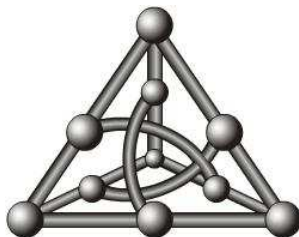

Temperatura de Melting: um estudo comparativo

Trabalho de Conclusão de Curso

Reginaldo Gomes de Arruda Júnior

Faculdade de Computação
Universidade Federal de Mato Grosso do Sul



Orientadora: Prof. Dra. Luciana Montera

Campo Grande, MÊS de 2010

Sumário

1	Conceitos da Biologia Molecular	3
1.1	Material Genético	3
1.2	Reação de PCR	6
1.3	Projeto de <i>Primers</i> para Reações de PCR	10
1.3.1	<i>Repeats, Runs</i> e Estruturas Secundárias	10
1.3.2	Especificidade e Tamanho do <i>primer</i>	11
1.3.3	%CG e Extremidade 3'	11
1.3.4	Temperatura de Pareamento	11
2	Conceitos de Termodinâmica e o Modelo <i>Nearest-Neighbor</i>	13
2.1	Termodinâmica	13
3	Cálculo da Temperatura de Melting	18
3.1	Básicas	18
3.2	Dependentes de Sal	18
3.3	Baseadas na Termodinâmica	19
4	Metodologia	20
4.1	Instrumental Utilizado	20
4.2	Cronograma	20
5	Resultados	21
5.1	Experimentos com o tamanho do <i>primer</i>	21
5.1.1	Médias das diferenças entre as T_m 's para <i>primers</i> de diferentes tamanho	23
5.1.2	Correlação entre T_m 's para diferentes tamanhos de <i>primers</i>	28

5.2	Experimentos com a concentração de GC nos <i>primers</i>	32
5.2.1	Médias das diferenças entre as T _m 's para <i>primers</i> com diferentes concentrações de GC	34
5.2.2	Correlação entre as T _m 's para diferentes concentrações de GC	39
6	Conclusão	44

Resumo

A Bioinformática é uma área de pesquisa com raízes em Ciência da Computação, Estatística e Biologia Molecular. Ela foi criada para lidar, entre outros, com os dados resultantes de projetos do sequenciamento genômico. Projetos deste tipo visam, por exemplo, o sequenciamento do DNA e a identificação dos genes existentes em organismos. Um volume de dados cada vez maior vem sendo produzido por projetos genoma. A bioinformática atua nas tarefas de armazenamento, consulta e manipulação desses dados, sendo, muitas vezes necessária a aplicação de métodos estatísticos capazes de analisar grandes quantidades de dados biológicos.

A reação de PCR (*Polymerase Chain Reaction*), é um método que pode ser utilizado para rápida duplicação de sequências de DNA em grandes quantidades. Para que a amplificação de uma sequência genética ocorra, segmentos específicos de ácidos nucleicos são necessários para iniciar o processo. Estes segmentos são chamados de *primers* (iniciadores). Os *primers* são sintetizados de modo a possuir uma sequência complementar ao trecho inicial/final da molécula de DNA que se deseja amplificar.

Muitas das técnicas de Biologia Molecular envolvendo reações de PCR dependem da precisão de predição de um valor experimental de um parâmetro crítico envolvido nesta reação: a temperatura de melting (T_m). Muitos programas de computador, baseados em diferentes métodos, estão disponíveis para estimativas teóricas do valor da (T_m) dada alguma sequência genética. Entretanto, na maioria dos casos, significantes diferenças nas estimativas da T_m são obtidas utilizando-se diferentes métodos. Por isso, é difícil decidir qual é o mais preciso.

Este trabalho compreende um estudo dos diferentes métodos para o cálculo da T_m a fim de desenvolver uma ferramenta computacional que permita comparações dos resultados obtidos por estes.

Introdução

Muitas técnicas utilizadas na análise de moléculas de DNA requerem que a sequência esteja disponível em grandes quantidades para viabilizar a realização de diversos experimentos. A possibilidade de se criar inúmeras cópias de uma molécula de DNA em laboratório otimiza este trabalho, uma vez que a molécula em questão não precisa ser extraída diretamente do organismo ao qual pertence.

O processo de produção em laboratório (*in vitro*) de cópia de uma molécula de DNA é possível através de reações de PCR (*Polymerase Chain Reaction*). Uma das principais aplicações da PCR é na medicina forense, onde pequenas amostras de DNA retiradas da cena de um crime (pedaços de cabelo, gotas de sangue ou saliva ou até mesmo a minúscula quantidade de DNA deixada em uma impressão digital) são amplificadas (replicadas) para serem analisadas. A PCR também é rotineiramente utilizada em procedimentos científicos de Biologia Molecular tais como amplificação e identificação de patógenos (por exemplo, vírus) presentes nas amostras ou ainda preparação de fragmentos de DNA para clonagem.

Na reação de PCR, em primeiro lugar, deve-se extrair o material genético da célula ou do material a ser estudado. Depois de extraído, o DNA é submetido a uma mistura que contém os componentes necessários para que a duplicação ocorra. Um dos principais componentes é o *primer*, sequência iniciadora do processo de síntese e corresponde a uma curta cadeia de nucleotídeo complementar ao início da sequência alvo (sequência de DNA que se deseja ampliar).

A PCR corresponde a execução de uma série de reações que ocorrem em diferentes temperaturas pré-estabelecidas com tempos exatos e específicos para cada fragmento de DNA a ser amplificado. Cada ciclo é dividido em etapas bem definidas. Na primeira etapa a temperatura é elevada para que haja a separação da dupla cadeia de DNA (*desnaturação*). Na segunda etapa, a temperatura é reduzida para que os *primers* se pareiem com a fita molde de DNA (*pareamento*). Na última etapa do ciclo a temperatura é elevada para que uma enzima sintetizadora (enzima *polimerase*) atue na extensão do *primer* e assim dar origem a uma da nova molécula (*extensão*). Essas três etapas são executadas ciclicamente por um número definido de vezes.

Diversos parâmetros tais como, o número de ciclos, a temperatura e o tempo de duração de cada etapa, bem como a qualidade e a quantidade de *primers* utilizados podem interferir na reação de PCR.

Para cada uma das etapas de uma reação, (desnaturação, pareamento e extensão) existe uma temperatura ideal, na qual a performance da reação é máxima. Define-se

temperatura de melting (T_m) ou temperatura de pareamento, como a temperatura na qual metade das moléculas estão pareadas e a outra metade não. São encontradas na literatura diferentes formulações para o cálculo da T_m , todas elas sendo apenas uma aproximação da temperatura de melting real de um *primer*.

Como já exemplificado, diversas são as técnicas/procedimentos de Biologia Molecular que se baseiam na reação de PCR (clonagem, amplificação, mutagênese, etc) e, consequentemente envolvem o projeto de *primers* bem como o cálculo da temperatura de melting.

Este trabalho tem por objetivo geral o estudo e implementação dos diferentes métodos de cálculo da T_m a fim de avaliar quão divergentes são os resultados produzidos pelas mesmas. Deste estudo resultará a implementação de um software cuja funcionalidade específica é: dado um *primer*, ou um par de *primers*, o usuário terá como saída diferentes temperaturas de melting calculadas por diferentes fórmulas, bem como estatísticas como a média, variância e o desvio padrão dos valores obtidos. De posse dessas informações, o usuário será capaz de avaliar e decidir qual a temperatura mais adequada a ser aplicada no experimento ou ainda se outros *primers* devem ser avaliados, caso, por exemplo, valores altos para variância e o desvio padrão sejam obtidos.

O restante deste texto está organizado como segue: no Capítulo 1 são apresentados os conceitos da Biologia Molecular essenciais para o entendimento do trabalho. No Capítulo 2 são abordados conceitos de termodinâmica que são utilizados nas formulações de cálculo da temperatura de melting. O Capítulo 3 aborda as formulações e suas respectivas classificações. O Capítulo 4 é descreve a metodoliga seguida durante o desenvolvimento deste trabalho. No capítulo 5 os resultados. Por fim, no Capítulo 6 encontra-se a conclusão do trabalho.

Capítulo 1

Conceitos da Biologia Molecular

Os conceitos apresentados neste capítulo foram, na sua maioria, compilados de [1], [2], [3] e [4].

1.1 Material Genético

Ácido desoxirribonucléico, ou DNA, é uma substância química envolvida na transmissão de caracteres hereditários e na produção de proteínas, compostos estes que são os principais constituintes dos seres vivos. Uma molécula de DNA é um polímero linear de nucleotídeos conectados entre si via *ligações covalentes*¹ denominadas ligações de *fosfodiéster*. A seguir será dada uma breve descrição dos componentes e da estrutura do DNA.

Os nucleotídeos, unidades básicas dos ácidos nucleicos, são constituídos de (Figura 1.1):

- uma base nitrogenada: anel heterocíclico de átomos de carbono e nitrogênio;
- uma pentose: açúcar com cinco carbonos; e
- um grupo fosfato: molécula com um átomo de fósforo cercado por quatro oxigênios;

As bases nitrogenadas são de dois tipos: púricas (adenina (A) e guanina (G)) e pirimídicas (timina (T), citosina (C) e uracila (U)). As purinas (bases púricas) são constituídas de dois anéis fundidos de cinco e seis átomos e as pirimidinas de um único anel de seis átomos, veja a Figura 1.2. A base uracila não está presente em cadeias de DNA (apenas no RNA).

Dois tipos de pentoses (resíduos de açúcar) são encontrados nos ácidos nucleicos: **ribose** e **desoxirribose** as quais diferem uma da outra pela presença ou ausência do grupo hidroxila no carbono 2 (C2') da pentose, Figura 1.3. É baseado nesta característica, dentre outras, que os ácidos nucleicos recebem o nome de RNA (ribose) ou DNA (desoxirribose).

¹ligação química entre átomos que tende a ser estável, mais forte do que outros tipos de ligações

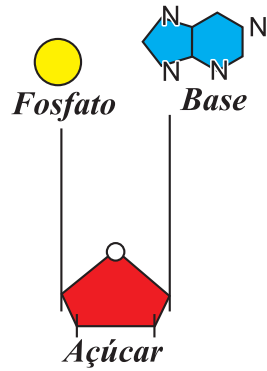


Figura 1.1: Estrutura básica de um nucleotídeo e sua composição: uma base nitrogenada, uma pentose (açúcar) e um grupo fosfato.

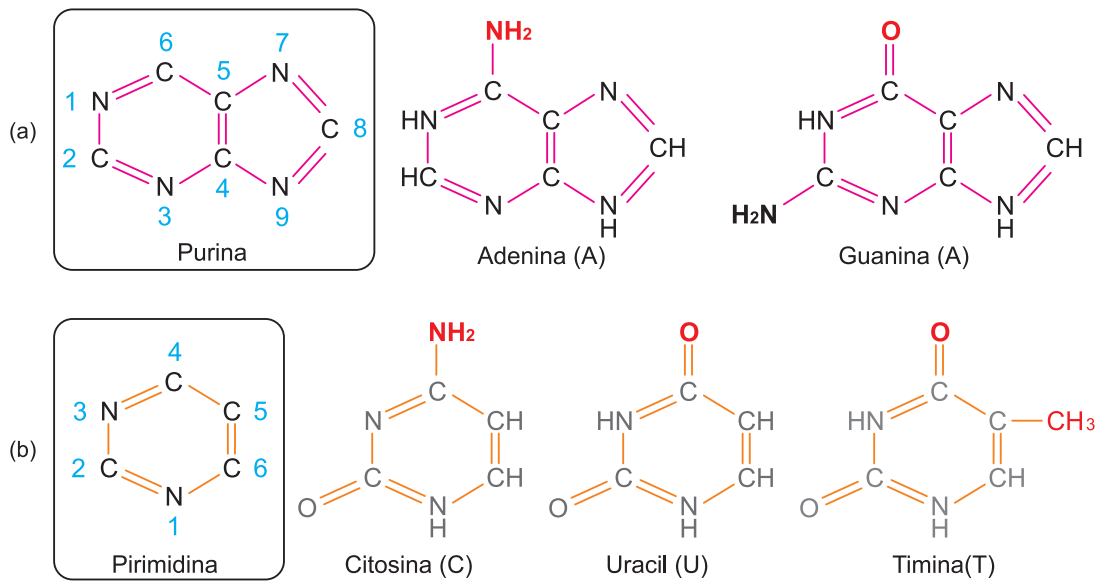


Figura 1.2: Estrutura das bases púricas (a) e das bases pirimídicas (b).

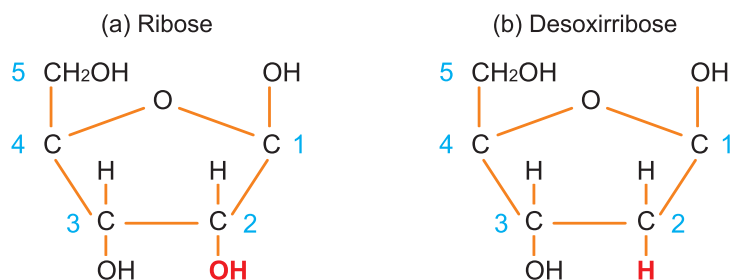


Figura 1.3: Dois tipos de pentoses. A pentose do RNA (a) possui hidroxila (OH) no C2'. A pentose do DNA (b) não possui o grupo hidroxila.

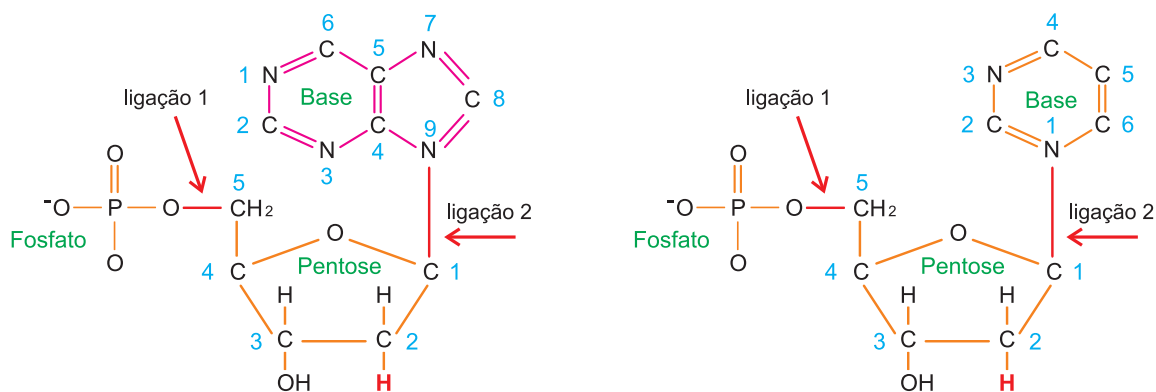


Figura 1.4: Ligações em um nucleotídeo

O ácido ribonucleico (RNA) é um polímero de nucleotídeos, assim como o DNA, responsável pela síntese de proteínas da célula. Porém, geralmente é encontrado em cadeia simples e possui dimensão muito inferior que o DNA. Este trabalho enfoca a replicação de DNA, assim, o conceito de RNA não será mais abordado.

A molécula de DNA é uma dupla hélice cujas cadeias estão unidas por **pontes de hidrogênios** estabelecidas entre purinas e pirimidinas das fitas opostas. Adenina sempre paria com timina (A-T) e guanina com citosina(G-C). Veja a Figura 1.5.

Como dito anteriormente, os nucleotídeos são formados por três diferentes tipos de moléculas (uma desoxirribose, um grupo fosfato e uma base nitrogenada). A orientação das ligações entre as três moléculas constituintes dos nucleotídeos é essencial para se determinar o sentido da fita de DNA. A ligação entre a base e a pentose é feita covalentemente através de uma ligação N-glicosídica com a hidroxila ligada ao C1' da pentose. Já a ligação entre o fosfato e a pentose é feita através de uma ligação fosfodiéster com a hidroxila ligada ao C5' da pentose. Veja a Figura 1.4.

Para a formação da molécula de DNA é necessário que ocorra a ligação entre os nucleotídeos. Eles estão ligados covalentemente por ligações fosfodiéster formando entre si pontes de fosfato. O grupo hidroxila do C3' da pentose do primeiro nucleotídeo se liga ao grupo fosfato ligado a hidroxila do C5' da pentose do segundo nucleotídeo através de

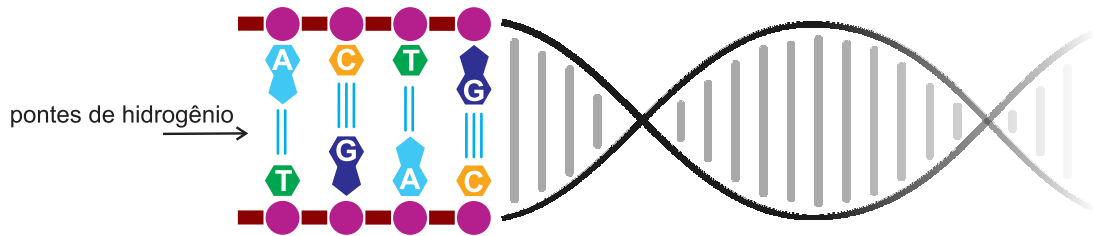


Figura 1.5: Estrutura básica da dupla fita de DNA. A base adenina (A) se liga co a timina (T) através de duas pontes de hidrogênio, enquanto que a citosina (C) se liga com a guanina (G) através de três pontes de hidrogênio.

uma ligação fosfodiéster. Devido a esta formação a cadeia de DNA fica com uma direção determinada. Em uma extremidade temos livre a hidroxila do C5' da primeira pentose e na outra temos livre a hidroxila do C3' da última pentose. Isto determina que o crescimento do DNA se faça na direção de 5' para 3'. Veja a figura 1.6.

