

Evolução Molecular Computacional

Caio Henrique Machado de Oliveira

**Projeto de pesquisa apresentado
para participação em Programa
De Iniciação Científica - PIC**

São Paulo

2018

Evolução Molecular Computacional

Caio Henrique Machado de Oliveira

**Projeto de pesquisa apresentado
para participação em Programa
De Iniciação Científica – PIC**

**Área de concentração:
Biotecnologia**

**Orientador: Prof. Wenderson A. S.
S.**

São Paulo

2018

RESUMO

No primórdio, a Terra era um ambiente desagradável com altas temperaturas, e este era um ambiente totalmente hostil para a reprodução de moléculas. Eram poucas as moléculas que estavam aptas a estas condições conseguindo sobreviver a este meio, crescer e aumentar o tamanho do Genoma. Desde o princípio as moléculas foram submetidas a Seleção natural, onde o ambiente as influenciaria decidindo assim quais as moléculas que prevaleceriam. Para que todo esse processo seja avaliado, as principais etapas a serem estudadas são: a aquisição de informação e o crescimento do genoma. Porém, ao realizar estas análises possuímos uma barreira, pois temos o longo período apagado em nossa história onde diversas informações não foram registradas através dos fósseis (utilizados para estudos), e com isso acabamos perdendo um enorme campo de informação. O projeto genoma foi uma explosão de informação onde foi possível realizar a análise de diversas teorias, mas elas ainda não trazem respostas concretas sobre o que aconteceu logo após o surgimento da Terra. Partindo desses princípios, através de programas elaborados em Python foi desenvolvido simulações que mimetizam a seleção natural com moléculas e utilizando como método principal ferramentas da Teoria da informação de Shannon para melhor entendimento dos cenários no qual foram analisados tanto seu crescimento genômico quanto sua estrutura e sua adaptabilidade ao meio.

Palavras-chave: Selex; Evolução molecular; Simulação; Entropia de Shannon; Seleção natural;

LISTA DE ILUSTRAÇÕES

FIGURA 1: DESENVOLVIMENTO DA ENTROPIA.....	8
FIGURA 2: FLUXOGRAMA PRINCIPAL.....	11
FIGURA 3: FLUXOGRAMA DAS SIMULAÇÕES DOS CENÁRIOS 1 E 2.	13
FIGURA 4: FLUXOGRAMA DA SIMULAÇÃO DO CENÁRIO 3.	13
FIGURA 5: CENÁRIO 1- TAXA DE MUTAÇÃO 0%.	15
FIGURA 6: CENÁRIO 1- TAXA DE MUTAÇÃO 5%.	16
FIGURA 7: CENÁRIO 1- TAXA DE MUTAÇÃO 15%.	16
FIGURA 8: CENÁRIO 1- TAXA DE MUTAÇÃO 20%.	16
FIGURA 9: CENÁRIO 1- RELAÇÃO ENTRE AFINIDADE E TAXA DE MUTAÇÃO.....	17
FIGURA 10: CENÁRIO 1- EFICIÊNCIA DE FILTRO 90%.	17
FIGURA 11: CENÁRIO 1- EFICIÊNCIA DE FILTRO 80%	18
FIGURA 12: CENÁRIO 1- EFICIÊNCIA DE FILTRO 50%.	18
FIGURA 13: CENÁRIO 1- EFICIÊNCIA DE FILTRO 0%.	18
FIGURA 14: CENÁRIO 1- RELAÇÃO ENTRE AFINIDADE E EFICIÊNCIA DE FILTRO.....	19
FIGURA 15: CENÁRIO 1- 1 FILTRO DE 20% COMPARADO A 2 FILTROS DE 10%.....	19
FIGURA 16: CENÁRIO 1- COMPARANDO MÉTODO SIMULADO COM CALCULADO.....	20
FIGURA 17: CENÁRIO 2- TAXA DE MUTAÇÃO 0%.	21
FIGURA 18: CENÁRIO 2- TAXA DE MUTAÇÃO 5%.	22
FIGURA 19: CENÁRIO 2- TAXA DE MUTAÇÃO 10%.	22
FIGURA 20: CENÁRIO 2- EFICIÊNCIA DE FILTRO 80%.	23
FIGURA 21: CENÁRIO 2- EFICIÊNCIA DE FILTRO 60%.	23
FIGURA 22: CENÁRIO 2- EFICIÊNCIA DE FILTRO 10%.	23
FIGURA 23: CENÁRIO 3- TAXA DE MUTAÇÃO 1%.	24
FIGURA 24: CENÁRIO 3- TAXA DE MUTAÇÃO 2%.	25
FIGURA 25: CENÁRIO 3- TAXA DE MUTAÇÃO 5%.	25
FIGURA 26: CENÁRIO 3- TAXA DE MUTAÇÃO 10%.	25

LISTA DE ABREVIATURAS

AMS	Amplificação, mutação e seleção
MC	Método Monte Carlo
A	Adenina
G	Guanina
U	Uracila
T	Timina
C	Citosina
Selex	Systematic evolution of ligands by exponential enrichment
PCR	Polymerase Chain Reaction
DNA	Ácido desoxirribonucléico
RNA	Ácido ribonucléico

Sumário

1. INTRODUÇÃO	7
1.1. Teoria da informação e evolução molecular	7
1.2. Selex.....	7
1.3. Entropia	8
2. JUSTIFICATIVA.....	9
3. OBJETIVO	10
3.1. Geral.....	10
3.2. Específico	10
4. METODOLOGIA	11
4.1. Protocolo	11
4.2. Código	12
5. ANÁLISE DAS SIMULAÇÕES	15
5.1. Cenário 1	15
5.1.1. Variando a taxa de mutação	15
5.1.2. Variando a eficiência do filtro.....	17
5.1.3. Presença de 2 filtros.....	19
5.1.4. Equação modelada	20
5.2. Cenário 2	21
5.2.1. Variando a taxa de mutação	21
5.2.2. Variando a eficiência do filtro.....	23
5.3. Cenário 3	24
5.3.1. Variando a taxa de mutação	24
6. REFERÊNCIAS BIBLIOGRÁFICAS	26

1. INTRODUÇÃO

Em meio a diversas hipóteses sobre o surgimento da Terra e a evolução genômica foram abertas a um longo período de discussão e debate que nos rodeia até hoje trazendo teorias de diferentes campos com intenção de esclarecer todo esse passado apagado. Uma teoria bem popular é a Panspermia cósmica [1].

Para ter uma ideia sobre a evolução molecular, modelos biológicos [17,18] e matemáticos [19] foram desenvolvidos. Na área de matemática, o modelo mais conhecido é o Modelo de Eigen [5,6] que descreve a evolução do fago Q β por equações diferenciais. Na área biológica suposição do “Mundo de RNA” [7] se tornando muito visado por trazer a descoberta sobre RNA catalítico [8,9]. Com esse ganho de conhecimento e popularidade foi um enorme passo entre a junção da biologia e tecnologia proporcionando o sequenciamento e o PCR (*polymerase chain reaction*) [10] com ciclos de AMS (amplificação mutação e seleção) [11], assim, podendo proporcionar diversos estudos e pesquisas na área de biotecnologia. O modelo de AMS que teve a intenção de ser mimetizado nessa pesquisa foi SELEX (*systematic evolution of ligands by exponential enrichment*) [12,13], um sistema usado no tratamento ou detecção de doenças[14,15,16].

1.1. Teoria da informação e evolução molecular

A teoria da informação [4] está tão presente no dia a dia, é a área que trabalha com símbolos, mensagens e canais, seu feito é em diversas áreas do conhecimento. Desenvolvida em 1948 por Shannon. O princípio básico da teoria da informação é estudar a transmissão de dados, onde, os símbolos podem ser considerados dados ou mensagens, o canal seria o meio ou fio onde seria feito o envio e recepção da mensagem. Como esse canal não é perfeito, pode haver interferências que atrapalhem no envio desses símbolos fazendo com que alguns deles sem percam no caminho. Quando aplicamos esse mesmo princípio na Biologia, as mensagens passam a serem os nucleotídeos ou aminoácidos, o canal seria a replicação pois é o momento onde acontece a transcrição e os erros mimetizam o processo de mutação.

1.2. Selex

Se baseando no estudo da evolução do RNA catalítico *in vitro*, Joyce desenvolveu um protocolo que trabalhavam com ciclos de amplificação, mutação e seleção [11]. No ano seguinte, Ellington e Szostak [12] e Tuerk e Gold [13] desenvolveram o SELEX (*systematic evolution of ligands by exponential enrichment*) [12,13]. Um sistema onde os alvos buscam moléculas aptas ao seu interesse fazendo esse processo por diversos ciclos, se baseando em etapas de amplificação, mutação e seleção.

1.3.Entropia

Para a análise da grande massa de moléculas foi necessária uma ferramenta desenvolvida por Shannon em 1948 [3] chamada entropia de Shannon [4], com ela foi possível medir o grau de desordem entre cada índice das moléculas. A entropia de Shannon é dada pela seguinte expressão:

$$H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

$p(x_i)$ = probabilidade de ocorrer x_i em X.

Para calcular a entropia nas moléculas, foi feito por índice de base assim cada coluna tinha o seu valor de entropia calculado e logos após tirando a média dessas entropias, segue abaixo uma figura sobre o desenvolvimento da entropia:

Molécula 1	= "	A	C	T	T	G	A	C	"
Molécula 2	= "	G	T	A	C	A	T	C	"
Molécula 3	= "	T	G	T	A	A	C	T	"
Molécula 4	= "	A	A	C	G	T	C	C	"
Molécula 5	= "	C	A	G	A	T	C	G	"
Entropia por índice	=	H1	H2	H3	H4	H5	H6	H7	

Figura 1: Desenvolvimento da entropia.

