, p. 2234–2244, MAY 1 2003. Citado na página 10.
- SWEETSER, D.; NONET, M.; YOUNG, R. A. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. **Proc Natl Acad Sci U S A**, v. 84, n. 5, p. 1192–1196, Mar 1987. Citado na página 9.

TENNYSON, C. N.; KLAMUT, H. J.; WORTON, R. G. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. **Nature Genetics**, v. 9, n. 2, p. 184–190, 1995. Citado na página 13.

WANG, M. D. et al. Force and Velocity Measured for Single Molecules of RNA Polymerase. **Science**, v. 282, n. 5390, p. 902–907, 1998. Disponível em: <<http://www.sciencemag.org/content/282/5390/902.abstract>>. Citado 2 vezes nas páginas 10 e 11.

WERNER, F. Structural evolution of multisubunit RNA polymerases. **Trends in Microbiology**, 16, n. 6, p. 247–250, JUN 2008. Citado na página 9.

YIN, H. et al. Transcription Against an Applied Force. **Science**, v. 270, n. 5242, p. 1653–1657, 1995. Disponível em: <<http://www.sciencemag.org/content/270/5242/1653.abstract>>. Citado 2 vezes nas páginas 10 e 11.

ZHOU, J.; RUDD, K. E. EcoGene3.0. **Nucleic Acids Research**, 41, n. , p. 613–624, 2013. Citado na página 22.