As bases nitrogenadas estão localizadas na parte interna da molécula (devido a sua característica hidrofóbica). O pareamento das bases de cada fita se dá de maneira padronizada, sendo possível apenas a ligação entre A–T e C–G pela formação de pontes de hidrogênio entre as bases. Adenina forma duas pontes de hidrogênio com a timina e a citosina forma três pontes de hidrogênio com a guanina. Veja a figura 1.5.

Replicação do DNA é o processo de auto-duplicação do material genético de um organismo vivo mantendo assim o padrão de herança ao longo das gerações. Para que a replicação se inicie é necessário que a enzima *DNA helicase* ligue-se à cadeia de DNA e deslize sobre esta, quebrando as ligações (pontes de hidrogênio, figura 1.5) entre as duas cadeias de nucleotídeos, ficando então as duas cadeias de DNA separadas. Também, em altas temperaturas ou pH extremos o DNA sofre **desnaturação** isto porque ocorre ruptura das pontes de hidrogênio entre os pares de bases. Em seguida liga-se às cadeias de DNA, já separadas, uma enzima chamada *RNA primase* que vai sintetizar um *primer*. Logo após, a enzima *DNA polimerase* vai continuar o processo de sintetização onde os nucleotídeos livres ligam-se, por complementaridade de bases, às cadeias de DNA. De uma cadeia original de DNA formam-se duas novas.

1.2 Reação de PCR

O processo conhecido como síntese *in vitro* de DNA se refere à produção, em laboratório, de diversas cópias de uma molécula de DNA original. Em 1957 foi realizada a primeira síntese *in vitro* de DNA, por Arthur Kornberg, Uriel Littauer e sua equipe (LITTAUER; KORNBERG, 1957). A descoberta da enzima polimerase I isolada da bactéria *Escherichia coli*, viabilizou o desenvolvimento da técnica, uma vez que essa enzima atua na replicação do DNA.

Para que as enzimas do tipo polimerase atuem na replicação ou cópia de moléculas de

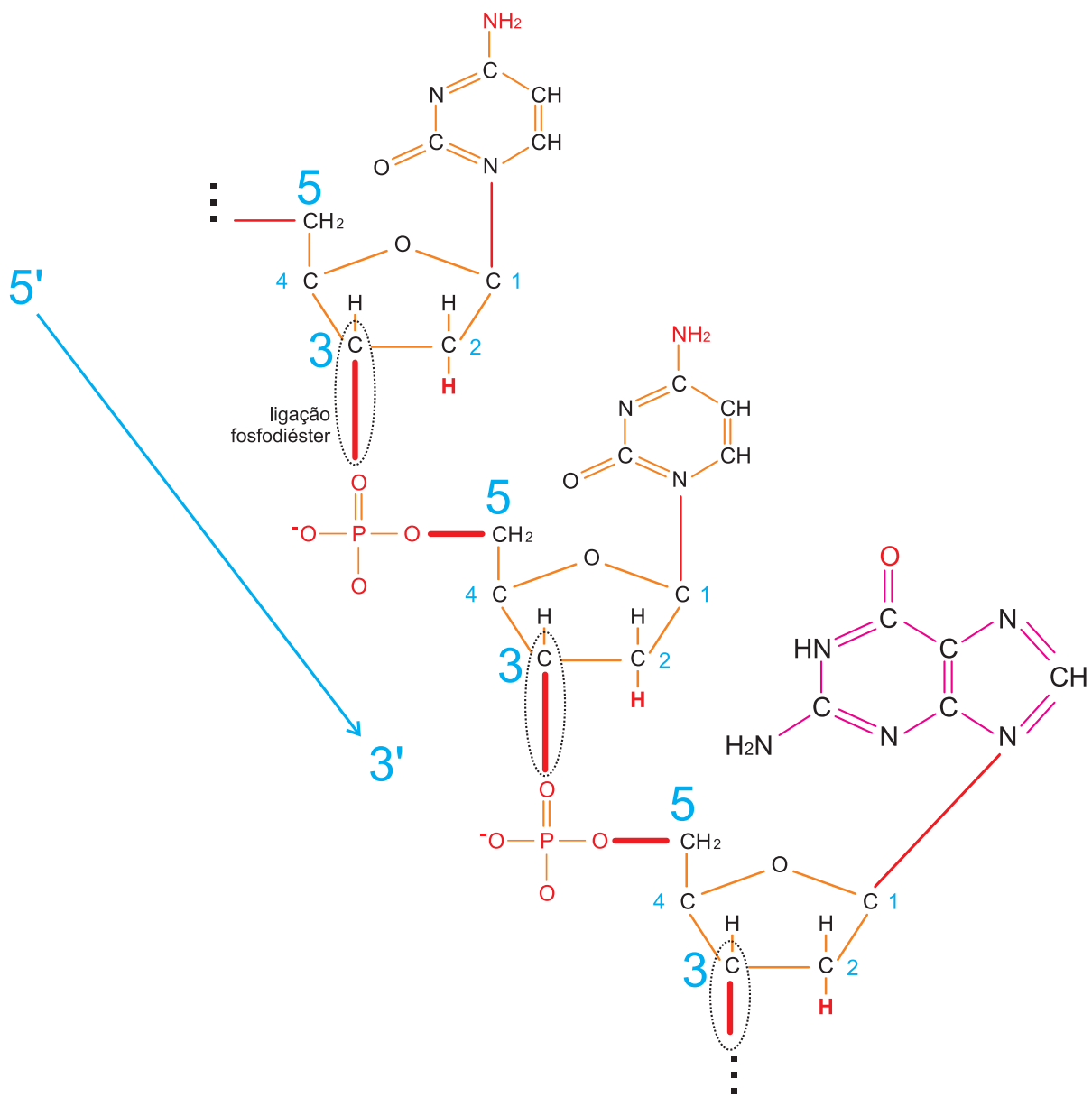


Figura 1.6: Direção da formação de cadeia de DNA

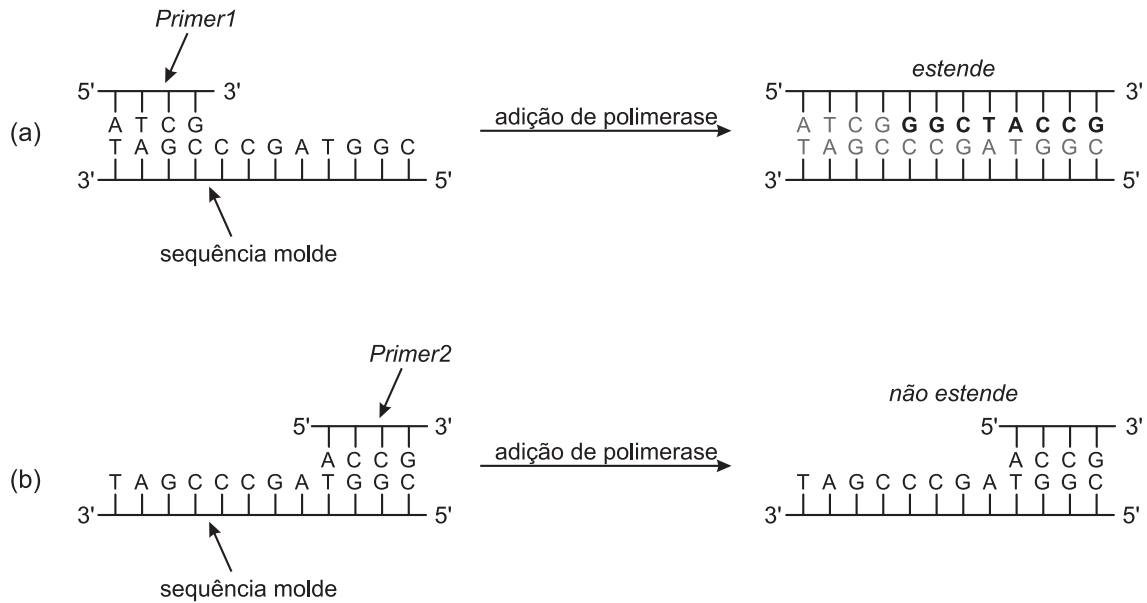


Figura 1.7: Atuação da enzima DNA polimerase na replicação do DNA. (a) A polimerase é capaz de estender o *Primer1* criando uma fita complementar ao molde. (b) A polimerase não é capaz de estender o *Primer2* – uma extensão não ocorre a partir da extremidade 5'. [Figura retirada de Montera (2008)]

DNA, é necessário a presença de três componentes básicos: *primers*, moldes e os quatro nucleotídeos (adenina (A), citosina (C), guanina (G) e timina (T)).

Primers são seqüências iniciadoras do processo de síntese e correspondem a curtas cadeias de nucleotídeos complementares à região inicial (*primer forward*) e região final (*primer reverse*) da seqüência que se deseja copiar, dita seqüência alvo. Em uma mesma reação, *primers* e seqüência alvo (a qual deve estar na forma desnaturada, ou seja, fita simples) se unem por complementaridade entre suas bases. Uma vez formado o complexo – *primer* e seqüência alvo – as enzimas do tipo polimerase ligam-se ao grupo OH da extremidade do carbono C3' do *primer* (extremidade 3' OH) para dar início ao processo de síntese. Para que a nova molécula seja sintetizada, a DNA polimerase utiliza a seqüência alvo como molde e, segundo o princípio da complementaridade entre as bases, adiciona novos nucleotídeos aos *primers*. Este processo recebe o nome de extensão. Ao final da extensão, uma nova molécula, complementar à seqüência molde, é produzida. Este tipo de reação, quando realizada em laboratório, é denominado PCR (*Polymerase Chain Reaction*) ou Reação Polimerase em Cadeia. Devido a necessidade do grupo 3' OH livre na seqüência do *primer* para que a enzima DNA polimerase se acople, tem-se que a síntese de uma nova molécula só ocorre no sentido 5' → 3'. Caso o *primer* pareie com a seqüência de DNA molde de forma que o sentido de extensão necessário seja 3' → 5', a polimerase não atua e, portanto, a extensão não ocorre. O esquema de atuação de enzimas do tipo polimerase é representado na Figura 1.7.

O termo amplificação é utilizado para caracterizar experimentos com vistas ao aumento

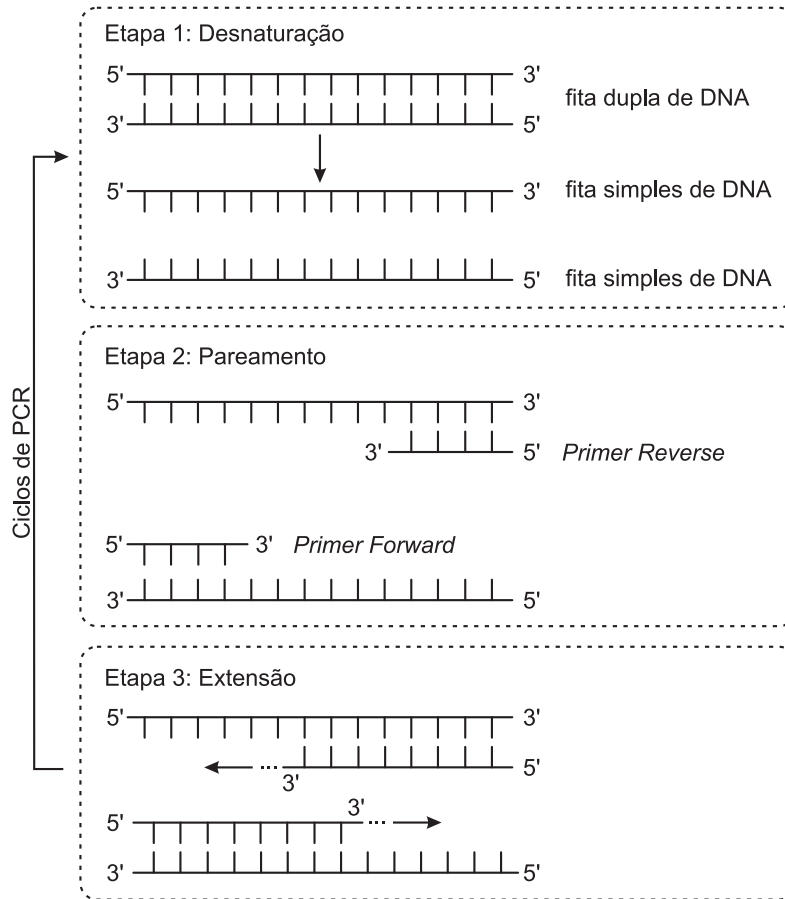
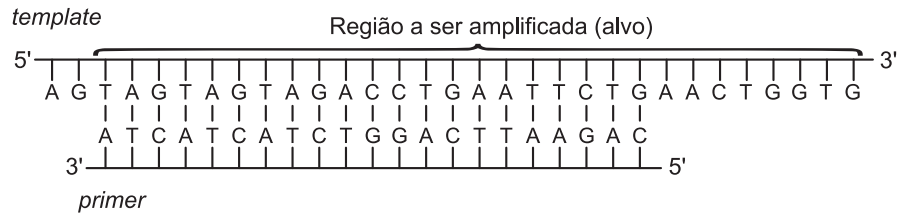


Figura 1.8: Ciclos de PCR com *primers*. A cada ciclo uma molécula de DNA dá origem a duas novas moléculas idênticas à molécula original. [Figura retirada de Montera (2008)]

(em número) de uma molécula. A amplificação de uma molécula de DNA em laboratório ocorre por meio de diversos ciclos de extensão por polimerase. No primeiro ciclo, a molécula de DNA alvo é desnaturada (veja a seção 1.1) resultando em duas sequências molde, as quais, após se parearem com os *primers*, servem como molde para extensão dos mesmos pela polimerase, resultando em duas novas moléculas de DNA idênticas à original. No segundo ciclo, as duas moléculas resultantes do ciclo anterior são desnaturadas, dando origem a quatro molde que, após o pareamento com os *primers* e a extensão, dão origem a oito novas moléculas de DNA idênticas à molécula original. Assim, após um número n de ciclos de PCR (desnaturação, pareamento e extensão) tem-se, idealmente, 2^n moléculas de DNA idênticas à molécula original. A Figura 1.8 ilustra as três etapas do ciclo de PCR.

Diversas variáveis e parâmetros, tais como o número de ciclos, a **temperatura** e o tempo de duração de cada ciclo, bem como a qualidade e a quantidade de *primers* utilizados podem interferir na reação de PCR e, conseqüentemente, o número 2^n de sequências de DNA resultantes pode não ser obtido após a execução de n ciclos de PCR.

(a) correto pareamento entre o *primer* e o *template*



(b) *misprimer* entre o *primer* e o *template*

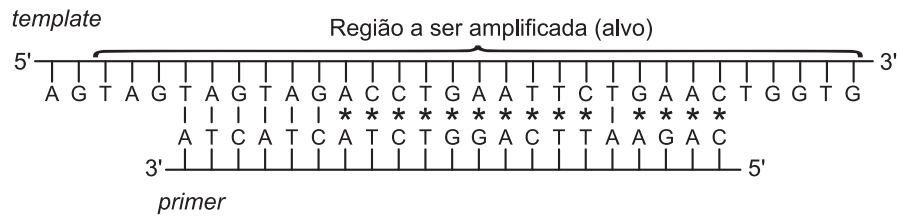


Figura 1.9: *Primer* com *repeats* (ATC). Os *matches* ocorridos no pareamento entre o *primer* e o *template* são representados pelo símbolo “|”, enquanto que os *mismatches* são representados por “*”. [Figura retirada de Montera (2008)]

1.3 Projeto de *Primers* para Reações de PCR

A determinação e escolha de um *primer* ou par de *primers* a ser utilizado em um experimento de PCR, ou baseado em PCR, requer uma série de cuidados quanto ao tamanho, composição, temperatura de pareamento (ou melting) e especificidade dos *primers*, entre outros. Não existem valores específicos pré-definidos para os parâmetros envolvidos no projeto de *primers*, contudo, intervalos de valores para esses parâmetros são de senso comum. Segue uma breve descrição dos parâmetros envolvidos no projeto de *primers*.

1.3.1 *Repeats*, *Runs* e Estruturas Secundárias

A ocorrência repetida de uma subsequência de nucleotídeos em posições consecutivas na sequência de um *primer* é chamada *repeat*. *Repeats* devem ser evitados uma vez que eles podem favorecer o pareamento entre o primer e o template em posições não desejadas (evento também chamado de *misprimer*, como mostrado na Figura 1.9). *Runs* são definidos como a repetição consecutiva de uma única base na sequência do *primer*.

Se os *primers forward* (F) e *reverse* (R) (veja a seção 1.2) apresentarem trechos complementares entre si, esses podem se parear, em detrimento do pareamento entre *primer* e o *template*, formando uma estrutura secundária chamada *hetero-dimer*. Caso o pareamento ocorra entre dois *primers* F ou dois *primers* R, ocorre a formação de uma estrutura secundária denominada *self-dimer*. *Primers* longos podem se auto-parear, resultando em

estrutura secundárias denominadas *hairpins*.