2. JUSTIFICATIVA

O projeto genoma foi um dos grandes avanços na biologia com intenção de mapear o corpo humano com intenção de explorar o código genético de um organismo e entender esse sistema complexo, iniciado por volta do ano 2000, foi produzido uma enorme quantidade de informação, onde se pode avaliar as diferentes teorias. No entanto, essas informações apenas não trazem respostas para as perguntas sobre o que aconteceu com a vida na terra logo após o seu surgimento. Sabe-se, por meio de registros fósseis, que os seres existentes logo após o surgimento da vida eram muito simples, mas que em 300 milhões de anos já havia seres muito mais complexos[1,2]. Há um "elo perdido" entre esses registros, onde muitos cenários possíveis podem ter acontecido e onde simulações computacionais do SELEX[3] e a teoria de informação[4,5] podem lançar uma luz.

3. OBJETIVO

3.1. Geral

O principal objetivo foi dar continuação ao projeto de pesquisa do Professor orientador Wenderson Alexandre. As etapas de simulação *in silico* envolvem ciclos onde são trabalhados com o protocolo de AMS (amplificação, mutação e seleção) de moléculas que consiste em analisar o processo de seleção natural e deixar o sistema de modo mais real possível.

3.2.Específico

O objetivo da pesquisa é analisar cenários de amplificação, mutação e seleção (protocolo AMS) de moléculas por meio de simulações computacionais (*in silico*) com o intuito de explorar diversas variáveis como: crescimento genômico da molécula, adaptação ao ambiente, sistemas de quebra de molécula, sistema de junção de moléculas, estrutura com base tanto nas moléculas da sua geração quanto nas moléculas da geração anterior e por fim analisar seu efeito natural de seleção com base nas próximas gerações de moléculas, a intenção é não fugir muito da realidade, mas também explorar o limite da simulação. Para a análise e avaliação das moléculas, foram usados métodos que ajudavam na avaliação sem ter que analisar cada uma delas a olho nu, dentre eles, os principais foram a entropia de Shannon que teria por finalidade medir o grau de desordem das moléculas e a distância de Hamming e Levenshtein que servia para dizer a quantidade de bases que eram diferentes ao pegar índice por índice da molécula e sua replicação, o motivo desse método ser inserido na simulação é que cada replicação tinha sua porcentagem de erro na replicação.

4. METODOLOGIA

4.1. Protocolo

O protocolo utilizado foi baseado na metodologia de Amplificação-Mutação-Seleção (AMS) onde os processos de amplificação e mutação eram representados pelo PCR e a seleção era representada pelo próprio SELEX que buscava sequências de afinidade. Os códigos foram desenvolvidos em Python por questão de facilidade em tratamento de “strings”, pois eram assim que as moléculas eram representadas, outra questão era a facilidade em desenvolver gráficos para o fácil entendimento e comportamento das moléculas.

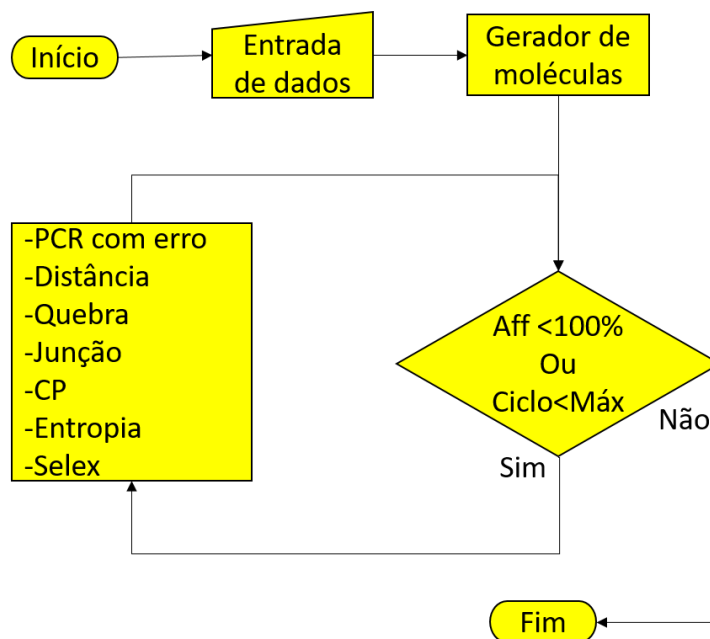


Figura 2: Fluxograma principal.

As primeiras etapas do ciclo eram amplificação e mutação, etapas essas que eram feitas em somente um passo. O PCR com erro (PCR error prone) era processo que realizava as replicações, colocando assim, uma molécula sucessora que nessa pesquisa vai ser chamada de molécula-filha. Essa molécula-filha pode conter pequenos erros no seu código que podem simplesmente favorecer ou prejudicar ela na seleção natural.

A etapa que diz respeito sobre a distância foi utilizada para análise de cada molécula, pois com ela era possível saber a quantidade de bases que mudaram após a sua

replicação, o método usado para o tamanho constante foi a distância de Hamming e para as moléculas com tamanhos variados a distância de Levenshtein.

Em dois dos três cenários foram implementados métodos para a variação de tamanho, métodos esses que podiam tanto aumentar uma molécula quanto diminuir a mesma, trazendo assim, um modelo mais real e mais complicado pois ajustar esses dois processos emanaram muitos esforços e tentativas com intenção de equilibrar este processo.

O método de Monte Carlo chamado de Constant Population (CP) foi um algoritmo muito importante para a análise das moléculas pois sem ele as replicações iriam tomar uma proporção enorme em questão a quantidade de moléculas, com ele foi possível colocar um limite máximo na quantidade de moléculas.

A função da entropia foi medir o grau de desordem das moléculas no ciclo em bits e ver o comportamento do resultado da entropia no gráfico, sendo ele, 2 que seria o maior grau de aleatoriedade por existirem somente 4 bases e 0 caso todas moléculas fossem exatamente iguais que é um fato impossível de acontecer.

A terceira etapa do ciclo foi a seleção que mimetizava o próprio SELEX. Foi escolhido uma sequência de bases que caracterizava a afinidade e se a molécula era “apta” ou não a seleção, na pesquisa essa sequência de bases são chamadas de sequencias afins, conservadas, alvos ou filtro. A intenção das sequências é mimetizar um vírus ou proteína e ver o quão forte ela é na seleção natural.

4.2. Código

O fluxo do programa desenvolvido para os cenários 1, 2 e 3 estão representados nas figuras abaixo:

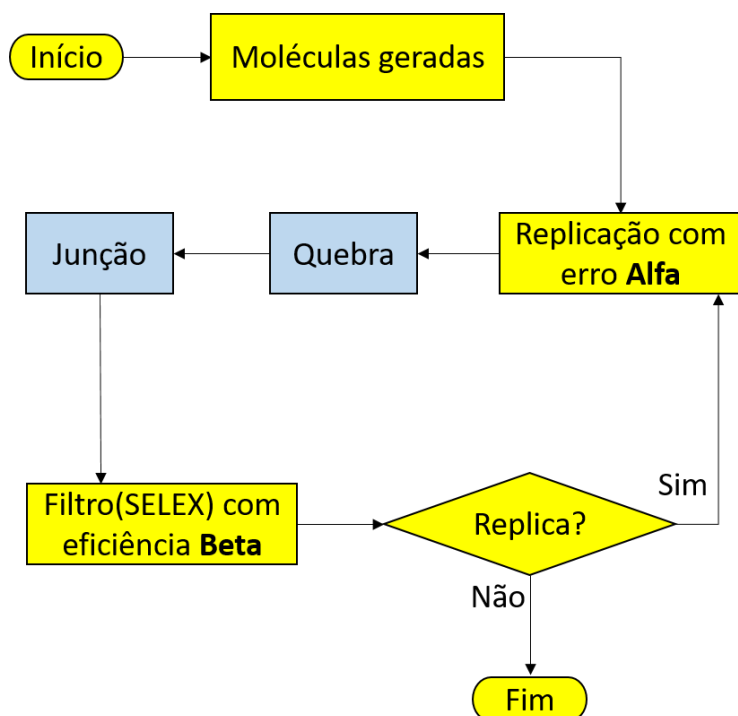


Figura 3: Fluxograma das simulações dos cenários 1 e 2.

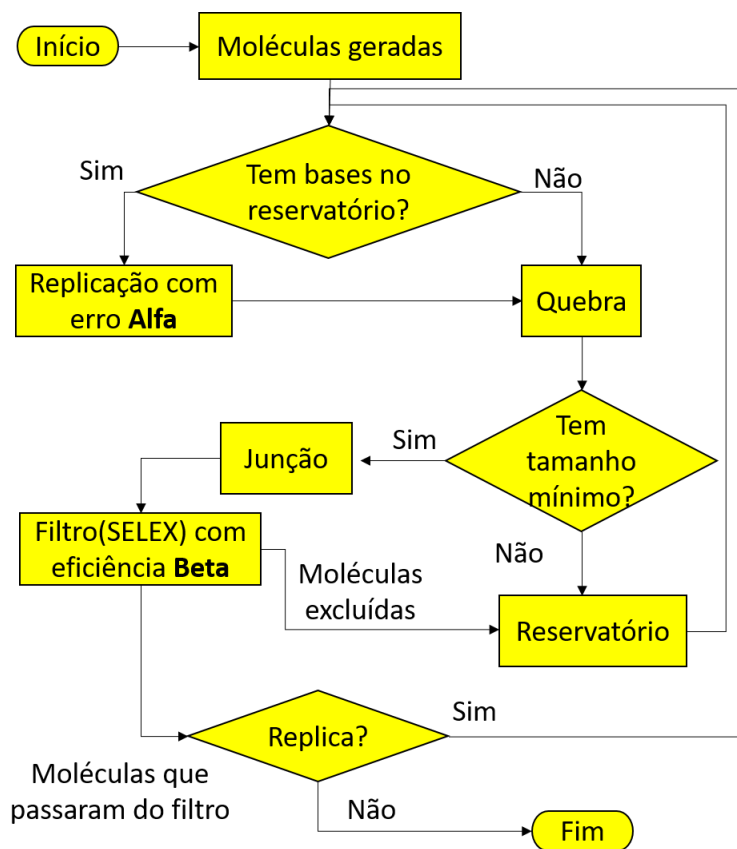


Figura 4: Fluxograma da simulação do cenário 3.