1.3.2 Especificidade e Tamanho do *primer*

Um *primer* é específico se ele se parecia com o *template* apenas na região específica para a qual ele foi projetado (ver região alvo da Figura 1.9). *Primers* não específicos acarretam na produção (amplificação) de fragmentos de DNA não correspondentes à região alvo.

O tamanho de um *primer* influencia a sua especificidade (*primers* maiores são mais específicos); custo de produção (quanto maior o *primer*, mais elevado é seu custo); estabilidade da formação *template-primer* (quanto maior o *primer*, mais fortemente ele estará unido ao *template* devido ao maior número de pontes de hidrogênio resultantes desta ligação); e formação de estruturas secundárias (*primers* maiores são mais propensos à formação de estruturas secundárias). Não existe um tamanho fixo ótimo para um *primer*. *Primers* variando entre 18 e 30 bases são os mais indicados (KAMEL; ABD-ELSALAM, 2003).

1.3.3 %CG e Extremidade 3'

O conteúdo de bases Citosina (C) e Guanina (G) de um *primer* determina a temperatura na qual a reação de pareamento deve ocorrer basicamente pelo fato do pareamento destas bases ocorrer devido à formação de três pontes de hidrogênio enquanto que o pareamento entre as bases A e T ocorre a formação de duas pontes de hidrogênio. Quanto maior o número de bases C e G, mais fortemente “ligados” estarão o *primer* e o *template*. Valores entre 40% e 60% de CG na composição dos *primers* são preferidos.

Preferencialmente, uma base C ou G deve compor a extremidade 3' de um *primer* por ser esta a extremidade na qual a enzima polimerase inicia a extensão do *primer*. Como o pareamento C–G é mais estável, espera-se que a polimerase inicie o processo de síntese mais eficientemente neste caso.

1.3.4 Temperatura de Pareamento

Cada uma das etapas da reação de PCR – desnaturação, pareamento e extensão (ver Figura 1.8) – ocorrem sob temperaturas “ideais”. A desnaturação, responsável por transformar a molécula de DNA de fita dupla em duas moléculas de fita simples ocorre por volta de 94°C. A temperatura de pareamento é dependente dos *primers* utilizados e a temperatura de extensão é dependente do tipo de enzima polimerase utilizada.

A temperatura de *melting* (T_m) é definida como a temperatura na qual metade dos fragmentos de DNA está na forma desnaturada, ou seja, não pareados, e a outra metade está pareada. A temperatura de pareamento ou *annealing* (T_a) é a temperatura na qual ocorre o pareamento entre *primer* e *template*. Tipicamente, a T_a difere de 3 a 5°C a menos da T_m dos primers.

Existem diversas maneiras de se calcular a T_m de um *primer*. De uma maneira geral, podemos classificar as fórmulas para o cálculo da T_m como:

1. Básicas: consideram apenas a composição do *primer*
2. Dependentes do Sal: consideram a concentração de sal na reação onde pareamentos ocorrem; e
3. Baseadas na Termodinâmica da reação: utilizam-se do modelo *Nearest Neighbor*.

O objetivo deste trabalho é fazer um levantamento das diferentes formulações para o cálculo da T_m bem como a implementação das mesmas a fim de comparar os resultados. Antes de apresentar a revisão bibliográfica sobre as formulações para o cálculo da T_m de *primers* (o que é feito no Capítulo 3), é necessário que conceitos de termodinâmica e do Modelo *Nearest Neighbor* sejam apresentados (Capítulo 2) já que diversas formulações para o cálculo da T_m se baseiam nestes conceitos.

Capítulo 2

Conceitos de Termodinâmica e o Modelo *Nearest-Neighbor*

Neste capítulo serão apresentados conceitos relacionados à termodinâmica que, segundo o Modelo *Nearest Neighbor* (NN), regem os eventos de pareamentos entre nucleotídeos de fitas complementares de DNA. Tais conceitos são necessários ao entendimento de fórmulas para o cálculo da T_m que se utilizam do modelo NN e que são apresentadas mais adiante no Capítulo 3.

O modelo *Nearest-Neighbor* (vizinho mais próximo) descreve a contribuição da entalpia e da entropia de cada par de bases vizinhas existentes em uma região de pareamento. Como referência básica para a descrição dos conceitos de termodinâmica apresentados a seguir utilizou-se [4].

2.1 Termodinâmica

A termodinâmica é um ramo da física que estuda as relações entre o trabalho e calor. Dentre os principais conceitos da termodinâmica estão os conceitos de entalpia e entropia de um sistema.

A entalpia H mede o conteúdo de energia térmica de um sistema, cuja variação (ΔH), mede o calor liberado ou absorvido por este sistema para uma dada transformação ocorrida sob pressão constante. A entropia S mede a desordem de um sistema e sua variação (ΔS) revela se o sistema sofreu aumento ou diminuição da desordem após a ocorrência de determinada reação. As variações de entalpia e entropia de um sistema ou reação são estabelecidas pelas equações (2.1) e (2.2), respectivamente.

$$\Delta H = H_{\text{produtos}} - H_{\text{reagentes}} \quad (2.1)$$

$$\Delta S = S_{\text{produtos}} - S_{\text{reagentes}} \quad (2.2)$$

nas quais:

ΔH	=	variação da entalpia do sistema
ΔS	=	variação da entropia do sistema
$H_{produtos}$	=	entalpia dos produtos
$H_{reagentes}$	=	entalpia dos reagentes
$S_{produtos}$	=	entropia dos produtos
$S_{reagentes}$	=	entropia dos reagentes

Com base nos valores das equações (2.1) e (2.2) tem-se:

$\Delta H < 0,$	indicativo que a reação libera calor, e por isso é dita exotérmica
$\Delta H > 0,$	indicativo que a reação absorve calor, e por isso é dita endotérmica
$\Delta S < 0,$	indicativo que o sistema, após a reação se torna menos desordenado
$\Delta S > 0,$	indicativo que o sistema, após a reação se torna mais desordenado

Uma reação ocorre mais espontaneamente sempre que:

- mais calor liberar, ou seja, quanto menor for seu ΔH ;
- mais desordenado se tornar o sistema após a reação, ou seja, quanto maior for seu ΔS .

É importante notar, entretanto, que a espontaneidade de uma reação depende da entropia total, isto é, da entropia do sistema e de sua vizinhança, e não apenas da variação associada ao sistema. Se for considerado, por exemplo, o universo como vizinhança, o cálculo se torna impraticável. Como solução, considera-se que as reações ocorrem sob pressão e temperatura constantes pois, desta forma, a entropia da vizinhança depende somente da quantidade de calor transferido da reação para a vizinhança e da temperatura na qual esse calor é transferido, representados pelas variáveis calor absorvido pela vizinhança (*calor_viz*) e temperatura (T), como mostra a eq. (2.3).

$$\Delta S_{vizinhanca} = \frac{calor_viz}{T} \quad (2.3)$$

Sob pressão e temperatura constantes, o calor absorvido pela vizinhança corresponderá ao calor liberado pelo sistema. Se o sistema libera calor para a vizinhança, ΔH é negativo. Desta forma, a eq. (2.3) pode ser reescrita como a eq. (2.4).

$$\Delta S_{vizinhanca} = \frac{-\Delta H_{sistema}}{T} \quad (2.4)$$

Assim, é possível calcular a entropia total como a soma das entropias do sistema mais a entropia da vizinhança:

$$\begin{aligned} \Delta S_{total} &= \Delta S_{sistema} + \Delta S_{vizinhanca} \\ \Delta S_{total} &= \Delta S_{sistema} - \frac{\Delta H_{sistema}}{T} \end{aligned}$$

$$-T\Delta S_{total} = \Delta H_{sistema} - T * \Delta S_{sistema} \quad (2.5)$$

A fim de facilitar a definição de reação espontânea, utiliza-se a função de Energia Livre de Gibbs, denotada por G , que estabelece uma relação entre entalpia e entropia:

$$G = H - T * S \quad (2.6)$$

Define-se a variação de energia livre de Gibbs (ΔG) como:

$$\Delta G = R * T * \ln(T) \quad (2.7)$$

Onde T é a temperatura expressa em Kelvin.

É possível agora definir a espontaneidade de uma reação com base apenas no valor da variação de sua energia livre. As relações entre variação de energia livre e espontaneidade de uma reação, considerando-se pressão (P) e temperatura (T) constantes, estão sumarizadas na tabela 2.1.

Tabela 2.1: Relação entre a variação da energia livre de Gibbs e a espontaneidade de uma reação.

ΔG (P e T constantes)	Reação
< 0	Espontânea
$= 0$	Em equilíbrio
> 0	Não Espontânea

Uma vez definida a energia livre de Gibbs, é possível apresentar o modelo de *Nearest-Neighbor* (NN). O modelo NN, descrito para ácidos nucleicos, assume que a estabilidade da ligação entre um determinado par de bases depende da identidade e da orientação dos pares de bases vizinhos a este em uma região de sobreposição (SANTALUCIA, 1988). Segundo o modelo, dois pares de bases vizinhos contribuem com um determinado valor de variação de entalpia (ΔH) e entropia (ΔS) para o cálculo da energia livre total correspondente à região de pareamento. Mais precisamente, de acordo com o modelo NN, a energia livre deve ser calculada como a soma de três termos distintos (FREIER, 1986):

- a energia livre de *iniciação* de uma hélice (fita dupla de DNA, seção 1.1) associada com a formação do primeiro par de base da fita dupla;
- soma da energia livre associada com cada um dos pares de bases subsequentes, sendo este evento denominado *propagação*; e
- energia associada à *simetria* das moléculas, ou seja, se estas são ou não complementares entre si.

Considere os pares de bases complementares A–T e C–G (seção 1.1). A notação utilizada para representar a ocorrência consecutiva $\begin{smallmatrix} 3' & TG & 5' \\ 5' & AC & 3' \end{smallmatrix}$ em uma região de sobreposição (fita dupla de DNA) é AC/TG. Dadas as quatro bases A, C, G e T que compõem a molécula de DNA, se uma base se alinhar com uma base não complementar (A–C ou A–T por exemplo), ocorre o que chamamos de **mismatch**, dezesseis pares de bases

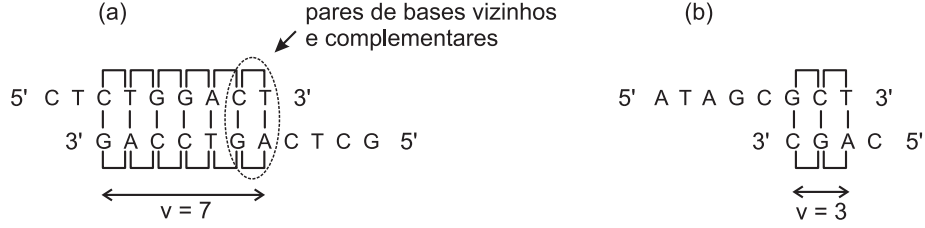


Figura 2.1: Duas regiões de sobreposição entre dois pares de fragmentos de DNA distintos. Os pares de bases vizinhos e consecutivos estão indicados pelos símbolos \square e \sqcup .

distintos e perfeitamente pareados, isto é, sem mismatches, podem ocorrer em uma região de sobreposição, são eles: AA/TT, TT/AA, AT/TA, TA/AT, CA/GT, TG/AC, GT/CA, AC/TG, CT/GA, AG/TC, GA/CT, TC/AG, CG/GC, GC/CG, GG/CC e CC/GG. Pares de bases com mismatches também podem ocorrer em uma região de pareamento.

A Figura 2.1 apresenta dois exemplos de regiões de sobreposição entre dois pares de fragmentos de DNA distintos, nas quais os pares de bases vizinhos que contribuem para a estabilidade do pareamento estão destacados dois a dois.

Seja v o tamanho de uma região de pareamento entre duas sequencias de fita simples de DNA. A variação de energia livre de Gibbs para este pareamento é dada pela eq. (2.8):

$$\Delta G = \sum_{i=1}^{v-1} (\Delta H_{i,i+1} - T \Delta S_{i,i+1}) + (\Delta H_{iniciacao} - T \Delta S_{iniciacao}) + (\Delta H_{simetria} - T \Delta S_{simetria}) \quad (2.8)$$

A Tabela 2.2 apresenta os valores de ΔS e ΔH propostos por Allawi e SantaLucia (1997), Sugimoto et al. (1996), Santa Lucia et al. (1996) e Breslauer (1986), para uma temperatura constante de 310K (ou 37°C) e 1M NaCl para os dezesseis pares de bases sem *mismatches*. A Tabela 2.2 traz ainda os valores sugeridos em cada uma das referências para os parâmetros *iniciação* e *simetria*. Valores de ΔS e ΔH para pares de bases com *mismatches* podem ser encontrados nas seguintes referências: (ALLAWI; SANTALUCIA, 1998), (ALLAWI; SANTALUCIA, 1997) e (HE et al., 1991).

Como exemplo, considere o cálculo da energia livre da região de sobreposição mostrada na Figura 2.1(a), considerando que a reação ocorre à temperatura de 310K e pressão constante, e os valores de ΔS e ΔH sugeridos por Allawi e SantaLucia (1997).

$$\begin{aligned} \Delta G &= (\Delta H_{CT/GA} - T * \Delta S_{CT/GA}) + (\Delta H_{TG/AC} - T * \Delta S_{TG/AC}) + (\Delta H_{GG/CC} - T * \Delta S_{GG/CC}) \\ &+ (\Delta H_{GA/CT} - T * \Delta S_{GA/CT}) + (\Delta H_{AC/TG} - T * \Delta S_{AC/TG}) + (\Delta H_{CT/GA} - T * \Delta S_{CT/GA}) \\ &+ (\Delta H_{iniciacao} - T * \Delta S_{iniciacao}) + (\Delta H_{simetria} - T * \Delta S_{simetria}) \\ &= (-7800 - 310 * -21, 0) + (-8500 - 310 * 22, 7) + (-8000 - 310 * -19, 9) + (-8200 - 310 * -22, 2) \\ &+ (-8400 - 310 * 22, 4) + (-7800 - 310 * 21, 0) + (100 - 310 * -2, 8) + (0 - 310 * -1, 4) \\ &= -8982 \text{ cal} = -8,982 \text{ kcal} \end{aligned}$$

Tabela 2.2: Valores de variação de entropia (ΔS) e entalpia (ΔH) para os pares de bases vizinhos.

Par NN	ALLAWI; SANTALUCIA 1997		SUGIMOTO et al., 1996		SANTALUCIA et al., 1996		BRESLAUER et al., 1986	
	ΔH	ΔS	ΔH	ΔS	ΔH	ΔS	ΔH	ΔS
AA/TT TT/AA	-7,9	-22,2	-8,0	-21,9	-8,4	-23,6	-9,1	-24,0
AT/TA	-7,2	-20,4	-5,6	-15,2	-6,5	-18,8	-8,6	-23,9
TA/AT	-7,2	-21,3	-6,6	-18,4	-6,3	-18,5	-6,0	-16,9
CA/GT TG/AC	-8,5	-22,7	-8,2	-21,0	-7,4	-19,3	-5,8	-12,9
GT/CA AC/TG	-8,4	-22,4	-9,4	-25,5	-8,6	-23,0	-6,5	-17,3
CT/GA AG/TC	-7,8	-21,0	-6,6	-16,4	-6,1	-16,1	-7,8	-20,8
GA/CT TC/AG	-8,2	-22,2	-8,8	-23,5	-7,7	-20,3	-5,6	-13,5
CG/GC	-10,6	-27,2	-11,8	-29,0	-10,1	-25,5	-11,9	-27,8
GC/CG	-9,8	-24,4	-10,5	-26,4	-11,1	-28,4	-11,1	-26,7
GG/CC CC/GG	-8,0	-19,9	-10,9	-28,4	-6,7	-15,6	-11,0	-26,6
Iniciação C/G	0,1	-2,8	0,6	-9,0	0	$-5,9 \pm 0,8$	0	-16,77
Iniciação T/A	2,3	4,1	0,6	-9,0	0	$-9,0 \pm 3,2$	0	-20,13
Simetria*	0	-1,4	0	-1,4	0	-1,4	0	-1,34

Unidade dos parâmetros: ΔH é dado em kcal/mol e ΔS é dado em cal/Kmol.

*a simetria só se aplica a sequências que são complementares entre si (*self*-complementares). Caso contrário o valor do parâmetro simetria é zero.

Capítulo 3

Cálculo da Temperatura de Melting

Neste capítulo, diferentes formulações são apresentadas para calcular a T_m de primers. Sendo essas formulações organizadas em três classes distintas: Básicas, Dependentes de Sal e Baseadas na Termodinâmica.

3.1 Básicas

Os cálculos básicos para o cálculo da T_m consideram apenas a composição do *primer*, ou seja, apenas a quantidade de cada base que compõe o *primer*. Marmur e Doty (1962) propuseram a seguinte fórmula para o cálculo da T_m :

$$T_m = 64.9 + 41.0 * \left(\frac{|G| + |C| - 16.4}{|A| + |T| + |G| + |C|} \right) \quad (3.1)$$

A fórmula sugerida por Wallace e colaboradores (Wallace *et al.*, 1979) é:

$$T_m = 2 * (|A| + |T|) + 4 * (|C| + |G|) \quad (3.2)$$

Ambas as equações (3.1) e (3.2) assumem que o pareamento ocorre em condições padrão de 50 nM de concentração de *primers*, 50mM de concentração de sal (Na^+) e pH próximo de 7.0. Em ambas, $|A|$, $|C|$, $|T|$ e $|G|$ representam a quantidade de bases Adenina, Citosina, Timina e Guanina, respectivamente.