O código foi desenvolvido em Python e contém 5 classes. Sendo elas: **Classe_RNA.py**, **Classe_Filtro.py**, **Classe_Equação.py**, **Classe_Entropia.py** e **Classe_Distancia.py**. Todas as 5 classes são chamadas em um código principal que chama a função a cada ciclo e após o término da simulação é gerado quatro gráficos, sendo eles: Afinidade x Ciclo, Entropia x Ciclo, Quantidade de moléculas x Ciclo, Tamanho médio x Ciclo. O código principal também é capaz de gerar o desvio padrão, a variância, tamanho da menor e maior molécula e por fim mostrar os pontos de máximo e mínimo da afinidade.

A classe RNA faz todo o processo de *Amplificação e Mutação*. É responsável pela maior parte do processo tendo funções que fazem as seguintes operações: Gera a quantidade de moléculas desejadas com o tamanho desejado para a primeira geração de moléculas entrar no ciclo; faz a replicação com erro **Alfa**, sendo esse erro aplicado base a base em cada molécula; limita a quantidade de moléculas com o método de Monte Carlo chamado Constant Population (CP) que exclui moléculas aleatoriamente, assim, não influenciando muito no resultado e mantendo a quantidade de moléculas em uma quantidade desejável, o método foi que adaptado para deletar aleatoriamente metade das moléculas afins e metade das moléculas não-afins reduzindo o erro a zero. Como essa classe fala de tudo que diz respeito a estrutura da molécula nada mais justo do que colocar as funções de quebra e junção nessa classe. O processo de quebra e junção só foram implementados nos cenários 1 e 2, no caso da junção, foi desenvolvida escolhendo uma sequência de 5 bases que caso a molécula tivesse no seu *códon* teria uma porcentagem de chance de junção com uma molécula afim, no cenário 2 essa porcentagem foi de 100%, já no cenário 3 foi de 10%. No processo de quebra foi escolhido uma sequência de 3 bases que se caso uma molécula tivesse no seu *códon* essa molécula ela se quebrava logo após, a chance de quebra foram as mesmas, 100% no cenário 2 e 10% no cenário 3.

A classe Distância é responsável por parte da análise das moléculas, é com ela que analisamos uma molécula e a sua duplicata ou molécula-filha sendo colocada uma embaixo da outra e olhando base por base, caso tenha diferença é incrementado 1 na contagem da distância dessa molécula-filha. Esse processo é feito a cada alteração na molécula sendo ela replicação, quebra, junção ou filtragem para que se tenha total controle da molécula. Vale lembrar que a distância usada no primeiro cenário é a distância de Hamming pois as moléculas têm o mesmo tamanho e no segundo e terceiro cenário é usado a distância de Levenshtein pois tem tamanhos variados ao decorrer dos ciclos.

A classe Entropia foi responsável pelo cálculo da entropia de Shannon no primeiro cenário como teve tamanho médio das moléculas sempre constante, foi exato o valor da medida. O cenário 2 e 3 teve um grau de entropia aproximado por conta dos tamanhos variados, por conta disso foi escolhido pelo menor tamanho que uma molécula poderia ter, que no caso desses cenários o tamanho mínimo era 10, foi testado também pegando como referência a menor molécula do ciclo, mas graficamente os valores se mostravam com enormes saltos.

A classe Filtro seria o próprio *SELEX*, onde moléculas afins tinham vantagens sobre as não-afins, no cenário 1 e 2 a probabilidade de as moléculas afins saírem é de 0% já no cenário 3 é de 5%. Com exceção do cenário 1, os outros cenários contêm filtragem por distância da molécula replicada com a original, ou seja, caso uma molécula-filha tenha

10% de diferença da sua descendente, ela tem 10% de chance morrer. A eficiência **Beta** se refere a probabilidade que uma molécula não-afim sobreviver ao filtro. Lembrando que para saber o valor da afinidade foi executada a equação:

$$(Aff = \frac{Mol. \text{ afins}}{Total \text{ de moléculas}}) \quad (2)$$

A classe Equação tem o intuito de trazer uma equação modelada para o cenário 1 com intenção de mostrar os resultados do método calculado e simulado para uma comparação entre quantidade de moléculas e eficiência de filtro trazendo assim uma segurança maior quanto aos resultados exibidos.

5. ANÁLISE DAS SIMULAÇÕES

5.1. Cenário 1

O cenário 1 foi o primeiro a ser desenvolvido e o mais básico, a intenção foi incrementar outras variáveis com o desenvolvimento de outros cenários e como o cenário 1 foi o menos complexo pode ser feita diversas análises tendo como a principal, uma modelagem matemática do cenário. Para termos noção das simulações e um bom equilíbrio dos resultados, foi considerado por padrão de valores fixos do cenário os seguintes: tamanho fixo de 50 bases, 500 moléculas iniciais, 20% eficiência de filtro (chance de uma molécula não-afim morrer), 0% de chance de uma molécula afim morrer, 5% de taxa de mutação, 500 moléculas de totais no *round* e o tamanho do filtro era de 5 bases onde a sequência era escolhida aleatoriamente no início.

5.1.1. Variando a taxa de mutação

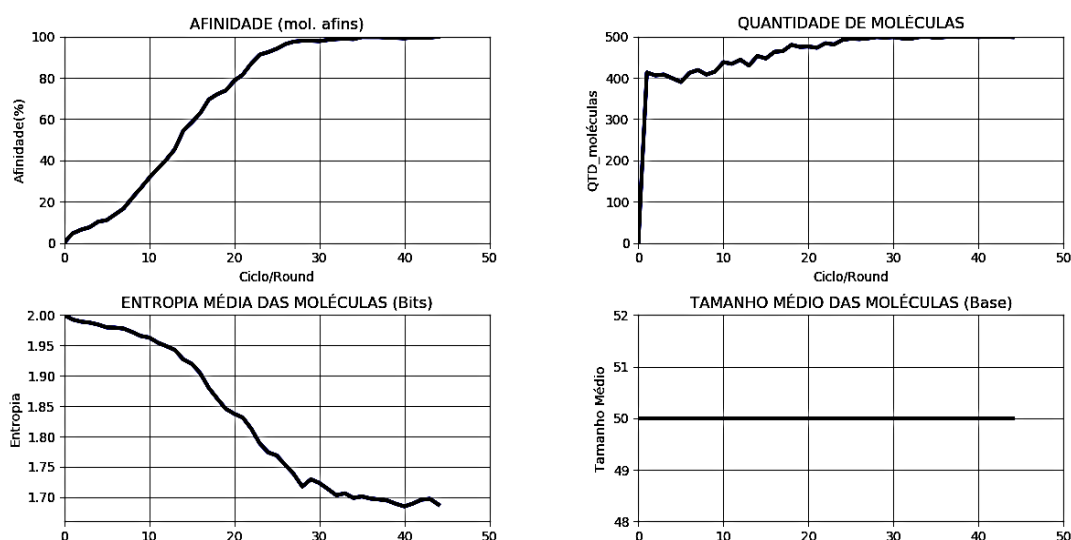


Figura 5: Cenário 1- Taxa de mutação 0%.

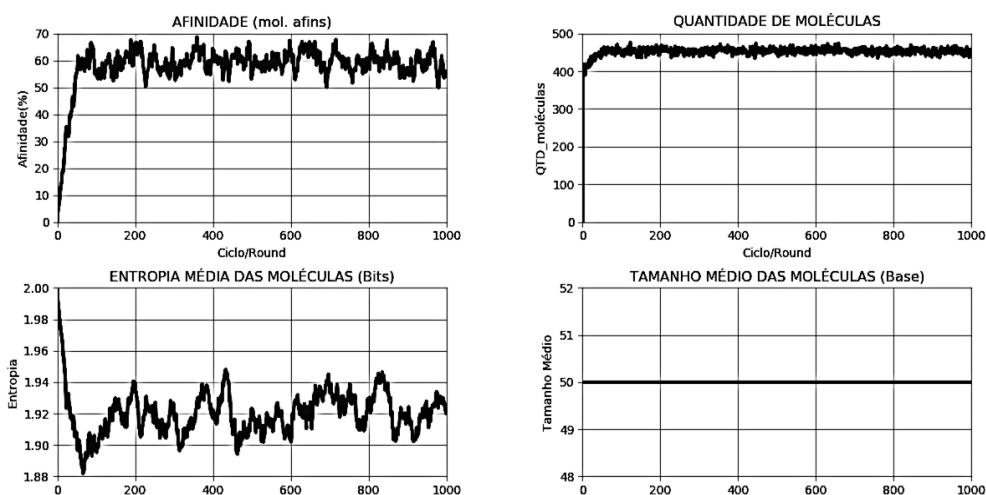


Figura 6: Cenário 1- Taxa de mutação 5%.