3.2 Dependentes de Sal

Formulações dependentes do sal consideram a concentração de sal no buffer da reação de PCR.

Howley e colaboradores (Howley *et al.*, 1979) apresentaram a seguinte fórmula para o

cálculo da T_m :

$$T_m = 100.5 + 41.0 * \left(\frac{|G| + |C| - 16.4}{|A| + |T| + |G| + |C|} \right) - \left(\frac{820.0}{|A| + |T| + |G| + |C|} \right) + 16.6 * \log([Na^+]) \quad (3.3)$$

O termo $16.6 * \log([Na^+])$ ajusta a T_m para mudanças na concentração de sal.

3.3 Baseadas na Termodinâmica

As formulações para o cálculo da T_m que se baseiam no modelo NN consideram a composição da sequência do *primer* em termos das interações entre os pares de bases vizinhos definidas pelos valores de variação de entalpia e entropia, como descrito na seção 2.1.

SantaLucia e colaboradores (1996) descrevem a seguinte fórmula para o cálculo da T_m :

$$T_m = \frac{\Delta H + \Delta H_{iniciacao}}{\Delta S + \Delta S_{iniciacao} + \Delta S_{simetria} + R * \ln \frac{C_T}{b}} + 16.6 * \log([Na^+]) \quad (3.4)$$

onde ΔH e ΔS é a soma da variação da entalpia e da entropia, respectivamente, sobre todos os pares de bases internos; $\Delta S_{simetria}$ é a penalidade da entropia por sequências autocomplementares (simetria na Tabela 2.2, e $\Delta H_{iniciacao}$ e $\Delta S_{iniciacao}$ são as somas iniciais de variações entalpias e entropias, respectivamente (Tabela 2.2); R é a constante dos gases (fixada em 1.987 cal/K* mol) e C é a margem de concentração do *primer* na solução dada em mol. A constante b recebe o valor 4 para sequências não-autocomplementares ou valor 1 para autocomplementares.

Rychilik e colaboradores (1990) propôs a seguinte fórmula para o cálculo da T_m :

$$T_m = \frac{\Delta H}{\Delta S + R * \ln \left[\frac{C}{4} \right]} - 273,15 + 16.6 * \log([Na^+]) \quad (3.5)$$

onde ΔH , ΔS , R e C são definidos como no parágrafo anterior.

Capítulo 4

Metodologia

Neste capítulo serão mostradas as etapas e uma forma de conduzir o processo de construção da aplicação que realizará o cálculo da temperatura de melting com as diferentes formulações, como visto na seção anterior.

4.1 Instrumental Utilizado

Após terminada a pesquisa aplicada para levantamento das diferentes formulações de temperatura de melting, uma nova parte do trabalho é iniciada: a construção de uma aplicação web. Java será a linguagem utilizada para desenvolver o projeto utilizando a IDE NETBeans com o sistema operacional windows XP. O programa será construído na forma de um *applet*, que é uma forma de software aplicativo executado no contexto de outro programa, neste caso, dentro de uma página web.

4.2 Cronograma

Tabela 4.1: Etapas do processo de construção da aplicação do trabalho

Mês	Atividade
agosto	estudo das linguagens e ferramentas que serão utilizadas
setembro	início da implementação prática do projeto
outubro	desenvolvimento da aplicação
novembro	finalização da aplicação e realização de testes
dezembro	verificação dos resultados encontrados

Capítulo 5

Resultados

Neste capítulo encontram-se os resultados de diferentes experimentos, todos com o objetivo de comparar os diferentes resultados e comportamentos das Tm's. As onze fórmulas são enumeradas de acordo com a tabela 5.1.

Tabela 5.1: Tabela de representação das Tm's para os resultados expostos neste capítulo

Representação	Fórmula
Tm1	Básica (3.1)
Tm2	Básica (3.2)
Tm3	Dependente do Sal (3.3)
Tm4	Termodinâmica (3.4) utilizando-se ΔH e ΔS propostos por Breslauer
Tm5	Termodinâmica (3.5) utilizando-se ΔH e ΔS propostos por Breslauer
Tm6	Termodinâmica (3.4) utilizando-se ΔH e ΔS propostos por Sugimoto
Tm7	Termodinâmica (3.5) utilizando-se ΔH e ΔS propostos por Sugimoto
Tm8	Termodinâmica (3.4) utilizando-se ΔH e ΔS propostos por SantaLucia
Tm9	Termodinâmica (3.5) utilizando-se ΔH e ΔS propostos por SantaLucia
Tm10	Termodinâmica (3.4) utilizando-se ΔH e ΔS propostos por Allawai;'SantaLucia
Tm11	Termodinâmica (3.5) utilizando-se ΔH e ΔS propostos por Allawai;'SantaLucia

O capítulo está dividido em duas seções. A seção 5.1 aborda os resultados em relação a variação do tamanho dos *primers* e na seção 5.2 são encontrados os resultados dos mesmo experimentos da seção 5.1 porém, agora, em relação a concentração de %GC nos *primers*. Além disso, cada seção está dividida em outras duas subseções: os resultados das médias das diferenças entre as fórmulas e a correlação entre cada par de fórmulas. Esses experimentos são descritos com mais detalhes mais a frente.

5.1 Experimentos com o tamanho do *primer*

Para os experimentos a seguir foram gerados mil *primers* para cada tamanho: 15, 20, 25, 30 e 35. Totalizando-se cinco mil *primers* gerados. Em seguida, para cada *primer* foram

calculadas as onze Tm's de acordo com a tabela 5.1. E, com esses valores foram feitas comparações entre as fórmulas, e os resultados são apresentadas na forma de tabelas e/ou representação gráfica.

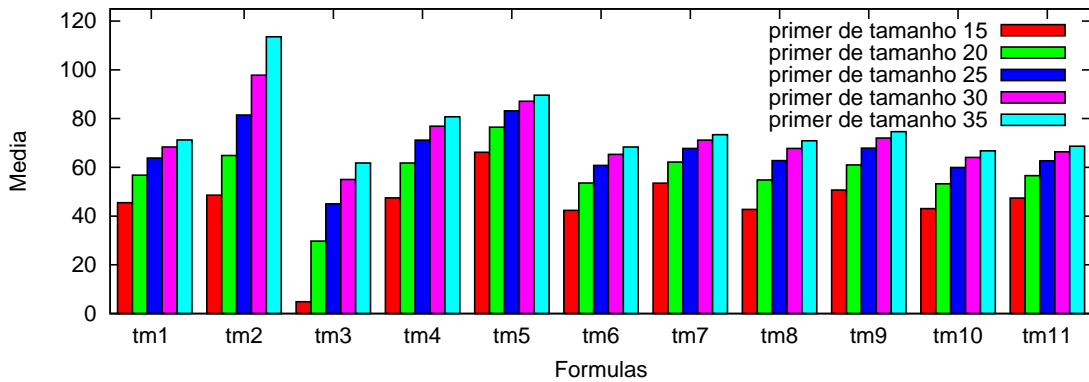


Figura 5.1: comparação das médias das Tm's em relação ao tamanho dos *primers*

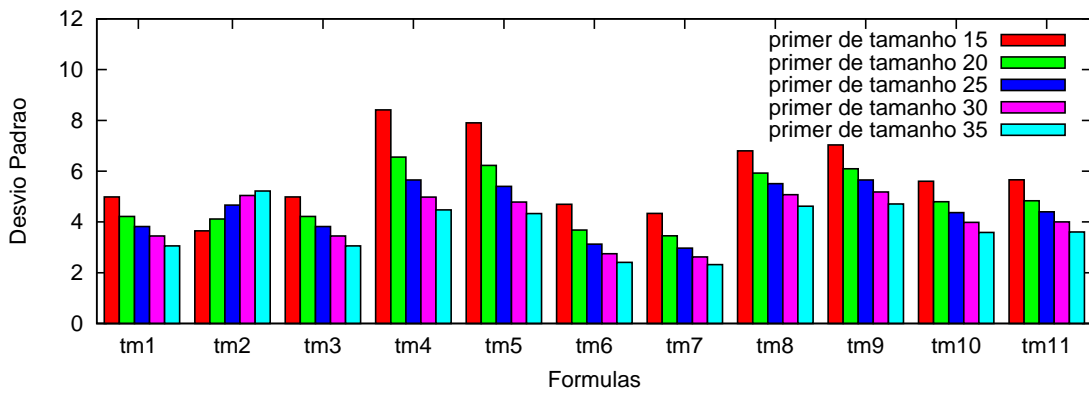


Figura 5.2: comparação das desvios padrões das Tm's em relação ao tamanho dos *primers*

Na figura 5.1, as médias das temperaturas calculadas pelas Tm's em relação ao tamanho dos *primers* mostram que as fórmulas apresentam o mesmo comportamento: maiores temperaturas para *primers* maiores. Podemos observar também que, para *primers* de tamanho 15, a tm3 é a fórmula que apresenta valores mais distantes de resultados das outras Tm's, e para *primers* de tamanho 35, a tm2 também não apresenta uma média próxima das outras fórmulas. Assim, para determinados tamanhos, esses valores podem não ser uma boa aproximação para a predição da temperatura real.

Já a figura 5.2 mostra os desvios padrão para os dados utilizados no cálculo das médias da figura 5.1. Assim, vemos que a tm2 apresenta um comportamento diferenciado das outras fórmulas, quanto maior o tamanho do *primer* maior o desvio padrão. Já as outras fórmulas apresentam um menor desvio padrão para *primers* maiores, ou seja, quando maior o *primer*, mais as temperaturas convergem para próximo da média.

5.1.1 Médias das diferenças entre as Tm's para *primers* de diferentes tamanho

A média da diferença entre duas Tm's é calculada de forma muito simples. Por exemplo, Seja t1 e t2 as temperaturas calculadas para um dos *primers* de tamanho 15, pela Tm1 e Tm2 respectivamente. A média da diferença entre a Tm1 e a Tm2 para *primers* de tamanho 15 nada mais é do que a média aritmética dos valores $|t1-t2|$ calculados para cada um dos *primers* de tamanho 15. Esse valor nos mostra quais fórmulas apresentam temperaturas mais próximas das outras.

Tabela 5.2: médias das diferenças entre as Tm's para *primers* de tamanho 15

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	3,11	40,66	3,99	20,66	3,28	7,96	3,03	5,20	2,55	2,05
Tm2	3,11	0,00	43,76	4,39	1,76	6,28	4,86	5,97	3,34	5,56	2,16
Tm3	40,66	43,76	0,00	42,64	61,33	37,49	48,62	37,90	45,83	38,21	42,55
Tm4	3,99	4,39	42,64	0,00	18,69	5,70	6,38	4,83	3,46	4,83	2,61
Tm5	20,66	1,76	61,33	18,69	0,00	23,84	12,71	23,43	15,49	23,12	18,78
Tm6	3,28	6,28	37,49	5,70	23,84	0,00	11,14	2,24	8,34	1,31	5,06
Tm7	7,96	4,86	48,62	6,38	12,71	11,14	0,00	10,73	3,85	10,42	6,08
Tm8	3,03	5,97	37,90	4,83	23,43	2,24	10,73	0,00	7,93	1,17	4,65
Tm9	5,20	3,34	45,83	3,46	15,49	8,34	3,85	7,93	0,00	7,63	3,29
Tm10	2,55	5,56	38,21	4,83	23,12	1,31	10,42	1,17	7,63	0,00	4,34
Tm11	2,05	2,16	42,55	2,61	18,78	5,06	6,08	4,65	3,29	4,34	0,00
Total	92,50	81,19	438,99	97,52	219,80	104,68	122,74	101,88	104,37	99,13	91,57

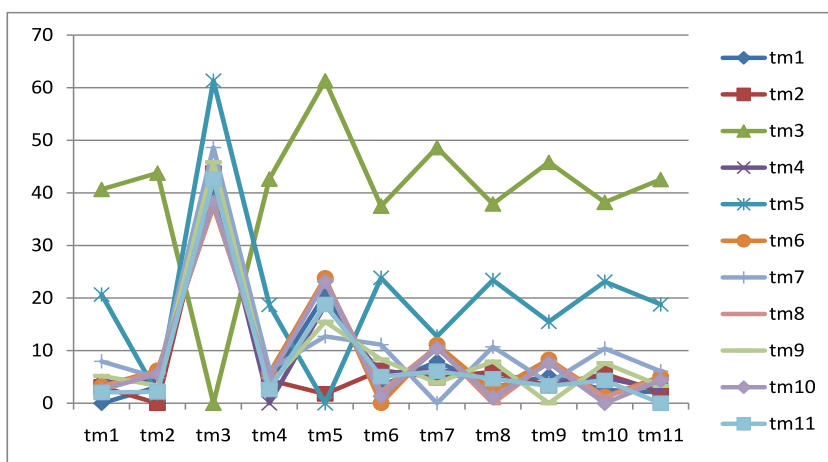


Figura 5.3: médias das diferenças entre as Tm's para *primers* de tamanho 15

Na tabela 5.2 podemos ver que a coluna da Tm3 apresenta grandes valores de diferença em relação as outras Tm's. A Tm5 também apresenta valores um pouco distantes das outras fórmulas, porém, mais próximos que a Tm3. Isso pode ser observado na figura 5.3.

Assim, a Tm3 e a Tm5 podem não apresentar um resultado apropriado para *primers* de tamanho 15. Vemos também, através dos totais das diferenças, que a Tm2 se aproxima mais das demais tm's podendo apresentar um bom resultado.

Tabela 5.3: médias das diferenças entre as Tm's para *primers* de tamanho 20

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	8,10	27,00	5,23	19,69	3,20	5,44	2,32	4,26	3,51	1,00
Tm2	8,10	0,00	35,10	3,54	11,59	11,26	2,77	10,06	3,98	11,60	8,26
Tm3	27,00	35,10	0,00	32,01	46,69	23,84	32,44	25,04	31,22	23,50	26,84
Tm4	5,23	3,54	32,01	0,00	14,68	8,20	3,01	6,97	1,77	8,52	5,21
Tm5	19,69	11,59	46,69	14,68	0,00	22,85	14,25	21,65	15,47	23,19	19,85
Tm6	3,20	11,26	23,84	8,20	22,85	0,00	8,60	2,51	7,39	1,46	3,02
Tm7	5,44	2,77	32,44	3,01	14,25	8,60	0,00	7,47	3,08	8,94	5,62
Tm8	2,32	10,06	25,04	6,97	21,65	2,51	7,47	0,00	6,19	1,67	1,96
Tm9	4,26	3,98	31,22	1,77	15,47	7,39	3,08	6,19	0,00	7,73	4,39
Tm10	3,51	11,60	23,50	8,52	23,19	1,46	8,94	1,67	7,73	0,00	3,34
Tm11	1,00	8,26	26,84	5,21	19,85	3,02	5,62	1,96	4,39	3,34	0,00
Total	79,74	106,26	303,68	89,14	209,91	92,31	91,61	85,85	85,46	93,45	79,50

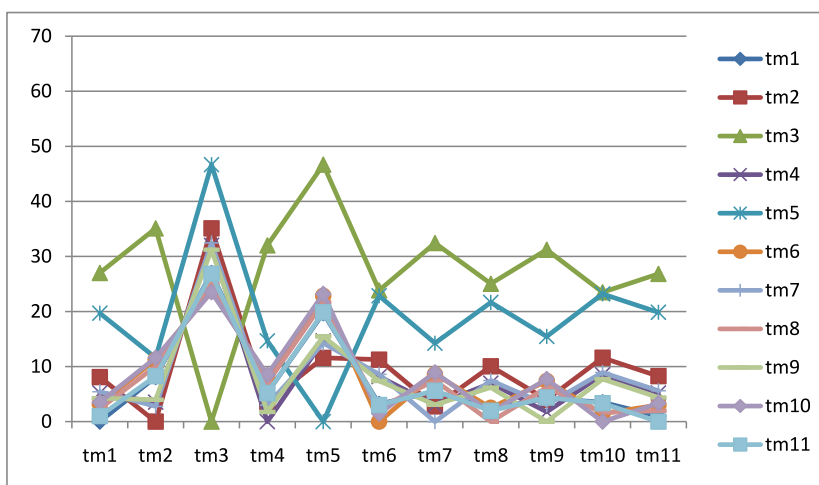


Figura 5.4: médias das diferenças entre as Tm's para *primers* de tamanho 20

Já na tabela 5.3 para *primers* de tamanho 20, os valores das diferenças entre a Tm3 e as outras fórmulas diminuí. Como podemos ver na figura 5.4, a Tm3 se aproxima mais das outras fórmulas mas ainda possui uma diferença considerável, enquanto que a Tm5 continua com um comportamento semelhante ao gráfico 5.3. É interessante observar que a Tm3 e a Tm5 apresentam os resultados mais distantes entre si, isso ocorre porque a Tm3 apresenta os menores valores para a temperatura de melting para *primers* pequenos enquanto que a Tm5 apresenta os maiores valores, isso pode ser observado na figura 5.1. Além disso, a tabela 5.3 mostra que a Tm1 e a Tm11 apresentam as menores diferenças, podendo apresentar bons resultados para a predição temperatura de melting.

Tabela 5.4: médias das diferenças entre as Tm's para *primers* de tamanho 25

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	17,67	18,79	7,33	19,33	3,10	3,95	1,77	4,06	3,86	1,29
Tm2	17,67	0,00	36,46	10,35	2,22	20,73	13,76	18,70	13,64	21,52	18,82
Tm3	18,79	36,46	0,00	26,11	38,12	15,73	22,70	17,75	22,82	14,93	17,64
Tm4	7,33	10,35	26,11	0,00	12,01	10,38	3,99	8,36	3,36	11,18	8,47
Tm5	19,33	2,22	38,12	12,01	0,00	22,38	15,42	20,36	15,29	23,18	20,47
Tm6	3,10	20,73	15,73	10,38	22,38	0,00	6,97	2,94	7,09	1,69	2,15
Tm7	3,95	13,76	22,70	3,99	15,42	6,97	0,00	5,29	2,91	7,76	5,09
Tm8	1,77	18,70	17,75	8,36	20,36	2,94	5,29	0,00	5,07	2,82	1,11
Tm9	4,06	13,64	22,82	3,36	15,29	7,09	2,91	5,07	0,00	7,89	5,18
Tm10	3,86	21,52	14,93	11,18	23,18	1,69	7,76	2,82	7,89	0,00	2,71
Tm11	1,29	18,82	17,64	8,47	20,47	2,15	5,09	1,11	5,18	2,71	0,00
Total	81,14	173,86	231,06	101,52	188,78	93,15	87,83	84,17	87,31	97,54	82,93

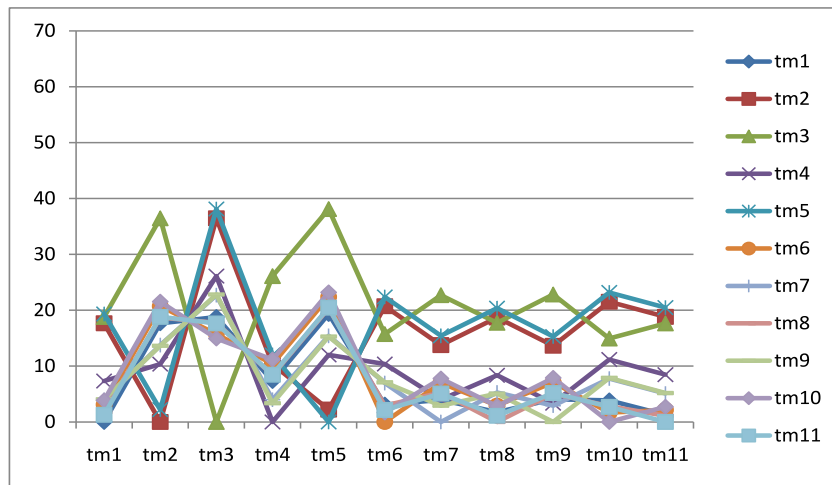


Figura 5.5: médias das diferenças entre as Tm's para *primers* de tamanho 25

Na tabela 5.4 para *primers* de tamanho 25, a diferença entre a Tm3 para as demais fórmulas diminuiu um pouco mais, enquanto que a Tm2 começou a apresentar maiores valores de diferença. Isso pode ser visto na figura 5.5, onde a representante da Tm3 está mais próxima e a Tm2 começa a se afastar das demais Tm's. A Tm1 e a Tm11 ainda apresentam as menores diferenças, representando resultados que podem ser mais confiáveis.

Tabela 5.5: médias das diferenças entre as Tm's para *primers* de tamanho 30

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	29,48	13,33	8,58	18,78	3,05	2,93	1,53	3,73	4,24	1,98
Tm2	29,48	0,00	42,81	20,90	10,70	32,50	26,64	30,07	25,78	33,72	31,45
Tm3	13,33	42,81	0,00	21,91	32,11	10,31	16,18	12,75	17,03	9,09	11,36
Tm4	8,58	20,90	21,91	0,00	10,20	11,60	5,82	9,17	4,89	12,82	10,55
Tm5	18,78	10,70	32,11	10,20	0,00	21,80	15,93	19,36	15,08	23,02	20,75
Tm6	3,05	32,50	10,31	11,60	21,80	0,00	5,87	3,12	6,74	1,83	1,70
Tm7	2,93	26,64	16,18	5,82	15,93	5,87	0,00	4,05	2,78	7,08	4,84
Tm8	1,53	30,07	12,75	9,17	19,36	3,12	4,05	0,00	4,28	3,65	1,55
Tm9	3,73	25,78	17,03	4,89	15,08	6,74	2,78	4,28	0,00	7,94	5,67
Tm10	4,24	33,72	9,09	12,82	23,02	1,83	7,08	3,65	7,94	0,00	2,27
Tm11	1,98	31,45	11,36	10,55	20,75	1,70	4,84	1,55	5,67	2,27	0,00
Total	87,63	284,06	186,89	116,43	187,72	98,51	92,10	89,53	93,92	105,66	92,12

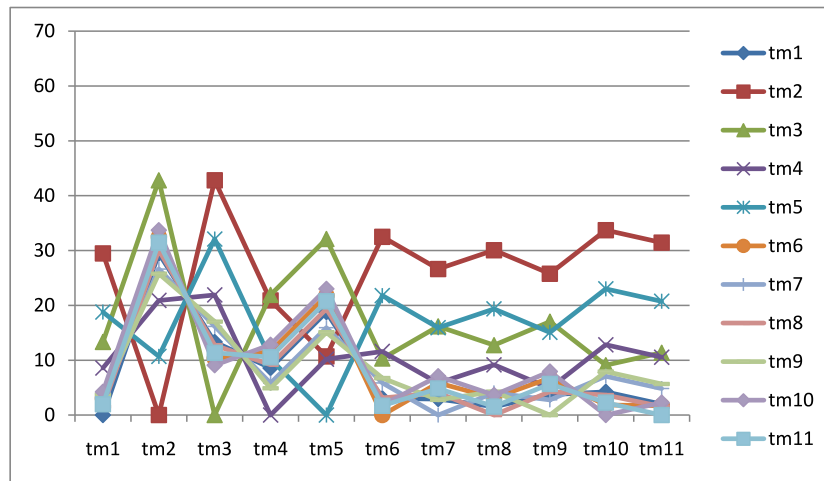


Figura 5.6: médias das diferenças entre as Tm's para *primers* de tamanho 30

Na tabela 5.5 para *primers* de tamanho 30, podemos observar que a Tm2 continua se afastando das demais fórmulas enquanto que a Tm3 apresenta valores mais coerentes com as outras tm's. Isso indica que para *primers* maiores a Tm2 pode não apresentar valores muito confiáveis.

Tabela 5.6: médias das diferenças entre as Tm's para *primers* de tamanho 35

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	42,38	9,43	9,53	18,39	2,89	2,36	1,42	3,43	4,49	2,55
Tm2	42,38	0,00	51,80	32,85	23,98	45,25	40,18	42,67	38,98	46,86	44,92
Tm3	9,43	51,80	0,00	18,95	27,82	6,56	11,62	9,13	12,83	4,94	6,88
Tm4	9,53	32,85	18,95	0,00	8,87	12,40	7,35	9,83	6,13	14,02	12,07
Tm5	18,39	23,98	27,82	8,87	0,00	21,26	16,20	18,69	14,99	22,88	20,94
Tm6	2,89	45,25	6,56	12,40	21,26	0,00	5,06	3,17	6,28	2,03	1,51
Tm7	2,36	40,18	11,62	7,35	16,20	5,06	0,00	3,34	2,75	6,68	4,76
Tm8	1,42	42,67	9,13	9,83	18,69	3,17	3,34	0,00	3,70	4,19	2,27
Tm9	3,43	38,98	12,83	6,13	14,99	6,28	2,75	3,70	0,00	7,89	5,95
Tm10	4,49	46,86	4,94	14,02	22,88	2,03	6,68	4,19	7,89	0,00	1,94
Tm11	2,55	44,92	6,88	12,07	20,94	1,51	4,76	2,27	5,95	1,94	0,00
Total	96,87	409,87	159,95	131,99	194,04	106,40	100,31	98,40	102,92	115,91	103,79

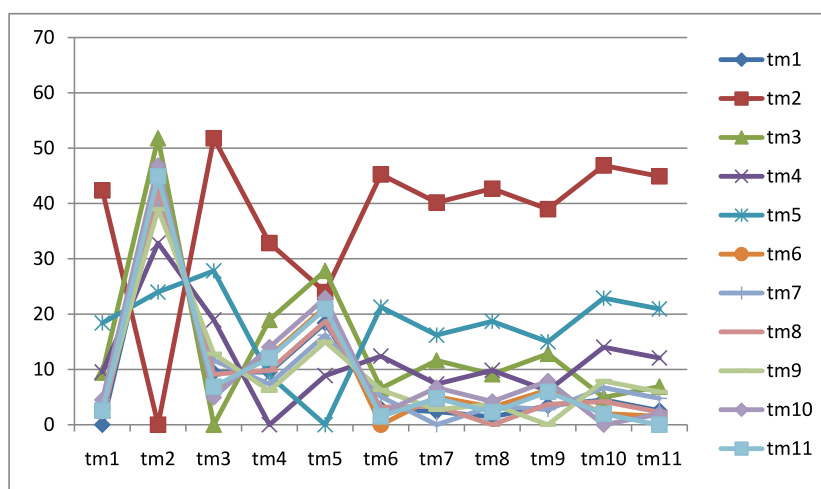


Figura 5.7: médias das diferenças entre as Tm's para *primers* de tamanho 35

A tabela 5.6 de médias das diferenças para *primers* de tamanho 35, mostra que a Tm2 se afasta completamente das outras Tm's e que seu resultado pode apresentar valores incorretos para um experimento. Já a Tm3 se aproxima bastante das outras fórmulas, podendo ser um bom referencial para esse tamanho de *primer*. Podemos ver em todos os gráficos desta seção que a Tm5 mantém um mesmo comportamento, ela não se aproxima completamente das outras fórmulas, mas não se afasta o suficiente para concluirmos se ela faz um boa predição.

5.1.2 Correlação entre Tm's para diferentes tamanhos de primers

Correlação, também chamada de coeficiente de correlação, indica a força e a direção do relacionamento linear entre duas variáveis aleatórias. Vários coeficientes são utilizados para situações diferentes. O mais conhecido, e utilizado aqui, é o coeficiente de correlação de Pearson, o qual é obtido dividindo a covariância de duas variáveis pelo produto de seus desvios padrão. Apesar do nome, ela foi apresentada inicialmente por Francis Galton.

Tabela 5.7: correlação entre as Tm's para primers de tamanho 15

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,91	0,91	0,92	0,91	0,97	0,97	0,96	0,96
Tm2	1,00	1,00	1,00	0,91	0,91	0,92	0,91	0,97	0,97	0,96	0,96
Tm3	1,00	1,00	1,00	0,91	0,91	0,92	0,91	0,97	0,97	0,96	0,96
Tm4	0,91	0,91	0,91	1,00	1,00	0,93	0,91	0,97	0,96	0,97	0,97
Tm5	0,91	0,91	0,91	1,00	1,00	0,94	0,92	0,97	0,96	0,97	0,97
Tm6	0,92	0,92	0,92	0,93	0,94	1,00	1,00	0,95	0,93	0,97	0,97
Tm7	0,91	0,91	0,91	0,91	0,92	1,00	1,00	0,92	0,91	0,95	0,95
Tm8	0,97	0,97	0,97	0,97	0,97	0,95	0,92	1,00	1,00	0,99	0,99
Tm9	0,97	0,97	0,97	0,96	0,96	0,93	0,91	1,00	1,00	0,99	0,99
Tm10	0,96	0,96	0,96	0,97	0,97	0,97	0,95	0,99	0,99	1,00	1,00
Tm11	0,96	0,96	0,96	0,97	0,97	0,97	0,95	0,99	0,99	1,00	1,00

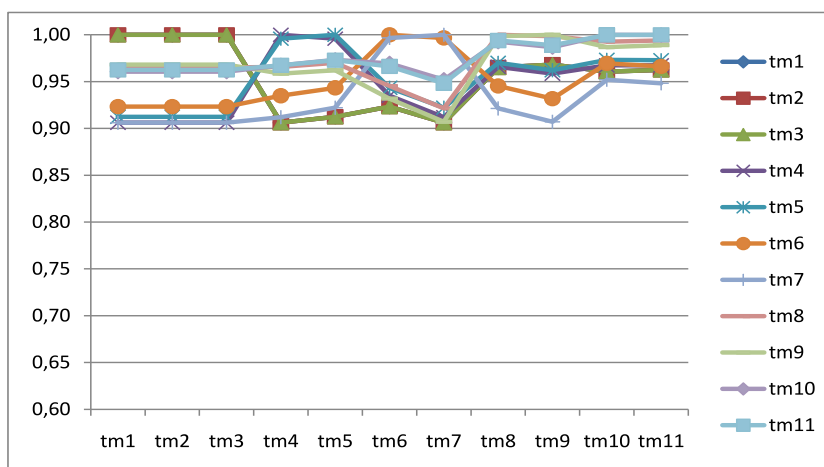


Figura 5.8: correlação entre Tm's para primers de tamanho 15

A tabela 5.7, apresenta a correlação entre as Tm's para primers de tamanho 15. Podemos observar que os valores são elevados, ou seja, as Tm's se comportam de maneira semelhante, isso pode ser visto na figura 5.8. Apesar de vermos na subseção 5.1.1 que algumas Tm's apresentam grandes diferenças das outras, por exemplo na figura 5.3 onde vemos a Tm3 destacada das demais, ela mostra um comportamento linear muito próximo

das demais. Ou seja, quando todas as Tm's apresentam um valor maior, ela provavelmente mostrará um valor maior e proporcional.

Tabela 5.8: correlação entre as Tm's para *primers* de tamanho 20

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,92	0,92	0,91	0,89	0,98	0,97	0,97	0,97
Tm2	1,00	1,00	1,00	0,92	0,92	0,91	0,89	0,98	0,97	0,97	0,97
Tm3	1,00	1,00	1,00	0,92	0,92	0,91	0,89	0,98	0,97	0,97	0,97
Tm4	0,92	0,92	0,92	1,00	1,00	0,93	0,90	0,96	0,95	0,97	0,97
Tm5	0,92	0,92	0,92	1,00	1,00	0,93	0,91	0,96	0,95	0,97	0,97
Tm6	0,91	0,91	0,91	0,93	0,93	1,00	1,00	0,90	0,89	0,95	0,94
Tm7	0,89	0,89	0,89	0,90	0,91	1,00	1,00	0,88	0,86	0,93	0,92
Tm8	0,98	0,98	0,98	0,96	0,96	0,90	0,88	1,00	1,00	0,99	0,99
Tm9	0,97	0,97	0,97	0,95	0,95	0,89	0,86	1,00	1,00	0,98	0,98
Tm10	0,97	0,97	0,97	0,97	0,97	0,95	0,93	0,99	0,98	1,00	1,00
Tm11	0,97	0,97	0,97	0,97	0,97	0,94	0,92	0,99	0,98	1,00	1,00

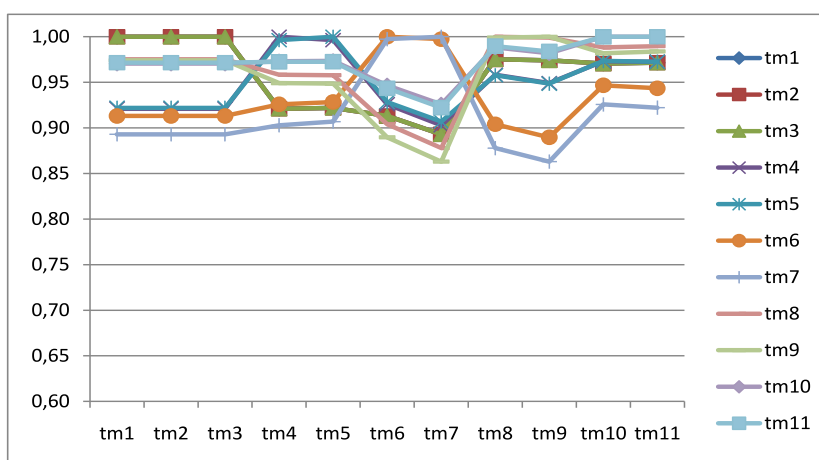


Figura 5.9: correlação entre Tm's para *primers* de tamanho 20

A tabela 5.8 nos mostra a correlação entre as Tm's para *primers* de tamanho 20. Podemos ver que os valores ainda são elevados, porém na figura 5.9 a Tm6 e a Tm7 começam a se afastar ligeiramente das demais, ou seja, para *primers* de tamanho 20 seu comportamento começa a se diferir linearmente das demais fórmulas.