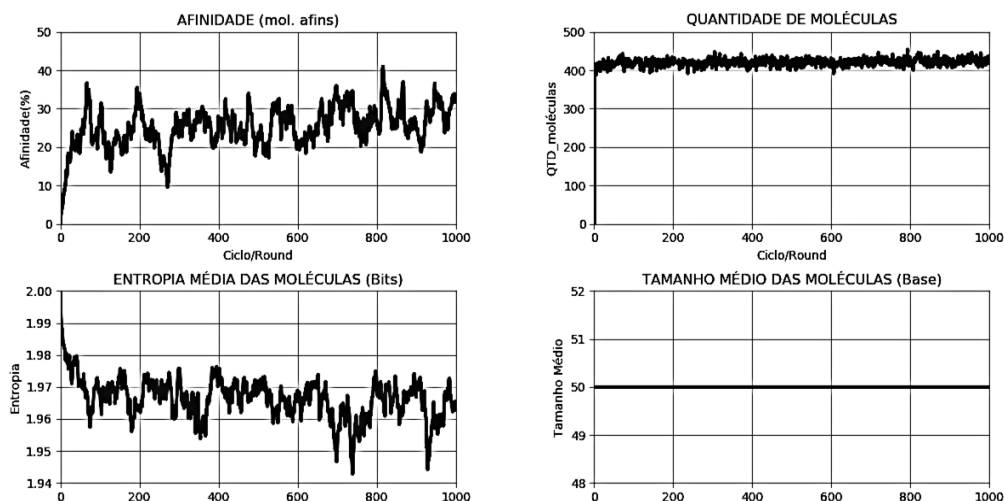


Figura 7: Cenário 1- Taxa de mutação 15%.

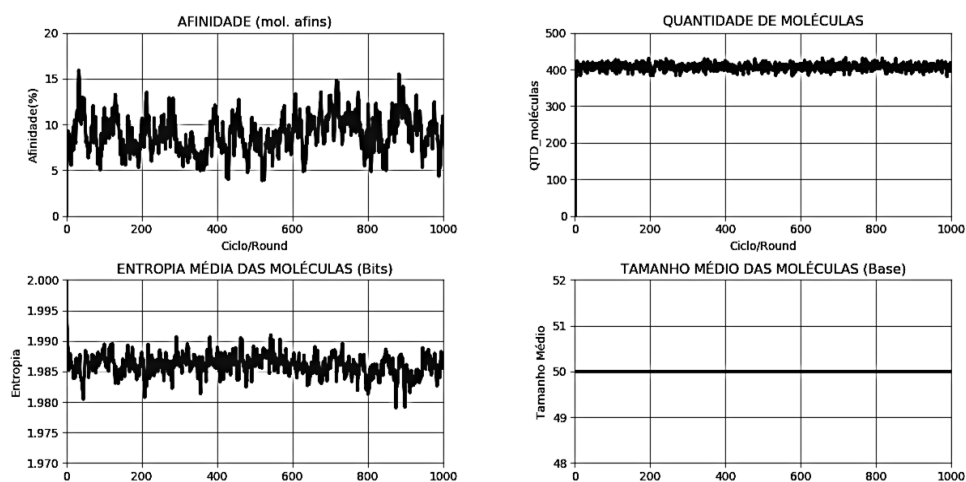


Figura 8: Cenário 1- Taxa de mutação 20%.

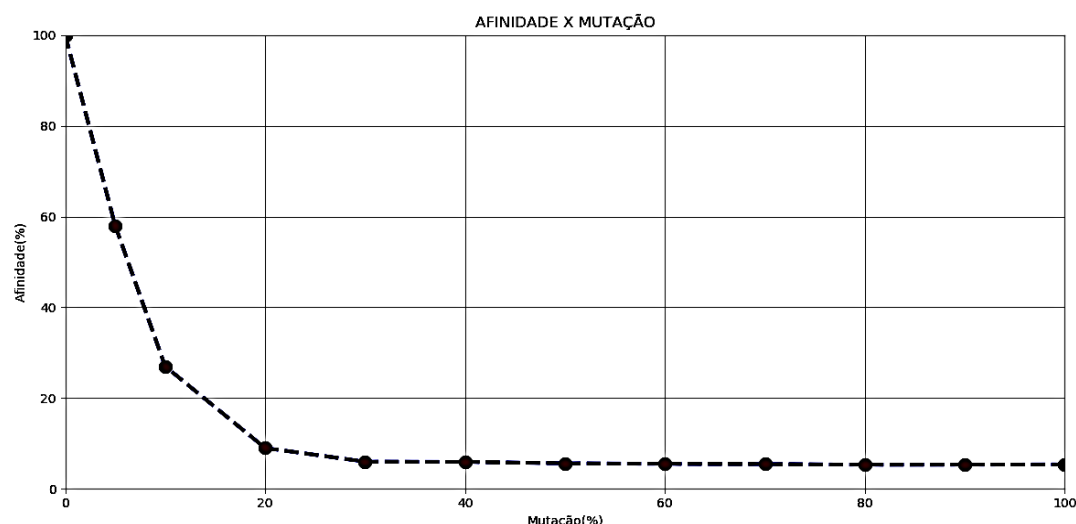


Figura 9: Cenário 1- Relação entre afinidade e taxa de mutação.

Como pode-se reparar, ao observar as figuras 4-7 podemos perceber que o único momento em que a afinidade bate 100% a simulação tem 0% de mutação e que com somente 5% de mutação, a afinidade fica em média de 60% mostrando que a porcentagem de afinidade é muito sensível a mutação. Na figura 8 temos uma referência que a partir de 20% de mutação a afinidade fica menor que 10% se tornando mínima a quantidade de moléculas aptas ao filtro.

5.1.2. Variando a eficiência do filtro

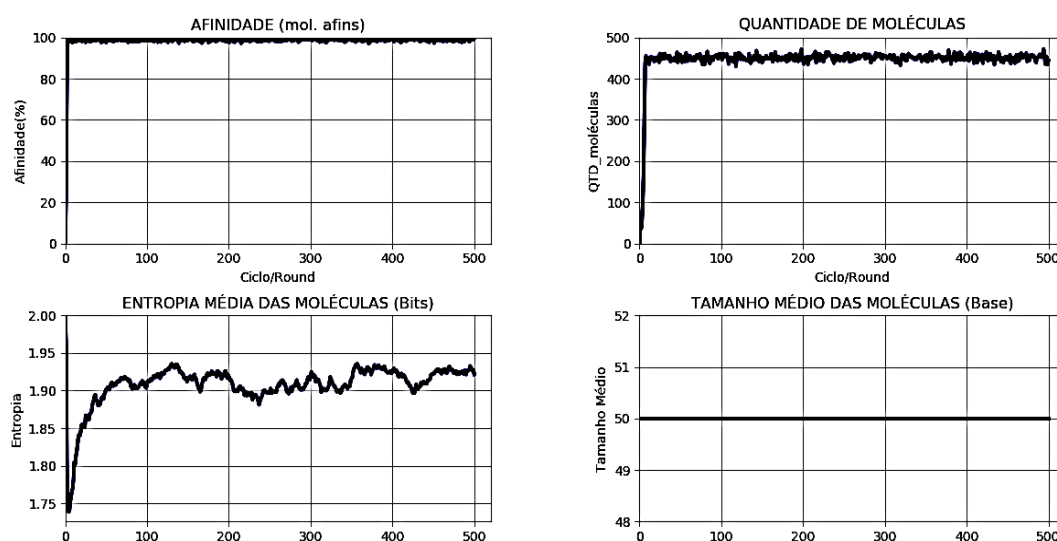


Figura 10: Cenário 1- Eficiência de filtro 90%.