Tabela 5.9: correlação entre as Tm's para *primers* de tamanho 25

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,93	0,92	0,90	0,87	0,98	0,98	0,98	0,98
Tm2	1,00	1,00	1,00	0,93	0,92	0,90	0,87	0,98	0,98	0,98	0,98
Tm3	1,00	1,00	1,00	0,93	0,92	0,90	0,87	0,98	0,98	0,98	0,98
Tm4	0,93	0,93	0,93	1,00	1,00	0,91	0,89	0,95	0,94	0,97	0,97
Tm5	0,92	0,92	0,92	1,00	1,00	0,92	0,89	0,94	0,93	0,97	0,97
Tm6	0,90	0,90	0,90	0,91	0,92	1,00	1,00	0,86	0,85	0,93	0,92
Tm7	0,87	0,87	0,87	0,89	0,89	1,00	1,00	0,84	0,82	0,90	0,90
Tm8	0,98	0,98	0,98	0,95	0,94	0,86	0,84	1,00	1,00	0,98	0,99
Tm9	0,98	0,98	0,98	0,94	0,93	0,85	0,82	1,00	1,00	0,98	0,98
Tm10	0,98	0,98	0,98	0,97	0,97	0,93	0,90	0,98	0,98	1,00	1,00
Tm11	0,98	0,98	0,98	0,97	0,97	0,92	0,90	0,99	0,98	1,00	1,00

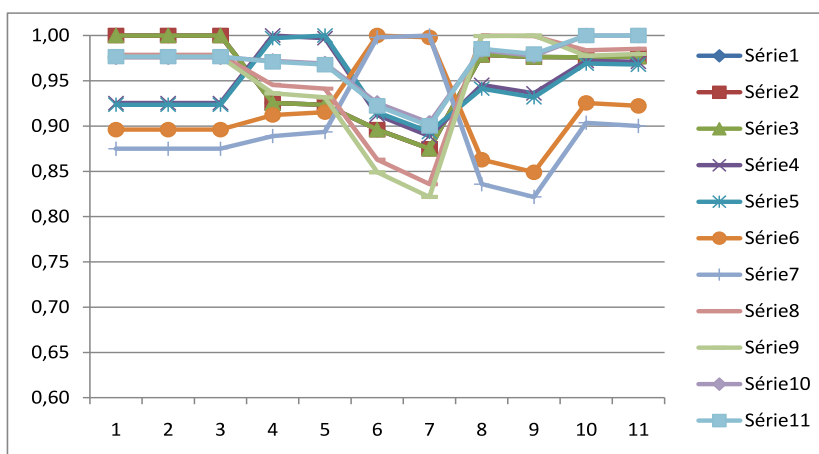


Figura 5.10: correlação entre Tm's para *primers* de tamanho 25

A tabela 5.9 de correlação entre as Tm's para *primers* de tamanho 25, nos mostra uma tendência em que a Tm6 e a Tm7 apresentem valores de correlação menores para *primers* maiores. Essa tendência pode ser vista na figura 5.10, onde as representantes da Tm6 e da Tm7 começam a se afastar das demais principalmente da Tm8 e da Tm9. Ou seja, seu comportamento pode começar a variar em relação as outras Tm's para *primers* grandes.

Tabela 5.10: correlação entre as Tm's para *primers* de tamanho 30

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,93	0,93	0,89	0,87	0,98	0,98	0,98	0,98
Tm2	1,00	1,00	1,00	0,93	0,93	0,89	0,87	0,98	0,98	0,98	0,98
Tm3	1,00	1,00	1,00	0,93	0,93	0,89	0,87	0,98	0,98	0,98	0,98
Tm4	0,93	0,93	0,93	1,00	1,00	0,90	0,88	0,94	0,94	0,97	0,97
Tm5	0,93	0,93	0,93	1,00	1,00	0,90	0,88	0,94	0,93	0,97	0,97
Tm6	0,89	0,89	0,89	0,90	0,90	1,00	1,00	0,84	0,83	0,91	0,91
Tm7	0,87	0,87	0,87	0,88	0,88	1,00	1,00	0,82	0,80	0,89	0,88
Tm8	0,98	0,98	0,98	0,94	0,94	0,84	0,82	1,00	1,00	0,98	0,98
Tm9	0,98	0,98	0,98	0,94	0,93	0,83	0,80	1,00	1,00	0,98	0,98
Tm10	0,98	0,98	0,98	0,97	0,97	0,91	0,89	0,98	0,98	1,00	1,00
Tm11	0,98	0,98	0,98	0,97	0,97	0,91	0,88	0,98	0,98	1,00	1,00

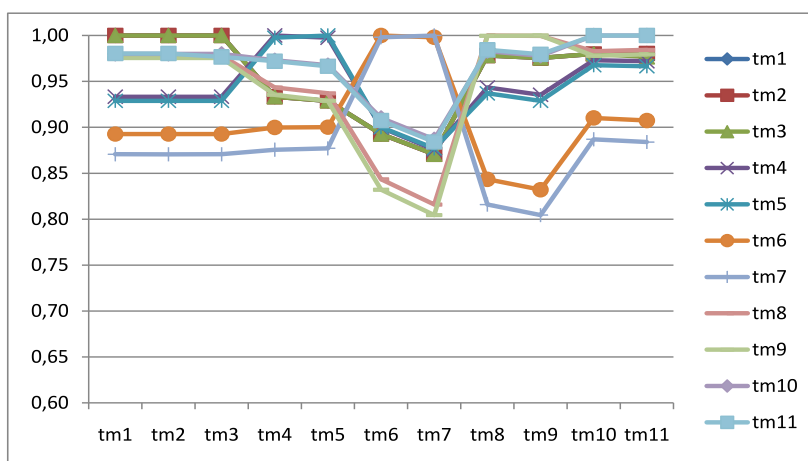


Figura 5.11: correlação entre Tm's para *primers* de tamanho 30

A tabela 5.10 nos dá a correlação entre as Tm's para *primers* de tamanho 30. Nela podemos observar que os valores das correlações da Tm6 e da Tm7 diminuem ainda mais. Na figura 5.11 vemos que suas representações se afastam um pouco mais em relação a figura 5.10. Isto não indica que essas fórmulas apresentem um valor inválido, apenas que elas variam de forma diferente das demais.

Tabela 5.11: correlação entre as Tm's para *primers* de tamanho 35

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,93	0,92	0,87	0,85	0,98	0,97	0,98	0,98
Tm2	1,00	1,00	1,00	0,93	0,92	0,87	0,85	0,98	0,97	0,98	0,98
Tm3	1,00	1,00	1,00	0,93	0,92	0,87	0,85	0,98	0,97	0,98	0,98
Tm4	0,93	0,93	0,93	1,00	1,00	0,88	0,85	0,93	0,92	0,97	0,97
Tm5	0,92	0,92	0,92	1,00	1,00	0,88	0,85	0,92	0,92	0,96	0,96
Tm6	0,87	0,87	0,87	0,88	0,88	1,00	1,00	0,80	0,79	0,88	0,88
Tm7	0,85	0,85	0,85	0,85	0,85	1,00	1,00	0,78	0,76	0,86	0,86
Tm8	0,98	0,98	0,98	0,93	0,92	0,80	0,78	1,00	1,00	0,98	0,98
Tm9	0,97	0,97	0,97	0,92	0,92	0,79	0,76	1,00	1,00	0,97	0,98
Tm10	0,98	0,98	0,98	0,97	0,96	0,88	0,86	0,98	0,97	1,00	1,00
Tm11	0,98	0,98	0,98	0,97	0,96	0,88	0,86	0,98	0,98	1,00	1,00

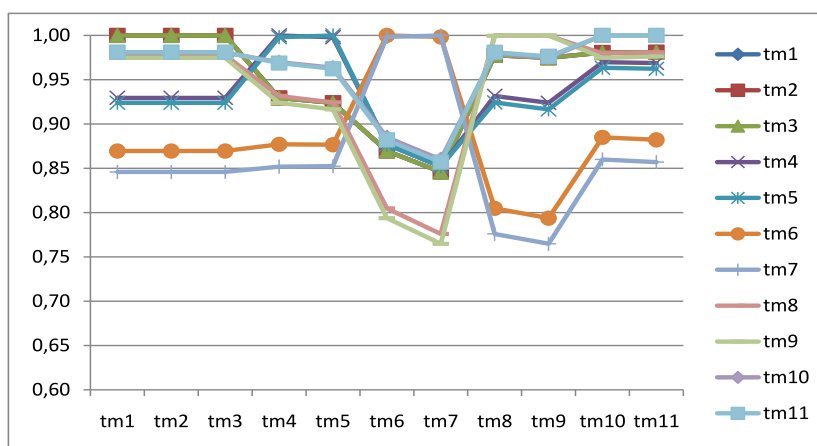


Figura 5.12: correlação entre Tm's para *primers* de tamanho 35

A tabela 5.11 nos dá a correlação entre as Tm's para *primers* de tamanho 35. Ela nos mostra que os valores das colunas da Tm6 e da Tm7 diminuem um pouco mais. Isso é observado no gráfico 5.12. Perceba que apesar de estes valores diminuírem, eles ainda são elevados (acima de 76%), ou seja, todas as fórmulas apresentam um comportamento linear semelhante, a maioria atingindo valores de correlação acima de 90%. Assim, quando uma das tm's apresenta um valor maior para algum *primer*, então provavelmente as outras também mostraram valores maiores e proporcionais para este *primer*.

5.2 Experimentos com a concentração de GC nos *primers*

Para os testes a seguir foram gerados mil *primers* de tamanho fixo 20 para cada uma das seguintes concentrações de GC: 0 a 20%, 20 a 40%, 40 a 60%, 60 a 80% e 80 a 100%.

Totalizando cinco mil *primers* gerados. Em seguida para cada *primer*, foram calculadas as onze T_m 's de acordo com a tabela 5.1. Com esses valores foram feitas as comparações entre as fórmulas.

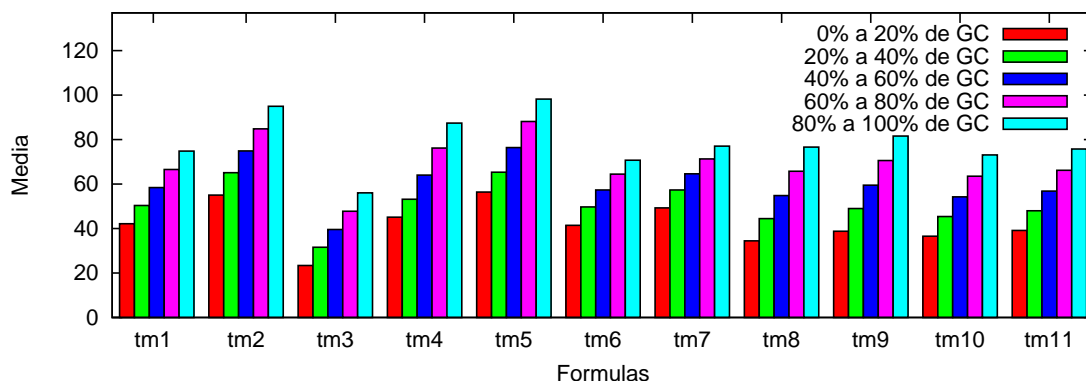


Figura 5.13: comparação das médias das T_m 's em relação a concentração de GC nos *primers*

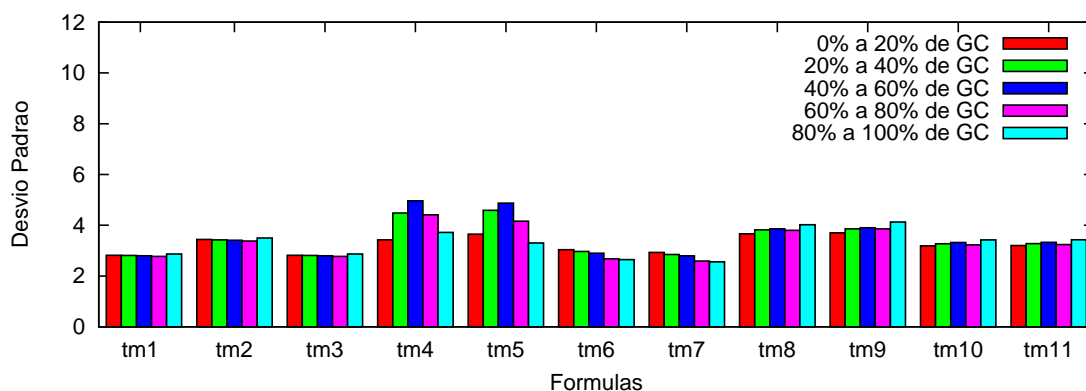


Figura 5.14: comparação das desvios padrões das T_m 's em relação a concentração de GC dos *primers*

A figura 5.13 nos dá a comparação entre as médias das temperaturas calculadas pelas T_m 's em relação as diferentes concentrações de GC nos *primers*. Podemos ver que o comportamento de todas as tm 's é semelhante. Do mesmo modo que o experimento em relação ao tamanho do *primer* (figura 5.1), quanto maior a concentração de GC maior a temperatura calculada pelas T_m 's. Vale notar que a $Tm3$ apresenta valores em média menores que as demais tm 's.

Já a figura 5.14 nos mostra os desvios padrão para os dados utilizados no cálculo das médias da figura 5.13. Assim, as únicas Tm 's que apresentam desvios padrão significativamente diferentes são a $Tm4$ e a $Tm5$. Para concentrações de GC intermediárias (40 a 60%) essas Tm 's variam mais em relação a média. Já as outras Tm 's possuem uma distribuição mais uniforme, mesmo que se varie a concentração de GC no *primer*.

5.2.1 Médias das diferenças entre as Tm's para *primers* com diferentes concentrações de GC

A seguir são apresentadas as médias das diferenças entre as Tm's na forma de tabelas e gráficos, assim como o experimento da subseção anterior que levava em consideração o tamanho do *primer*. Porém, agora, utilizando-se diferentes concentrações de GC nos *primers*.

Tabela 5.12: médias das diferenças entre as Tm's para *primers* com 0 a 20% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	12,91	18,79	3,38	14,22	0,95	7,16	7,70	3,38	5,58	3,04
Tm2	12,91	0,00	31,70	9,92	2,42	13,63	5,75	20,61	16,27	18,49	15,95
Tm3	18,79	31,70	0,00	21,78	33,01	18,08	25,95	11,09	15,44	13,21	15,76
Tm4	3,38	9,92	21,78	0,00	11,23	3,78	4,21	10,69	6,35	8,57	6,03
Tm5	14,22	2,42	33,01	11,23	0,00	14,93	7,06	21,92	17,57	19,80	17,25
Tm6	0,95	13,63	18,08	3,78	14,93	0,00	7,88	6,99	2,65	4,87	2,32
Tm7	7,16	5,75	25,95	4,21	7,06	7,88	0,00	14,86	10,52	12,74	10,20
Tm8	7,70	20,61	11,09	10,69	21,92	6,99	14,86	0,00	4,35	2,12	4,67
Tm9	3,38	16,27	15,44	6,35	17,57	2,65	10,52	4,35	0,00	2,23	0,55
Tm10	5,58	18,49	13,21	8,57	19,80	4,87	12,74	2,12	2,23	0,00	2,55
Tm11	3,04	15,95	15,76	6,03	17,25	2,32	10,20	4,67	0,55	2,55	0,00
Total	77,11	147,65	204,81	85,94	159,41	76,06	106,33	105,00	79,30	90,16	78,31

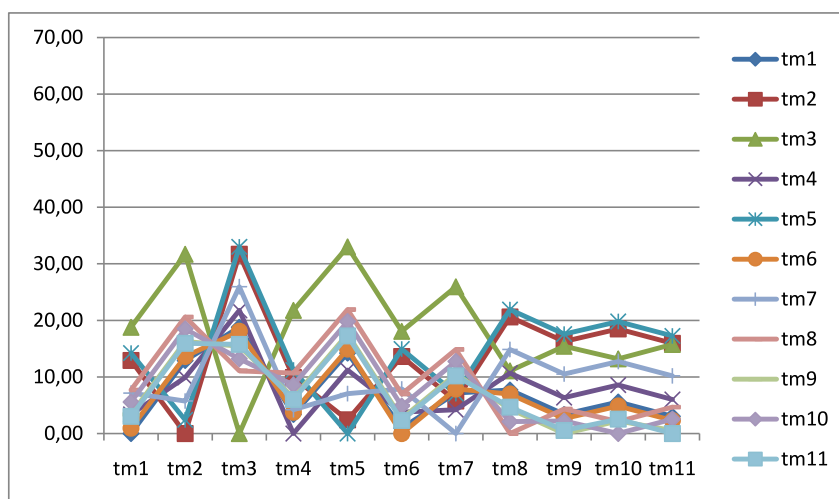


Figura 5.15: médias das diferenças entre as Tm's para *primers* com 0 a 20% de GC

Na tabela 5.12 das médias das diferenças para *primers* com 0 a 20% de GC, podemos ver que a Tm3 possui as maiores médias das diferenças seguida da Tm2 e da Tm5. O restante possui valores aproximados umas das outras. Isso pode ser visualizado na figura 5.15 onde a representante da Tm2, Tm3, e Tm5 se destacam das outras.

Tabela 5.13: médias das diferenças entre as Tm's para *primers* com 20 a 40% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	14,72	18,79	3,38	14,91	0,99	6,95	5,95	1,67	5,02	2,44
Tm2	14,72	0,00	33,51	11,94	2,29	15,37	7,77	20,67	16,11	19,73	17,15
Tm3	18,79	33,51	0,00	21,57	33,70	18,14	25,74	12,84	17,39	13,78	16,36
Tm4	3,38	11,94	21,57	0,00	12,13	3,63	4,25	8,73	4,19	7,79	5,21
Tm5	14,91	2,29	33,70	12,13	0,00	15,56	7,96	20,86	16,30	19,92	17,33
Tm6	0,99	15,37	18,14	3,63	15,56	0,00	7,60	5,30	1,23	4,36	1,79
Tm7	6,95	7,77	25,74	4,25	7,96	7,60	0,00	12,90	8,35	11,97	9,38
Tm8	5,95	20,67	12,84	8,73	20,86	5,30	12,90	0,00	4,56	0,99	3,52
Tm9	1,67	16,11	17,39	4,19	16,30	1,23	8,35	4,56	0,00	3,62	1,06
Tm10	5,02	19,73	13,78	7,79	19,92	4,36	11,97	0,99	3,62	0,00	2,59
Tm11	2,44	17,15	16,36	5,21	17,33	1,79	9,38	3,52	1,06	2,59	0,00
Total	74,81	159,27	211,81	82,83	160,95	73,97	102,87	96,32	74,48	89,76	76,84

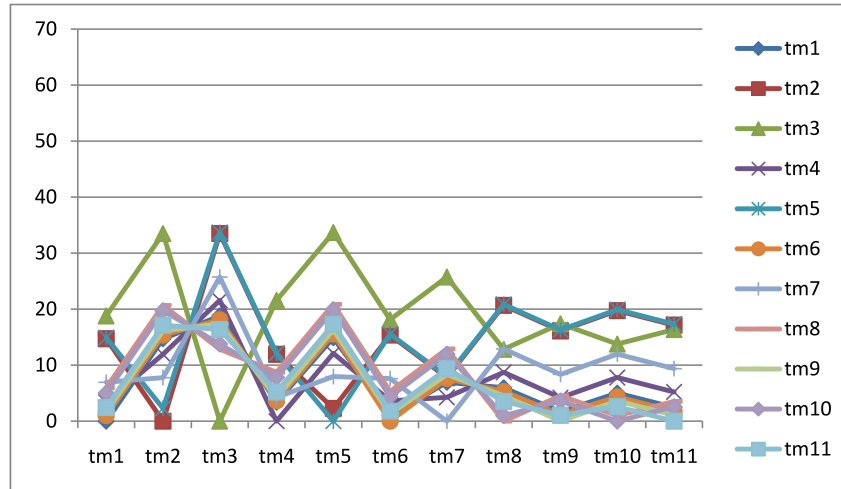


Figura 5.16: médias das diferenças entre as Tm's para *primers* com 20 a 40% de GC

A tabela 5.13 nos dá as médias das diferenças entre as Tm's para *primers* com 20 a 40% de GC em sua composição. Ela não apresenta grandes diferenças em relação a tabela anterior. Podemos observar que ambas as figuras 5.15 e 5.16 apresentam dados semelhantes. E, poderemos ver que as próximas tabelas e figuras apresentam resultados próximos destes. Isso indica que as diferentes concentrações de GC não caracterizam em uma mudança significativa nas médias das diferenças entre as Tm's. Ou seja, para um aumento na concentração de GC, todas as Tm's aumentam proporcionalmente.

Tabela 5.14: médias das diferenças entre as Tm's para *primers* com 40 a 60% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	16,48	18,79	5,73	18,01	1,34	6,21	3,60	1,50	4,17	1,63
Tm2	16,48	0,00	35,27	10,85	2,50	17,51	10,27	20,07	15,36	20,65	18,03
Tm3	18,79	35,27	0,00	24,42	36,80	17,75	25,00	15,19	19,91	14,62	17,24
Tm4	5,73	10,85	24,42	0,00	12,38	6,70	2,37	9,23	4,55	9,80	7,19
Tm5	18,01	2,50	36,80	12,38	0,00	19,05	11,80	21,61	16,89	22,18	19,56
Tm6	1,34	17,51	17,75	6,70	19,05	0,00	7,25	2,72	2,32	3,14	1,02
Tm7	6,21	10,27	25,00	2,37	11,80	7,25	0,00	9,81	5,10	10,38	7,76
Tm8	3,60	20,07	15,19	9,23	21,61	2,72	9,81	0,00	4,72	0,78	2,05
Tm9	1,50	15,36	19,91	4,55	16,89	2,32	5,10	4,72	0,00	5,29	2,67
Tm10	4,17	20,65	14,62	9,80	22,18	3,14	10,38	0,78	5,29	0,00	2,62
Tm11	1,63	18,03	17,24	7,19	19,56	1,02	7,76	2,05	2,67	2,62	0,00
Total	77,46	166,98	225,00	93,21	180,80	78,79	95,96	89,78	78,31	93,63	79,77

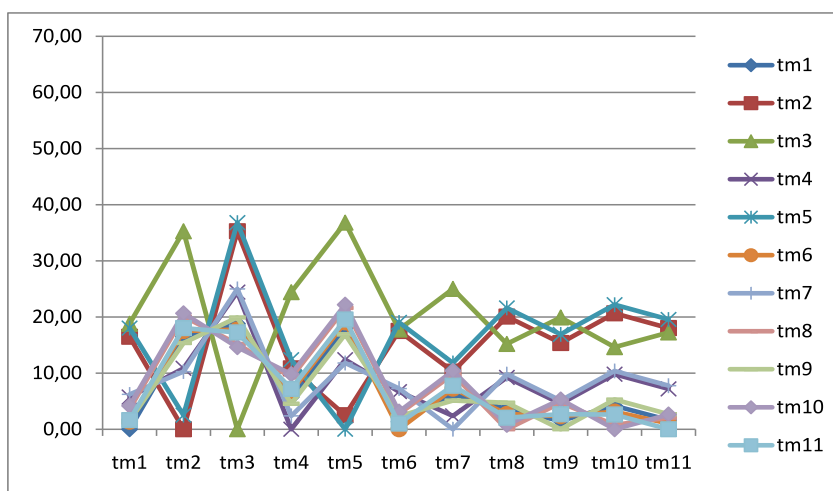


Figura 5.17: médias das diferenças entre as Tm's para *primers* com 40 a 60% de GC

A tabela 5.13 nos dá as médias das diferenças entre as Tm's para *primers* com 40 a 60% de GC. Ela mostra um pequeno aumento nas diferenças da Tm2, Tm3 e Tm5. Porém o resultado encontrado é semelhante ao das tabelas 5.11 e 5.12. A figura 5.17 mostra-se muito próxima das figuras 5.15 e 5.16.

Tabela 5.15: médias das diferenças entre as Tm's para *primers* com 60 a 80% GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	18,27	18,79	9,65	21,55	2,22	4,73	1,33	4,00	3,03	0,81
Tm2	18,27	0,00	37,06	8,63	3,38	20,38	13,55	19,12	14,27	21,30	18,64
Tm3	18,79	37,06	0,00	28,44	40,34	16,69	23,52	17,94	22,79	15,76	18,42
Tm4	9,65	8,63	28,44	0,00	11,90	11,75	5,00	10,49	5,65	12,68	10,02
Tm5	21,55	3,38	40,34	11,90	0,00	23,65	16,82	22,39	17,55	24,57	21,91
Tm6	2,22	20,38	16,69	11,75	23,65	0,00	6,83	2,19	6,11	1,50	1,93
Tm7	4,73	13,55	23,52	5,00	16,82	6,83	0,00	5,63	2,39	7,75	5,10
Tm8	1,33	19,12	17,94	10,49	22,39	2,19	5,63	0,00	4,85	2,19	0,95
Tm9	4,00	14,27	22,79	5,65	17,55	6,11	2,39	4,85	0,00	7,03	4,37
Tm10	3,03	21,30	15,76	12,68	24,57	1,50	7,75	2,19	7,03	0,00	2,66
Tm11	0,81	18,64	18,42	10,02	21,91	1,93	5,10	0,95	4,37	2,66	0,00
Total	84,38	174,61	239,75	114,20	204,05	93,24	91,31	87,07	89,00	98,47	84,81

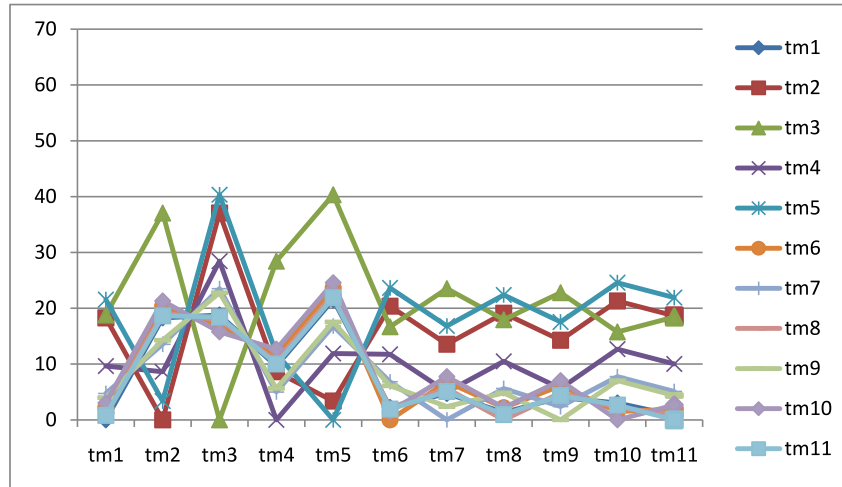


Figura 5.18: médias das diferenças entre as Tm's para *primers* com 60 a 80% de GC

Na tabela 5.15 temos as médias das diferenças entre as Tm's para *primers* com 60 a 80% de GC. Podemos ver que as médias das diferenças aumentaram de maneira praticamente uniforme para todas as Tm's para concentrações de 60 a 80% de GC. Porém, esse aumento é pouco significativo. Assim, o padrão das tabelas anteriores é mantido. Isso é observado também através da figura 5.18.

Tabela 5.16: médias das diferenças entre as Tm's para *primers* com 80 a 100% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	0,00	20,09	18,79	12,54	23,37	4,16	2,51	1,95	6,69	1,79	1,01
Tm2	20,09	0,00	38,88	7,55	3,28	24,24	17,89	18,31	13,40	21,88	19,19
Tm3	18,79	38,88	0,00	31,33	42,16	14,64	21,00	20,58	25,48	17,00	19,69
Tm4	12,54	7,55	31,33	0,00	10,83	16,69	10,34	10,75	5,88	14,33	11,64
Tm5	23,37	3,28	42,16	8,87	0,00	27,52	21,16	21,58	16,68	25,16	22,46
Tm6	4,16	24,24	14,64	12,40	27,52	0,00	6,36	5,99	10,84	2,58	5,06
Tm7	2,51	17,89	21,00	7,35	21,16	6,36	0,00	2,94	4,81	4,06	2,17
Tm8	1,95	18,31	20,58	9,83	21,58	5,99	2,94	0,00	4,91	3,58	1,32
Tm9	6,69	13,40	25,48	6,13	16,68	10,84	4,81	4,91	0,00	8,48	5,79
Tm10	1,79	21,88	17,00	14,02	25,16	2,58	4,06	3,58	8,48	0,00	2,70
Tm11	1,01	19,19	19,69	12,07	22,46	5,06	2,17	1,32	5,79	2,70	0,00
Total	92,91	184,71	249,55	122,08	214,20	118,08	93,22	91,90	102,96	101,56	91,03

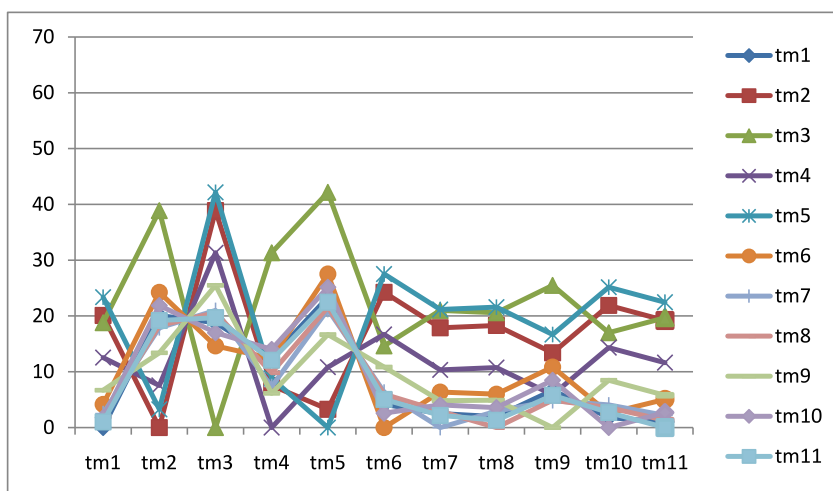


Figura 5.19: médias das diferenças entre as Tm's para *primers* com 80 a 100% de GC

A tabela 5.16 nos dá as médias das diferenças entre as Tm's para *primers* com 80 a 100% de GC. Podemos ver que, da mesma maneira que a tabela 5.15, com o aumento da concentração de GC, as médias das diferenças aumentam de maneira proporcional para todas as Tm's. Assim, a diferença desproporcional da Tm2, Tm3 e Tm5 em relação às outras Tm's está mais ligada ao tamanho do *primer*, no caso 20, do que com a concentração de GC. Olhando para a figura 5.19 e para as outras figuras desta subseção, veremos que, para diferentes concentrações as fórmulas possuem diferenças semelhantes.

5.2.2 Correlação entre as Tm's para diferentes concentrações de GC

A seguir são apresentados os índices de correlação entre as Tm's para diferentes concentrações de GC na forma de tabelas e gráficos. O cálculo é feito de maneira semelhante à subseção 5.1.2.

Tabela 5.17: correlação entre as Tm's para *primers* de 0 a 20% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,64	0,71	0,95	0,96	0,90	0,91	0,94	0,94
Tm2	1,00	1,00	1,00	0,64	0,71	0,95	0,96	0,90	0,91	0,94	0,94
Tm3	1,00	1,00	1,00	0,64	0,71	0,95	0,96	0,90	0,91	0,94	0,94
Tm4	0,64	0,64	0,64	1,00	0,99	0,82	0,79	0,89	0,88	0,85	0,84
Tm5	0,71	0,71	0,71	0,99	1,00	0,86	0,84	0,92	0,92	0,89	0,89
Tm6	0,95	0,95	0,95	0,82	0,86	1,00	1,00	0,98	0,98	0,99	0,99
Tm7	0,96	0,96	0,96	0,79	0,84	1,00	1,00	0,97	0,97	0,99	0,99
Tm8	0,90	0,90	0,90	0,89	0,92	0,98	0,97	1,00	1,00	0,99	0,99
Tm9	0,91	0,91	0,91	0,88	0,92	0,98	0,97	1,00	1,00	0,99	0,99
Tm10	0,94	0,94	0,94	0,85	0,89	0,99	0,99	0,99	0,99	1,00	1,00
Tm11	0,94	0,94	0,94	0,84	0,89	0,99	0,99	0,99	0,99	1,00	1,00

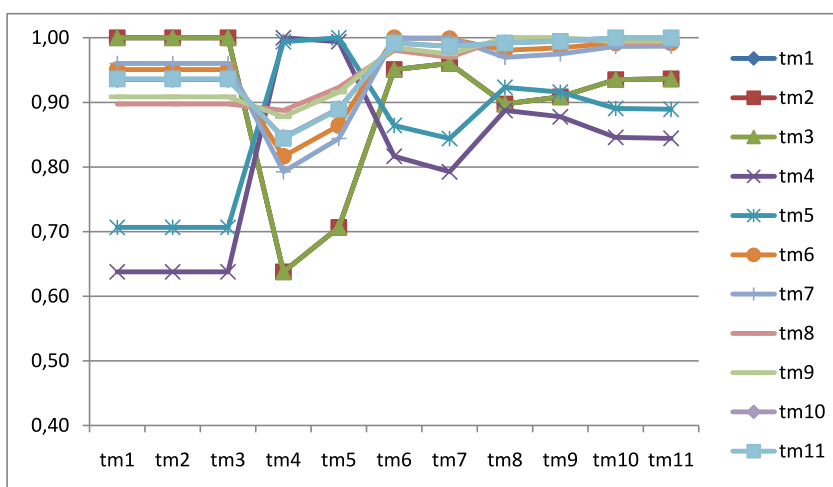


Figura 5.20: correlação entre Tm's para *primers* de tamanho 0 a 20% de GC

Na tabela 5.17 de correlação entre as Tm's para *primers* com 0 a 20% de GC, vemos que a Tm4 e a Tm5 tem menores índices de correlação com as outras Tm's principalmente com a Tm1, Tm2 e a Tm3. Isso pode ser observado na figura 5.20, onde a representante da Tm4 e da Tm5 se destacam das outras. Assim, a Tm4 e a Tm5 podem apresentar um comportamento desproporcional em relação as outras Tm's. Ou seja, quando as tm's apresentarem um aumento de temperatura para algum *primer*, elas apresentaram também um aumento, porém seu aumento pode estar desproporcional as outras Tm's.