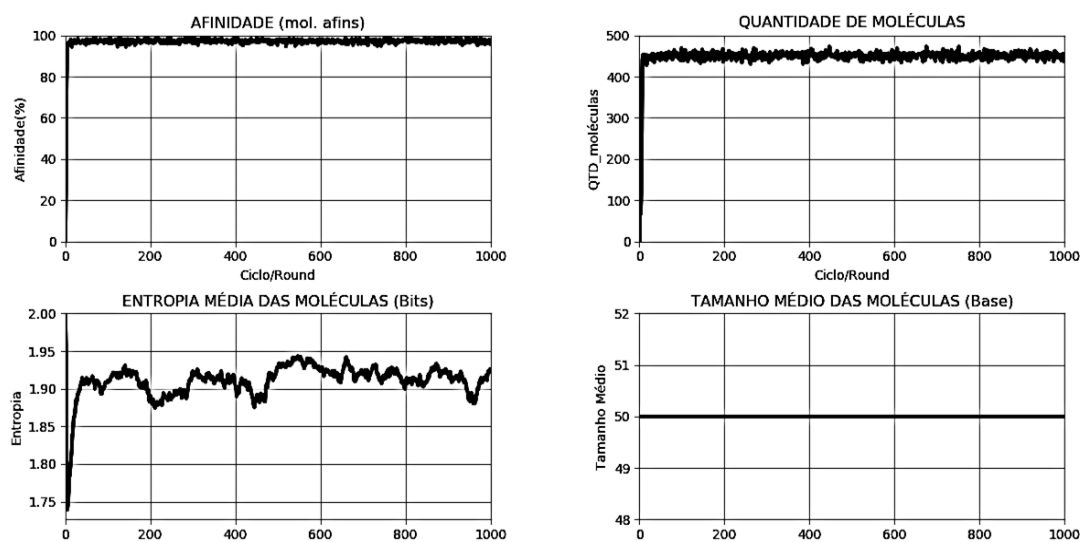


Figura 11: Cenário 1- Eficiência de filtro 80%

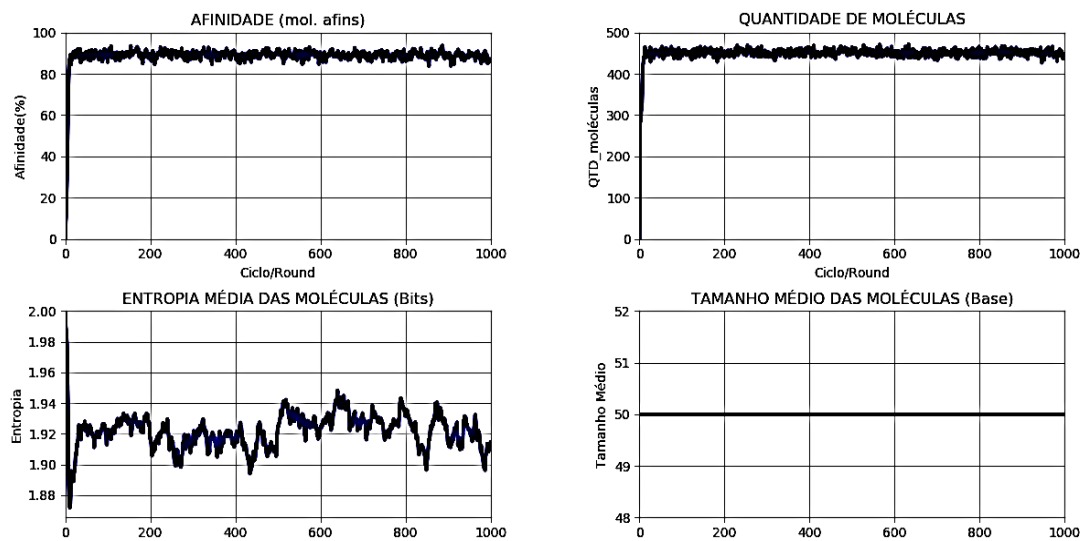


Figura 12: Cenário 1- Eficiência de filtro 50%.

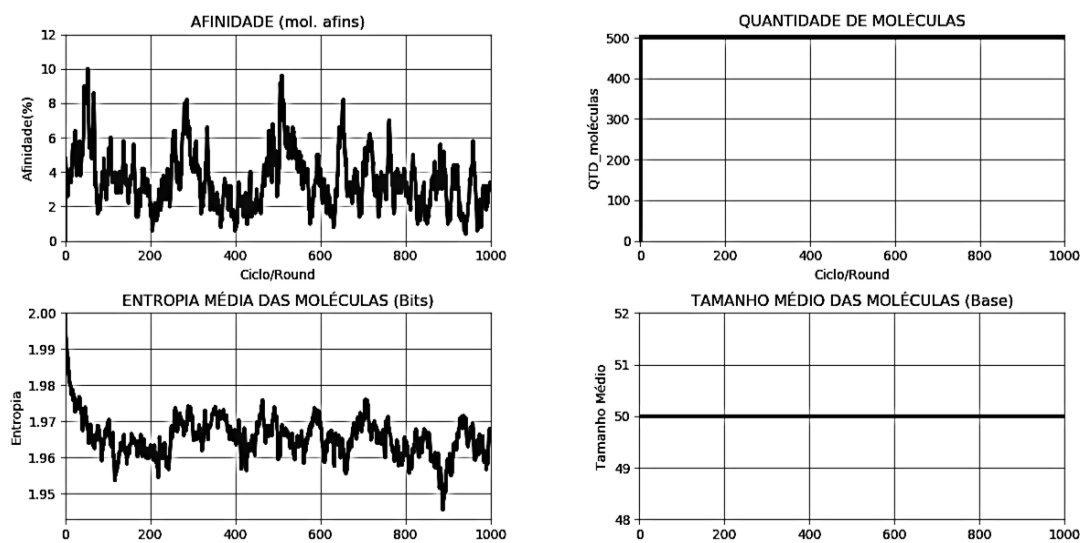


Figura 13: Cenário 1- Eficiência de filtro 0%.

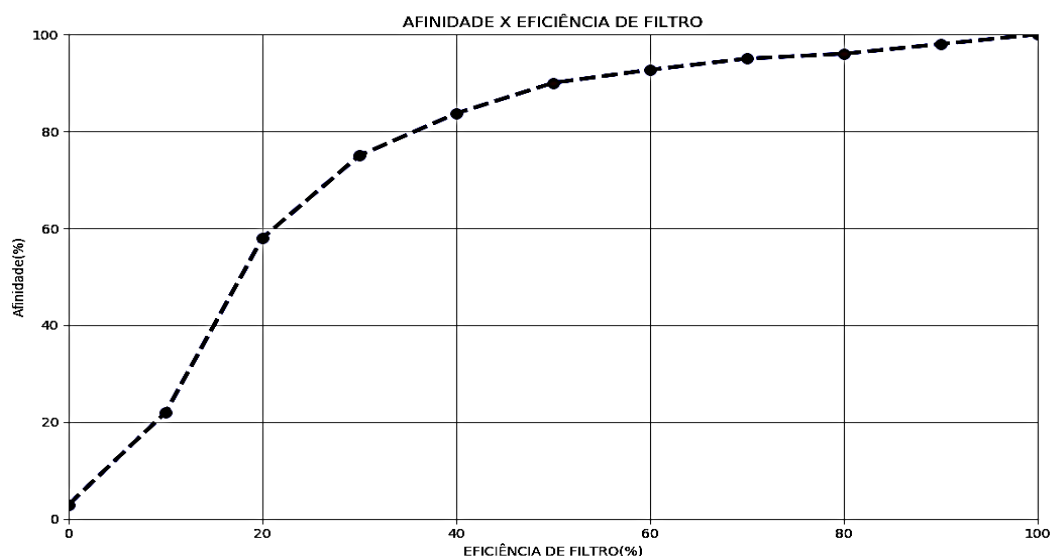


Figura 14: Cenário 1- Relação entre afinidade e eficiência de filtro.

Com bases nas figuras 9-13, pode-se perceber que a afinidade se decai muito quando é menor que 20% de eficiência de filtro, acima desse valor tanto a afinidade quanto a quantidade de moléculas não são tão sensíveis caso tenha uma mudança mínima de eficiência. Qualquer valor acima de 20% seria um bom valor para a simulação.

5.1.3. Presença de 2 filtros

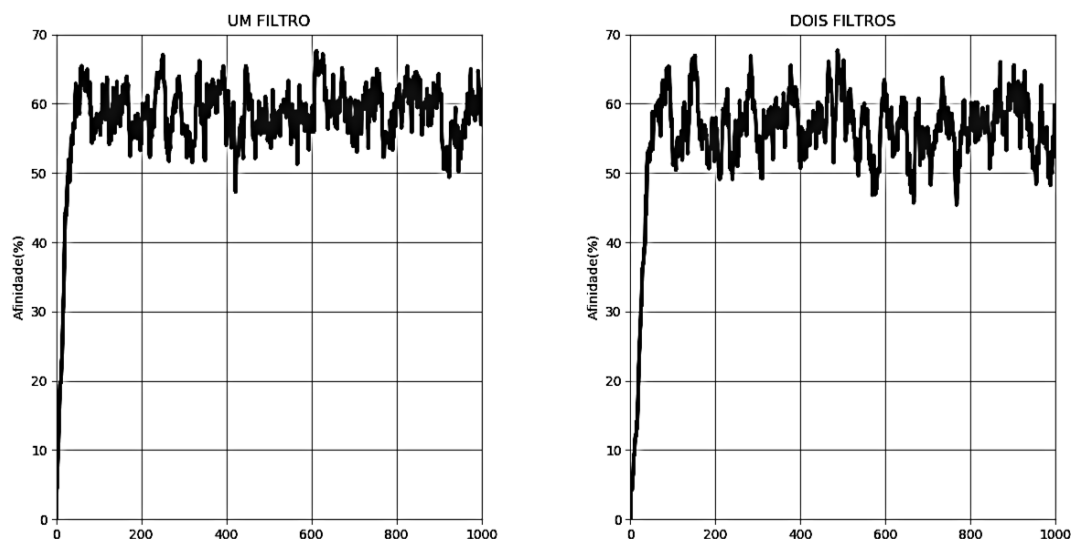


Figura 15: Cenário 1- 1 filtro de 20% comparado a 2 filtros de 10%

A intenção dessa simulação foi mostrar que não tem tanta diferença em acrescentar mais filtros se eles forem iguais (buscarem a mesma sequência) pois basta aumentar a força do filtro que você tem em mãos, na figura 14 pode-se perceber que os valores são idênticos.

5.1.4. Equação modelada

$$Q(t) = (1 - diff_i)[N_A + (1 - \alpha)N_A + 2N_B + \alpha N_A] \quad (3)$$

Onde:

$$diff = \begin{cases} 0 & , \text{ se } popMax \geq Q_{t+\frac{1}{2}} \\ \frac{Q_{t+\frac{1}{2}} - popMax}{Q_{t+\frac{1}{2}}} & , \text{ se } popMax < Q_{t+\frac{1}{2}} \end{cases} \quad (4)$$

e

$$Q_{t+\frac{1}{2}} = 2(N_A + N_B) \quad (5)$$

$Q(t)$ = Quantidade de moléculas no tempo “t” (ciclo ou round);

α = Taxa de mutação;

β = Taxa de eficiência de filtro;

N_A = Quantidade de moléculas afins;

N_B = Quantidade de moléculas não-afins;

Diff= Corte na quantidade de moléculas;

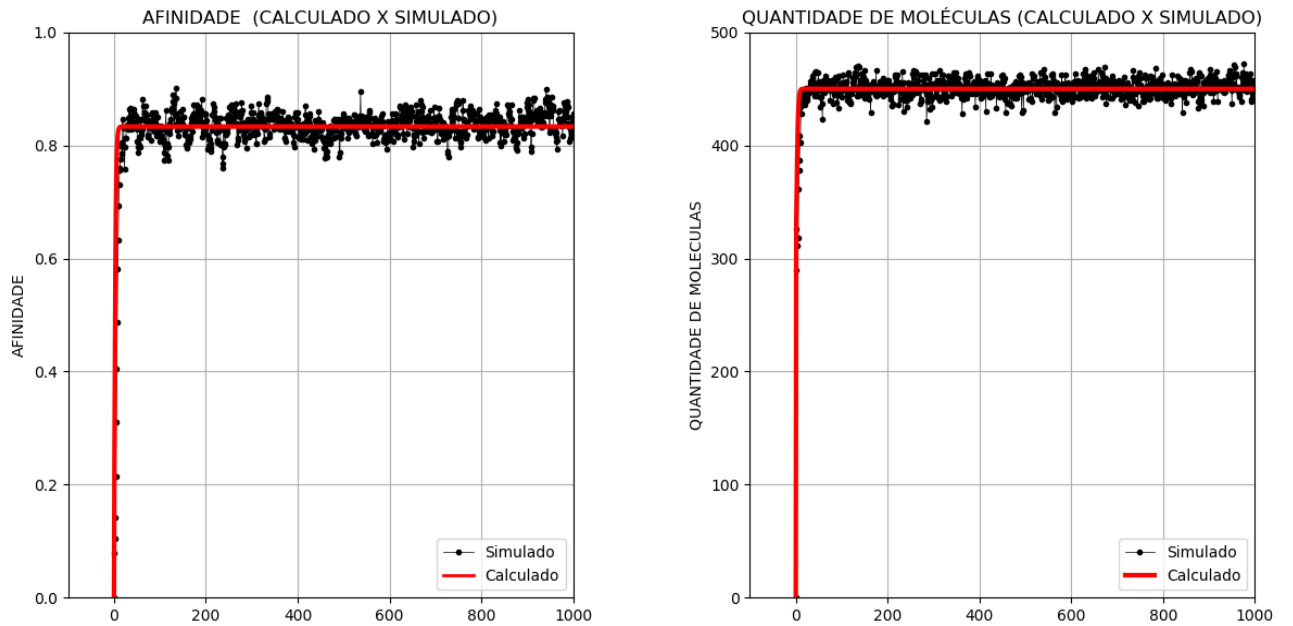


Figura 16: Cenário 1- Comparando método Simulado com Calculado.

Esse foi sem dúvidas o melhor resultado dos cenários, tanto a afinidade quanto a quantidade de moléculas são bem parecidas, foi testado com a eficiência de filtro 40% e variando a mutação entre 5%, 10%, 15% e 20%. Na figura 15 é referente a 5% de taxa de mutação. A equação 5 se refere a ação do ciclo onde temos sua replicação, já na equação 4 representa o fator de diminuição, caso as moléculas passem de 500 elas são cortadas até o seu limite. Na equação 3 é modelada com bases nos passos do método simulado.

5.2. Cenário 2

O cenário 2 teve incremento de duas novas variáveis, quebra e junção. A intenção é analisar o quão diferente vai ser essa mudança entre o cenário 1 e 2, como vai ser o comportamento do tamanho das moléculas e a afinidade e quantidade perante essas modificações. Novamente, para termos noção das simulações e um bom equilíbrio dos resultados, foi considerado por padrão de valores fixos do cenário os seguintes: tamanho inicial de 50 bases, 500 moléculas iniciais, 20% eficiência de filtro (chance de uma moléculas não-afim morrer), 0% de chance de uma molécula afim morrer, 5% de taxa de mutação, 500 moléculas de totais no *round* e o tamanho do filtro era de 5 bases onde a sequência era escolhida aleatoriamente no início. Quanto aos valores de quebra e junção foram: 100% de chance de uma molécula se quebrar caso encontre a sequência de 3 bases escolhida “*TTT*” o *stop códon*, já o processo de junção, foram escolhidas as 5 bases complementares da sequência escolhida pelo filtro para uma molécula ser apta (se a sequência afim fosse *AATTG* seu complemento seria *TTAAC*) com 100% de chance de junção.

5.2.1. Variando a taxa de mutação

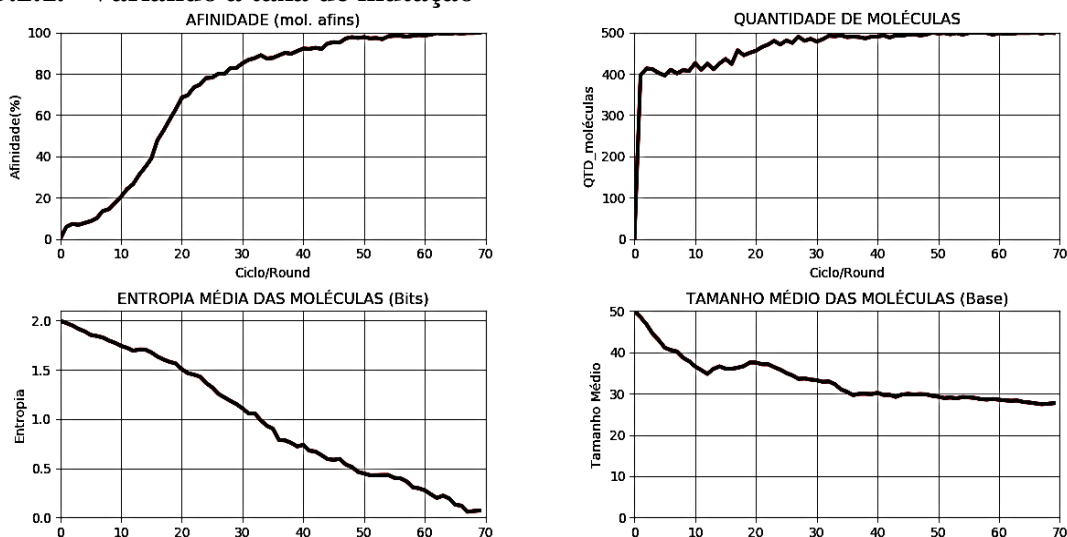


Figura 17: Cenário 2- Taxa de mutação 0%.

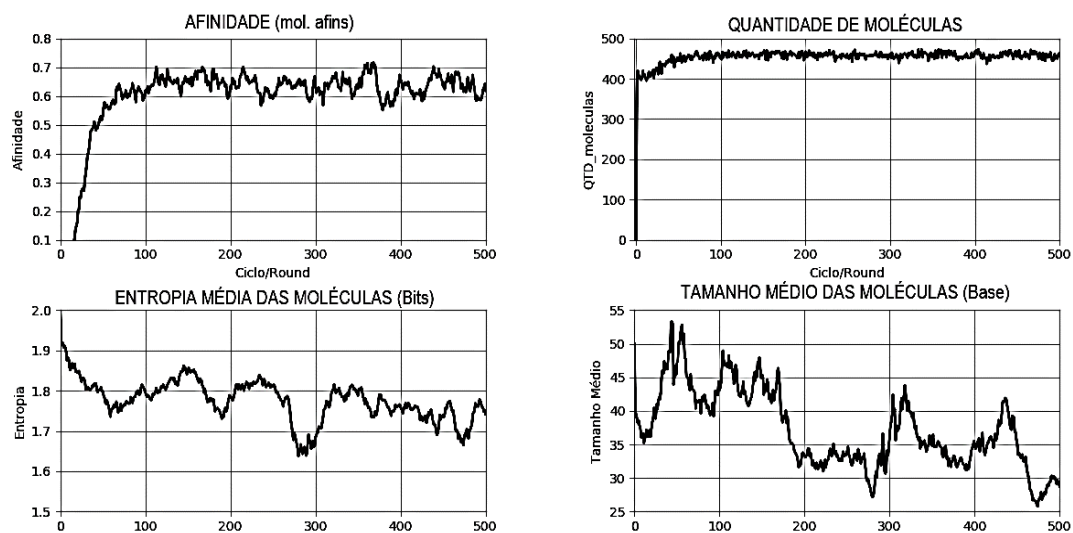


Figura 18: Cenário 2- Taxa de mutação 5%.

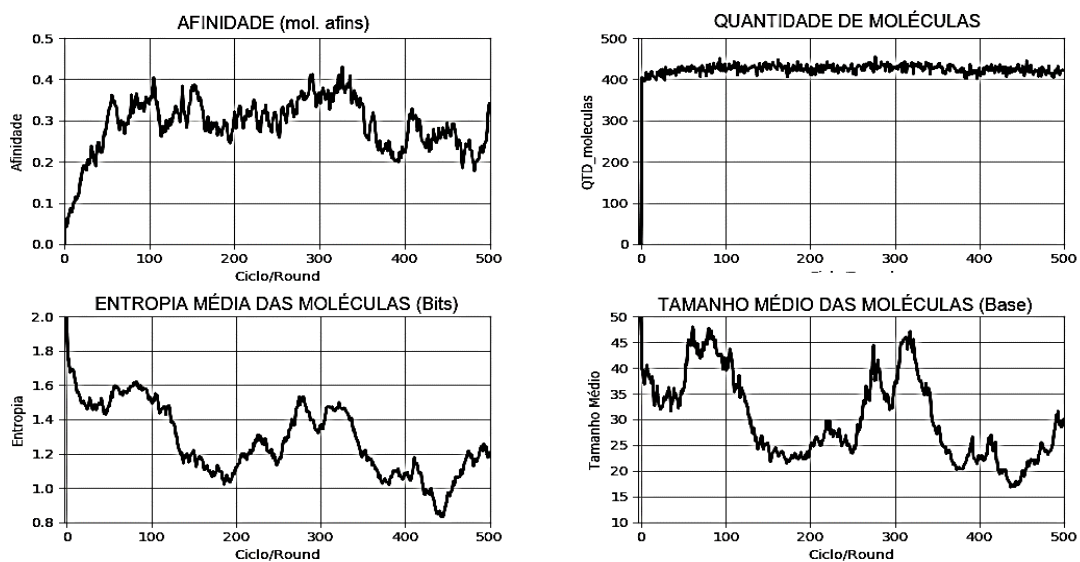


Figura 19: Cenário 2- Taxa de mutação 10%.