Tabela 5.18: correlação entre as Tm's para *primers* de 20 a 40% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,76	0,79	0,94	0,93	0,93	0,94	0,94	0,94
Tm2	1,00	1,00	1,00	0,76	0,79	0,94	0,93	0,93	0,94	0,94	0,94
Tm3	1,00	1,00	1,00	0,76	0,79	0,94	0,93	0,93	0,94	0,94	0,94
Tm4	0,76	0,76	0,76	1,00	1,00	0,87	0,85	0,91	0,90	0,90	0,90
Tm5	0,79	0,79	0,79	1,00	1,00	0,88	0,87	0,93	0,92	0,92	0,92
Tm6	0,94	0,94	0,94	0,87	0,88	1,00	1,00	0,96	0,96	0,97	0,97
Tm7	0,93	0,93	0,93	0,85	0,87	1,00	1,00	0,95	0,95	0,96	0,96
Tm8	0,93	0,93	0,93	0,91	0,93	0,96	0,95	1,00	1,00	0,99	0,99
Tm9	0,94	0,94	0,94	0,90	0,92	0,96	0,95	1,00	1,00	0,99	0,99
Tm10	0,94	0,94	0,94	0,90	0,92	0,97	0,96	0,99	0,99	1,00	1,00
Tm11	0,94	0,94	0,94	0,90	0,92	0,97	0,96	0,99	0,99	1,00	1,00

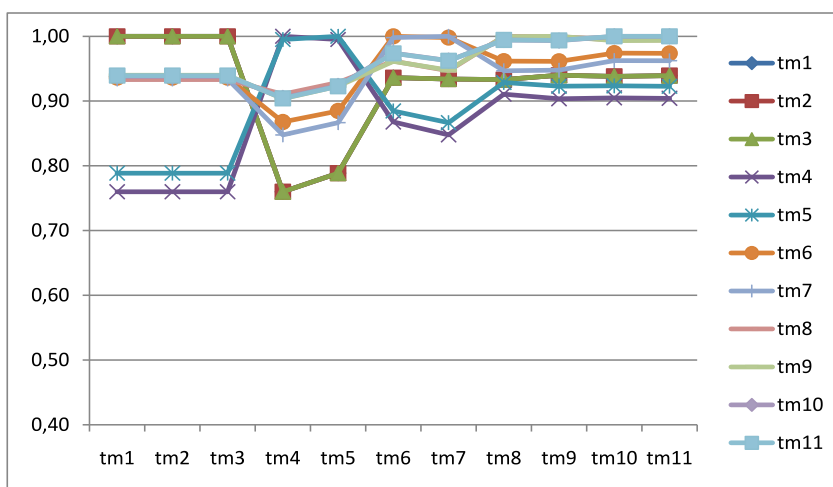


Figura 5.21: correlação entre Tm's para *primers* de tamanho 20 a 40% de GC

Já a tabela 5.18 nos dá a correlação entre as Tm's para concentrações de 20 a 40% de GC nos *primers*. Ela nos mostra que o índice de correlação entre a Tm4 e a Tm5 aumentam e relação as outras Tm's, enquanto que a Tm6, Tm7, Tm8 e Tm9 apresentaram valores de correlação ligeiramente menores. Isso pode ser melhor observado na figura 5.21.

Tabela 5.19: correlação entre as Tm's para *primers* de 40 a 60% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,85	0,85	0,90	0,88	0,96	0,96	0,95	0,95
Tm2	1,00	1,00	1,00	0,85	0,85	0,90	0,88	0,96	0,96	0,95	0,95
Tm3	1,00	1,00	1,00	0,85	0,85	0,90	0,88	0,96	0,96	0,95	0,95
Tm4	0,85	0,85	0,85	1,00	1,00	0,88	0,86	0,93	0,92	0,94	0,94
Tm5	0,85	0,85	0,85	1,00	1,00	0,88	0,86	0,93	0,93	0,95	0,95
Tm6	0,90	0,90	0,90	0,88	0,88	1,00	1,00	0,90	0,89	0,94	0,94
Tm7	0,88	0,88	0,88	0,86	0,86	1,00	1,00	0,88	0,87	0,93	0,92
Tm8	0,96	0,96	0,96	0,93	0,93	0,90	0,88	1,00	1,00	0,98	0,99
Tm9	0,96	0,96	0,96	0,92	0,93	0,89	0,87	1,00	1,00	0,98	0,98
Tm10	0,95	0,95	0,95	0,94	0,95	0,94	0,93	0,98	0,98	1,00	1,00
Tm11	0,95	0,95	0,95	0,94	0,95	0,94	0,92	0,99	0,98	1,00	1,00

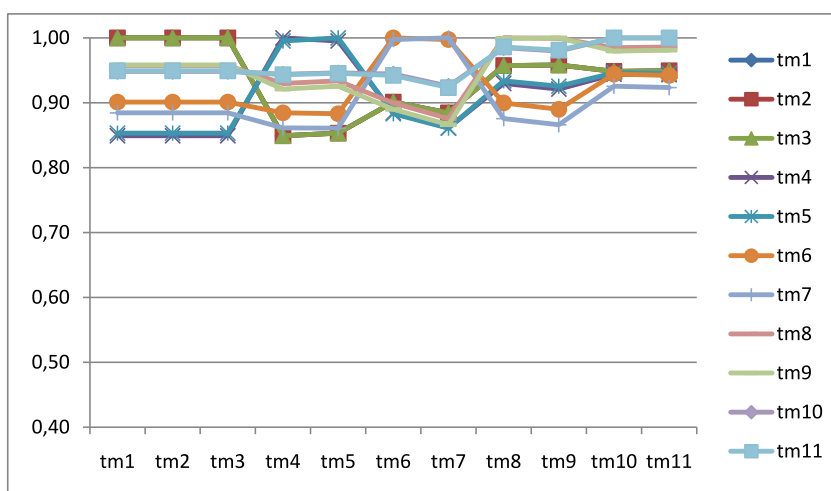


Figura 5.22: correlação entre Tm's para *primers* de tamanho 40 a 60% de GC

A tabela 5.19 nos mostra a correlação entre as Tm's para *primers* com 40 a 60% de GC. Todas as Tm's possuem um alto índice de correlação para esse intervalo de concentração de GC. Assim, as Tm's apresentam comportamento linear próximos e valores para temperatura de melting proporcionais. Veja que, na figura 5.22, as representates estão todas com valores próximos de 1 (todas maiores que 0,85).

Tabela 5.20: correlação entre as Tm's para *primers* de 60 a 80% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,90	0,88	0,83	0,80	0,96	0,95	0,96	0,96
Tm2	1,00	1,00	1,00	0,90	0,88	0,83	0,80	0,96	0,95	0,96	0,96
Tm3	1,00	1,00	1,00	0,90	0,88	0,83	0,80	0,96	0,95	0,96	0,96
Tm4	0,90	0,90	0,90	1,00	1,00	0,88	0,85	0,90	0,89	0,96	0,96
Tm5	0,88	0,88	0,88	1,00	1,00	0,87	0,84	0,89	0,87	0,95	0,95
Tm6	0,83	0,83	0,83	0,88	0,87	1,00	1,00	0,75	0,72	0,87	0,87
Tm7	0,80	0,80	0,80	0,85	0,84	1,00	1,00	0,71	0,68	0,84	0,84
Tm8	0,96	0,96	0,96	0,90	0,89	0,75	0,71	1,00	1,00	0,97	0,97
Tm9	0,95	0,95	0,95	0,89	0,87	0,72	0,68	1,00	1,00	0,95	0,96
Tm10	0,96	0,96	0,96	0,96	0,95	0,87	0,84	0,97	0,95	1,00	1,00
Tm11	0,96	0,96	0,96	0,96	0,95	0,87	0,84	0,97	0,96	1,00	1,00

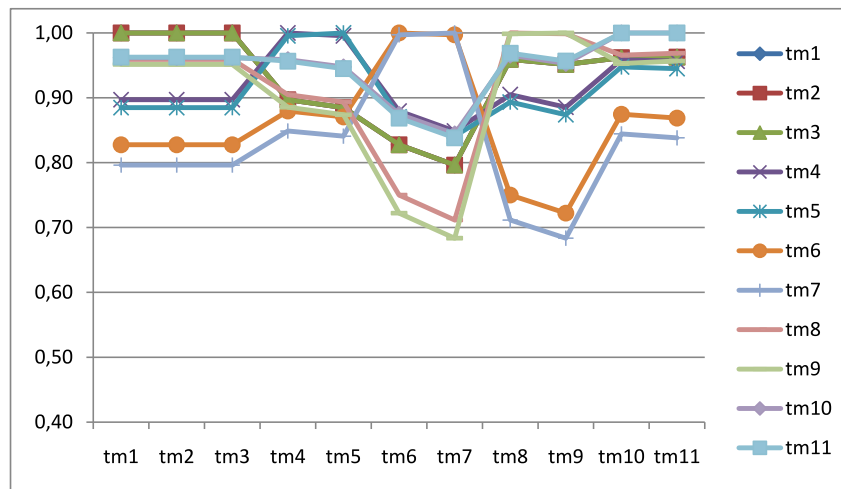


Figura 5.23: correlação entre Tm's para *primers* de tamanho 60 a 80% de GC

A tabela 5.20, de correlação entre as Tm's para *primers* de 60 a 80% de GC, nos mostra que a Tm4 e a Tm5 deixam o posto de fórmulas com menores índices de correlação e a Tm6, Tm7, Tm8 e a Tm9 tomam o seu lugar. Isso é fácil de ver através da figura 5.23, onde as representantes dessas Tm's estão abaixo das demais.

Tabela 5.21: correlação entre entre as Tm's para *primers* de 80 a 100% de GC

	Tm1	Tm2	Tm3	Tm4	Tm5	Tm6	Tm7	Tm8	Tm9	Tm10	Tm11
Tm1	1,00	1,00	1,00	0,95	0,93	0,76	0,72	0,94	0,92	0,99	0,99
Tm2	1,00	1,00	1,00	0,95	0,93	0,76	0,72	0,94	0,92	0,99	0,99
Tm3	1,00	1,00	1,00	0,95	0,93	0,76	0,72	0,94	0,92	0,99	0,99
Tm4	0,95	0,95	0,95	1,00	1,00	0,88	0,85	0,83	0,80	0,97	0,96
Tm5	0,93	0,93	0,93	1,00	1,00	0,89	0,86	0,81	0,77	0,95	0,95
Tm6	0,76	0,76	0,76	0,88	0,89	1,00	1,00	0,53	0,49	0,79	0,78
Tm7	0,72	0,72	0,72	0,85	0,86	1,00	1,00	0,49	0,44	0,76	0,74
Tm8	0,94	0,94	0,94	0,83	0,81	0,53	0,49	1,00	1,00	0,93	0,94
Tm9	0,92	0,92	0,92	0,80	0,77	0,49	0,44	1,00	1,00	0,91	0,92
Tm10	0,99	0,99	0,99	0,97	0,95	0,79	0,76	0,93	0,91	1,00	1,00
Tm11	0,99	0,99	0,99	0,96	0,95	0,78	0,74	0,94	0,92	1,00	1,00

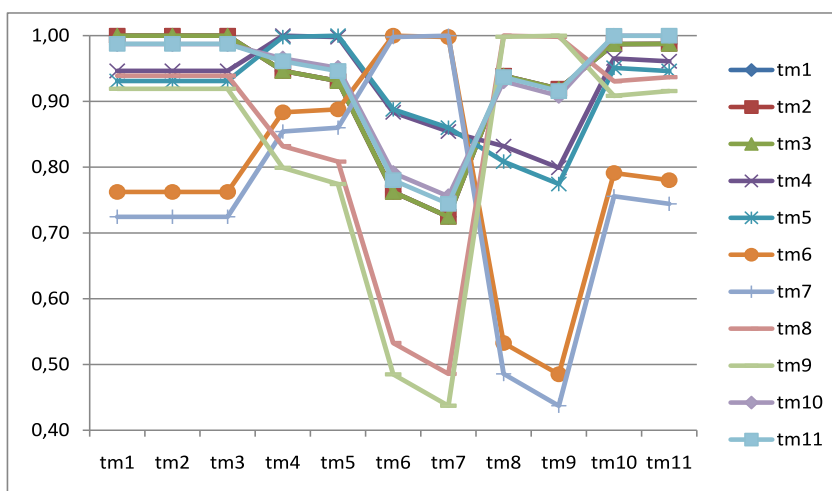


Figura 5.24: correlação entre Tm's para *primers* de tamanho 90 a 100% GC

Por fim, a tabela 5.21, nos dá a correlação entre as Tm's para *primers* com 80 a 100% de GC em sua composição. É fácil observar, que a Tm6, Tm7, Tm8 e a Tm9 apresentam índices de correlação inferiores às demais Tm's. Sendo que os menores índices ocorrem entre a Tm6 e a Tm7 em relação a Tm8 e a Tm9. Assim, para primers com alta concentração de GC, essas Tm's podem apresentar um comportamento muito desproporcional das demais.

Capítulo 6

Conclusão

Referências Bibliográficas

- [1] Arial, R. T. Predição de rna não codificadores no transcriptoma do fungo paracoccioides brasiliensis usando aprendizagem de máquina., Abril 2008.
- [2] Baldi, P. and Brunak, S. Bioinformatics - the machine learning approach. *MIT press*, 2001.
- [3] Chang, C. -C. and Lin C. -J. Libsvm: a library for support vector machines., February 2009.
- [4] Eddy, S. R. Non-coding rna genes and the modern rna world. *Nat Rev Genet*, 2(12):919–929, December 2001.
- [5] Frith, M. C.; Bailey, T. L.; Kasukawa, T.; Mignone, F.; Kummerfeld, S. K.; Madera, M.; Sunkara, S.; Furuno, M.; Bult, C. J.; Quackenbush, J.; Kai, C. Kawai, J.; Carninci, P.; Hayashizaki, Y.; Pesole, G. and Mattick, J. S. Discrimination of non-protein-coding transcripts from protein-coding mrna. *RNA Biol*, 2006.
- [6] Griffiths-Jones, S.; Moxon, S.; Marshall, M.; Khanna, A.; Eddy, S. R. and Bateman, A. Rfam: annotating non-coding rnas in complete genomes. *Nucleic Acids Research* 33, 2005.
- [7] Hsu, C. -W.; Chang, C. -C. and Lin, C. -J. . A practical guide to support vector classification., May 2009.
- [8] Hüttenhofer, A.; Schatther, P. and Polacek, N. Non-coding rnas: hope or hype?, 2005.
- [9] Joachims, T. Making large-scale svm learning practical. in: *Advances in kernel methods - support vector learning. MIT press*, 1999.
- [10] Kin, T.; Yamada, K.; Terai, G.; Okida, H.; Yoshinari, Y.; Ono, Y.; Kojima, A.; Kimura, Y.; Komori, T. and Asai, K. . frnadb: a platform for mining/annotating functional rna candidates from non-coding rna sequences. *Nucleic Acids Res*, 35(Database issue), January 2007.
- [11] Larrañaga, P.; Calvo, B.; Santana, R.; Bielza, C.; Galdiano, J.; Inza, I.; Lozano, J. A.; Armañanzas, R.; Santafé, G.; Pérez, A. and Robles, V. Machine learning in bioinformatics. *Briefings in bioinformatics*, 2006.

- [12] Lima, A. M. Predição de rnas não codificantes e sua aplicação na busca do componente rna da telomerase., 2006.
- [13] Liu, C.; Bai, B.; Skogerbo, G.; Cai, L.; Deng, W.; Zhang, Y.; Bu, D.; Zhao, Y. and Chen, R. Noncode: an integrated knowledge database of non-coding rnas. *Nucleic Acids Res*, 33(Database issue), January 2005.
- [14] Liu, J.; Gough, J. and Rost, B. Distinguishing protein-coding from non-coding rnas through support vector machines., 2006.
- [15] Lorena, A. C. *Investigação de estratégias para geração de máquinas de vetores de suporte multiclasse*. PhD thesis, Universidade de São Paulo (São Carlos), 2006.
- [16] Machado-Lima; Portillo H. A. and Durham A. M. Computational methods in non-coding rna research. *Mathematical Biology*, 2007.
- [17] Mattick, J. S. and Makunin I. V. Non-coding rna, australia research council for functional and applied genomics, February 2006.
- [18] Mitchell, T. M. . *Machine Learning*. McGraw-Hill, 1997.
- [19] Monard, M. C and Baranauskas, J. A. . *Sistemas Inteligentes: Fundamentos e Aplicações*. Manole, 2003.
- [20] Neves, M. I. S. C. and Lorena, C. R. M. . Desvendando aplicações da biotecnologia no ensino formal., 2009.
- [21] Pang, K. C.; Frith, M. C. and Mattick, J. S. Rapid evolution of noncoding rnas: lack of conservation does not mean lack of function. *Trends in Genetics*, 2006.
- [22] Schölkopf, B. and Smola, A. J. Learning with kernels: Support vector machines, regularization, optimization, and beyond. *The MIT press, London - England*, 2002.
- [23] Setubal, J. C. and Meidanis, J. Introduction to computational molecular biology., 2000.
- [24] Silva, T. C. C. Som-portrait: um método para identificar rna não codificador utilizando mapas auto organizáveis., 2009.
- [25] Szymanski, M.; Barciszewski J. and ErdmannV. A. Noncoding rnas: Molecular biology and molecular medicine., 2003.
- [26] Torarinsson, E.; Sawera, M.; Hacgaard, J. H.; Fredholm, M. and Gorodkin, J. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common rna structure., 2006.
- [27] Vapnik, V. N. Statistical learning theory, 1998.
- [28] Wahlestedt, C. Natural antisense and noncoding rna transcripts as potential drug targets., 2006.

- [29] Zhang, S.; Hass, B.; Eskin, E and Bafna, V. Searching genomes for non-coding rna using fastr. *IEEE/ACM Trans. on Comput. Biol. and Bioinfo*, 2005.