5.2.2. Variando a eficiência do filtro

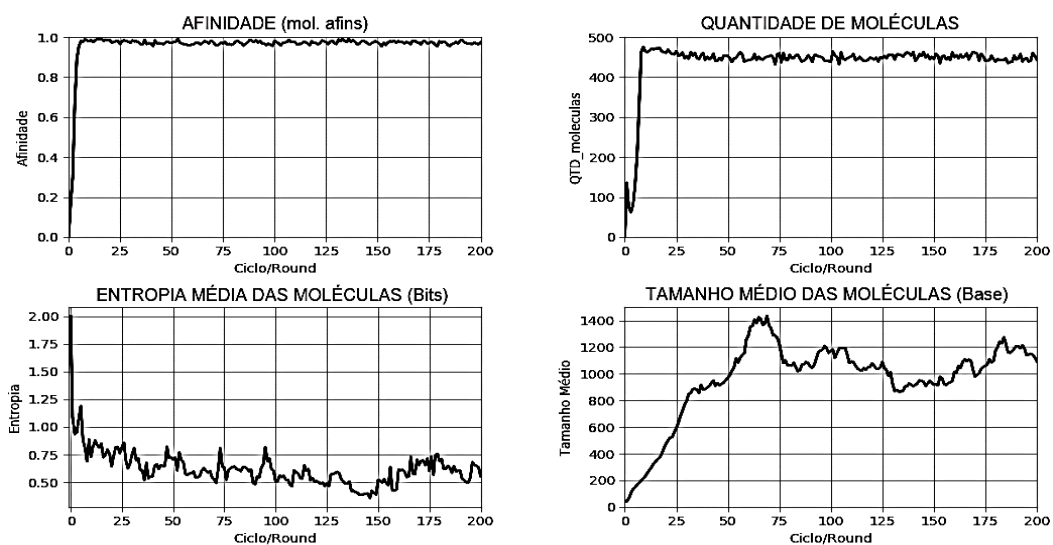


Figura 20: Cenário 2- Eficiência de filtro 80%.

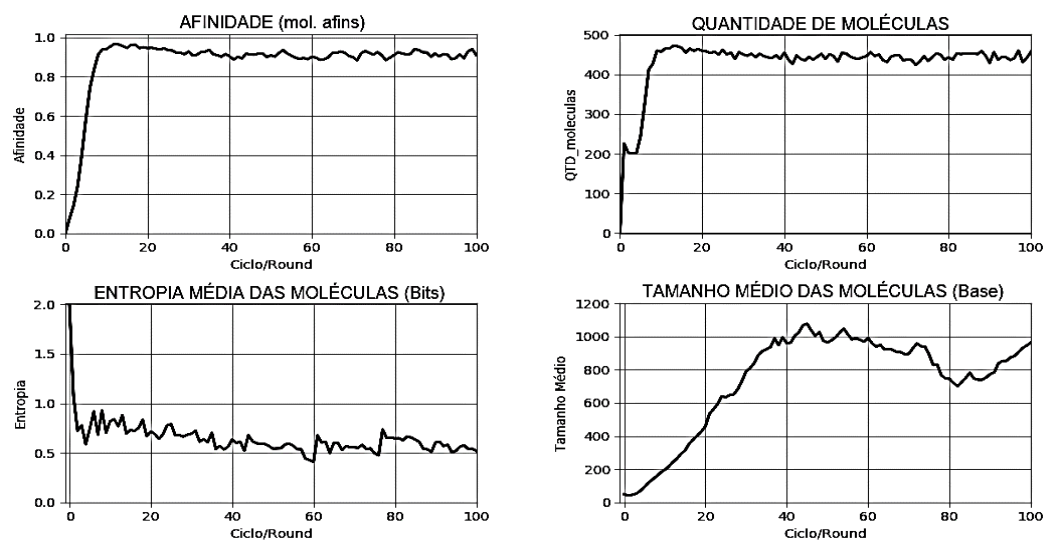


Figura 21: Cenário 2- Eficiência de filtro 60%.

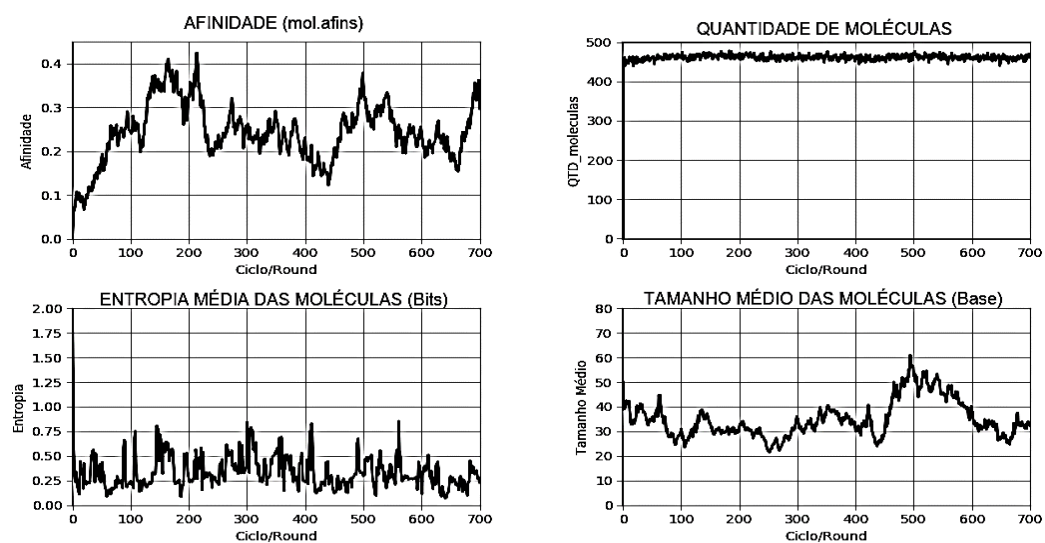


Figura 22: Cenário 2- Eficiência de filtro 10%.

As figuras 16-21 sobre o cenário 2 mostram que o tamanho médio das moléculas junto a quantidade de moléculas nunca se mostrava preciso a cada simulação e bem instável ao decorrer dos ciclos, mostrando assim, que não foi escolhido o melhor método para simulação.

5.3.Cenário 3

O último cenário foi bem mais complexo pois todas as moléculas que morrem durante os ciclos são recolhidas as suas bases e colocadas no reservatório, onde, para haver replicação de moléculas é necessário que tenha bases suficientes para a replicação, caso não tenha a quantidade de bases necessária, a etapa é pulada o sistema parte para a próxima. O sistema começa e termina com 50.000 moléculas na simulação, ou seja, o valor fixo do cenário são os seguintes: tamanho inicial de 100 bases, 500 moléculas iniciais, 20% eficiência de filtro (chance de uma moléculas não-afim morrer), 0% de chance de uma molécula afim morrer, 5% de taxa de mutação, 500 moléculas de totais no *round* e o tamanho do filtro era de 5 bases onde a sequência era escolhida aleatoriamente no início. Quanto aos valores de quebra e junção foram: 10% de chance de uma molécula se quebrar caso encontre a sequência de 2 bases no *stop códon*, para junção foram escolhidas as 5 bases complementares da sequência escolhida pelo filtro para uma molécula ser apta (se a sequência afim fosse *AATTG* seu complemento seria *TTAAC*) com 10% de chance de junção.

5.3.1. Variando a taxa de mutação

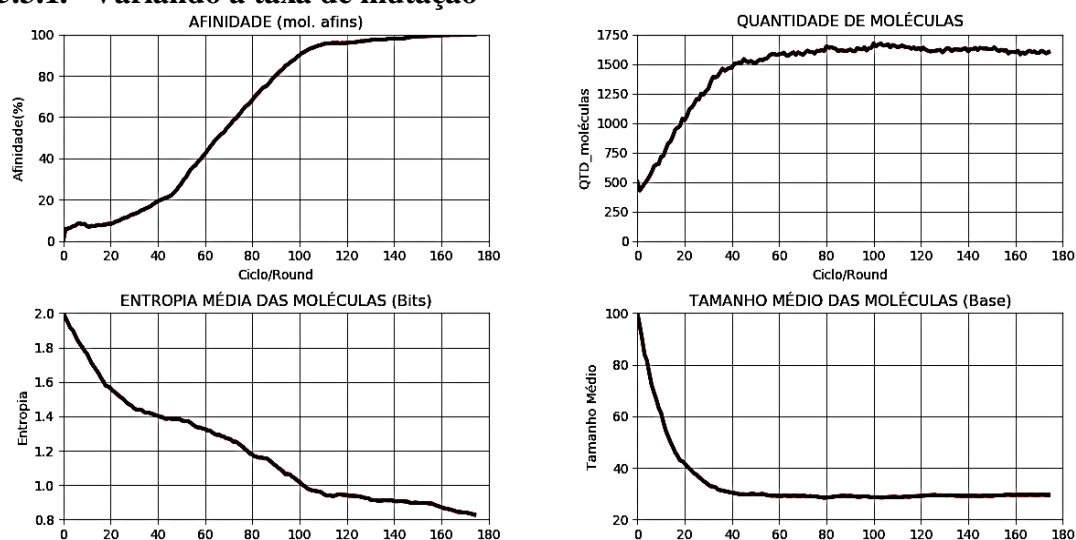


Figura 23: Cenário 3- Taxa de mutação 1%.

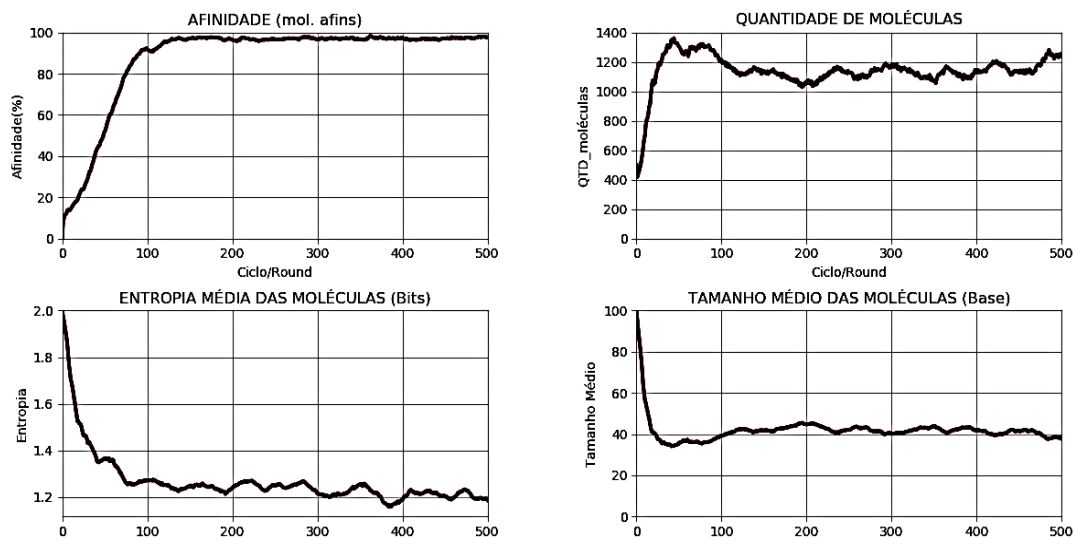


Figura 24: Cenário 3- Taxa de mutação 2%.

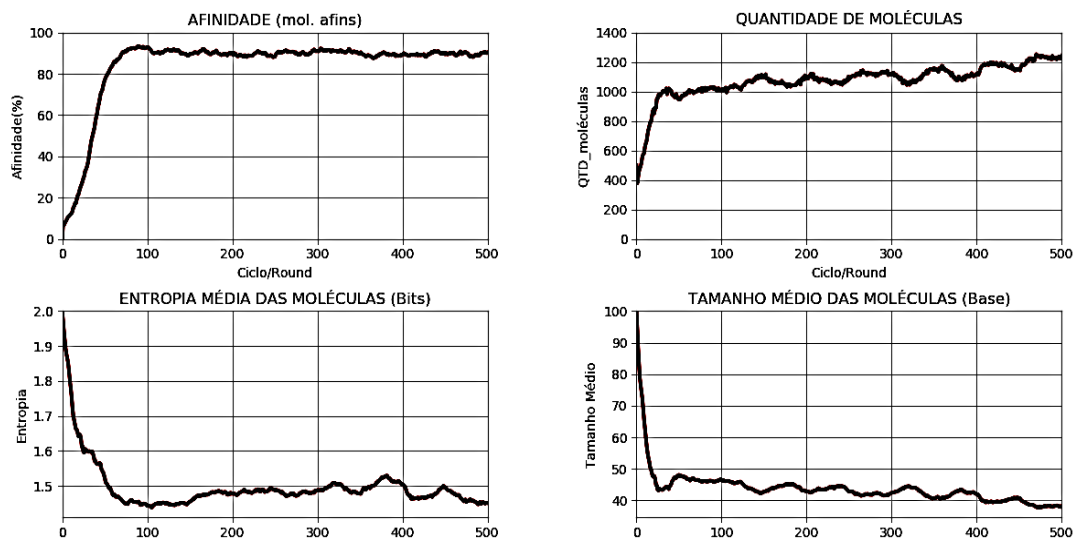


Figura 25: Cenário 3- Taxa de mutação 5%.

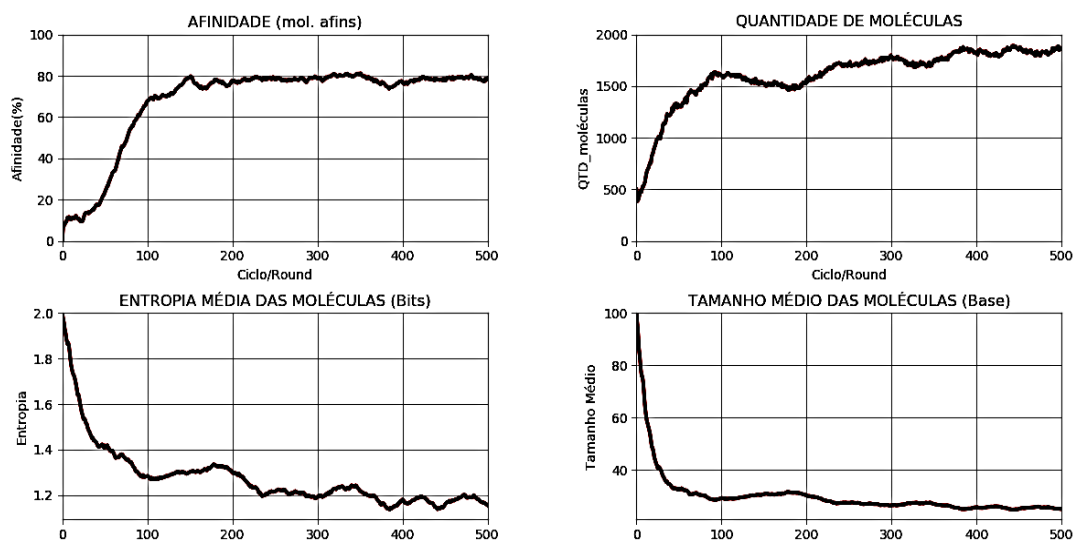


Figura 26: Cenário 3- Taxa de mutação 10%.

Após análise das figuras 22-25 é perceptível foi um avanço na questão da quebra e junção estarem bem mais suave no decorrer da simulação, porém podemos perceber pelo tamanho das moléculas que cada vez mais iam diminuindo lentamente após a instabilidade dos 100 primeiros ciclos, podemos tirar como conclusão 2 grandes evidencias, sendo a primeira delas que mesmo a quebra e junção tenho 10% de probabilidade de acontecer, o processo de quebra foi mais forte. A segunda é que, a molécula ser pequena tendo no mínimo suficiente para estar no ciclo pode algo benéfico pois estamos falando de uma molécula apta e pequena tornando-a forte e compacta.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Oró, J. Comets and formation of biochemical compounds on primitive Earth. *Nature* 190:380 (1961).
- [2] Tuerk, C., Gould, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, v.3, p. 505-510 (1990).
- [3] Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, p. 379-423, 623-656 (1948).
- [4] Cover, T. et al. *Elements of Information Theory*. New York: John Wiley (2001). (Wiley Series).
- [5] EIGEN, M. Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, v. 56: 465, 1971.
- [6] EIGEN, M.; SCHUSTER, P. *The Hypercycle, a Principle of Natural Self-Organization*. Berlin, Germany: Springer-Verlag, 1979.
- [7] GILBERT, W. The RNA world. *Nature*, v. 319, p. 618, 1986.
- [8] CECH, T.R. Self-splicing RNA: implications for evolution. *International Review of Cytology*, v. 93, p. 3-22, 1985.
- [9] CECH, T.R et al. Biological catalysis of RNA. *Annual Review of Biochemistry*, v. 55, p. 559-629, 1986.
- [10] MÜLLIS, K.B. "The unusual origin of the polymerase chain reaction". *Scientific American*, v. 262, n. 4, p. 56-61, 1990.
- [11] JOYCE, G. F. Amplification, mutation and selection of catalytic RNA. *Gene*, v.82, p. 83-87, 1989.
- [12] ELLINGTON, A. D.; SZOSTAK, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature*, v. 346, p. 818-822, 1990.
- [13] TUERK, C.; GOLD, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, v. 3, p. 505-510, 1990.

- [14] ULRICH, H.; MAGDESIAN, M. H.; ALVES, M. J.; COLLI, W.; HENNING, U. In vitro selection of RNA aptamers that bind to cell adhesion receptors of Trypanosoma cruzi and inhibit cell invasion. *Journal of Biological Chemistry*, v. 277, p. 20756-20762, 2002.
- [15] TANG, Z. et al. Selection of Aptamers for Molecular Recognition and Characterization of Cancer Cells. *Analytica Chemistry*, v. 79, p. 4900-4907, 2007.
- [16] ZHANG, Z. Advances in SELEX and application of aptamers in the central nervous system. *Biomolecular Engineering*, v. 24, p. 583–592, 2007.
- [17] JOYCE, G.F. RNA Evolution and origins of life. *Nature*, v. 338, p. 217-224, 1989.
- [18] ORGEL, L.E. Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in Biochemistry and Molecular Evolution*, v. 33, p. 99-123, 2004.
- [19] LEVINE, H.; NILSEN-HAMILTON, M. A. A mathematical analysis of SELEX.