

**Desenvolvimento de uma representação mesoscópica  
para marcadores fluorescentes em DNA e predição de  
intensidade de fluorescência**

**Dissertação de Mestrado**

PÂMELLA MIRANDA DE MOURA  
GERALD WEBER (ORIENTADOR)

Belo Horizonte  
Janeiro de 2020

PÂMELLA MIRANDA DE MOURA

**Desenvolvimento de uma representação mesoscópica  
para marcadores fluorescentes em DNA e predição de  
intensidade de fluorescência**

Dissertação apresentada ao Programa  
Interunidades de Pós-Graduação em Bi-  
oinformática da Universidade Federal de  
Minas Gerais para obtenção do grau de  
Mestre em Bioinformática.

Orientador: Gerald Weber

PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Belo Horizonte  
Janeiro de 2020

- 043      Moura, Pâmella Miranda de.  
Desenvolvimento de uma representação mesoscópica para marcadores fluorescentes em DNA e predição de intensidade de fluorescência [manuscrito] / Pâmella Miranda de Moura. – 2020.  
91 f. : il. ; 29,5 cm.
- Orientador: Gerald Weber.  
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.
1. Biofísica. 2. DNA. 3. Ácidos nucleicos. I. Weber, Gerald. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004



**ATA DA DEFESA DE DISSERTAÇÃO**

**Pâmella Miranda de Moura**

71/2020  
entrada  
1º/2018  
CPF:  
015.357.416-00

Às quatorze horas do dia **18 de fevereiro de 2020**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Desenvolvimento de uma representação mesoscópica para marcadores fluorescentes em DNA e predição de intensidade de fluorescência**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, o Presidente da Comissão, **Dr. Gerald Weber**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Dr. Gerald Weber	UFMG	02491500850	APROVADA
Dra. Mariana Torquato Quezado de Magalhaes	ICB/UFMG	905.464.151-72	APROVADA
Dr. Luiz Cláudio de Carvalho	UFMG	76578542604	APROVADA

Pelas indicações, a candidata foi considerada: APROVADA  
O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.  
**Belo Horizonte, 18 de fevereiro de 2020.**

Dr. Gerald Weber - Orientador

Dra. Mariana Torquato Quezado de Magalhaes

Dr. Luiz Cláudio de Carvalho

”A persistência realiza o impossível.”

Provérbio chinês

# 1 Agradecimentos

Obrigada, Gerald, por ter aceitado me orientar nessa caminhada e, apesar das pedras pelo caminho, acreditou em mim e teve toda a paciência para chegarmos até aqui.

Obrigada, mamãe, por me apoiar em meus sonhos, me ajudar de todas as formas que pudesse e por ter me dado força acreditando comigo que seria possível.

Obrigada, Gladystone, por segurar a minha mão durante a caminhada, pelo abraço nos momentos nublados e por compreender o quão importante meus sonhos são para mim.

Obrigada, Luciana, por todas as coisas que me ensinou e todo o incentivo ao longo da caminhada. Por toda a ajuda mesmo em meio à sua lista de tarefas tão grande quanto a minha.

Obrigada, Alessandra, por ter surgido no primeiro grande desafio e ter permanecido na minha vida mesmo em meio às nossas vidas atribuladas.

Obrigada ao meu grupo de pesquisa — GBC — pelas risadas, pelo apoio mútuo, pela compreensão nos momentos de desespero e loucura, enfim, obrigada por estarem lá e aqui.

Obrigada aos meus amigos e pessoas maravilhosas que atravessaram essa caminhada e me ofereceram boa sorte, seja em palavras ou com o mais discreto sorriso. Isso foi muito importante para mim!

Obrigada à Profa. Kira Astakhova, da Universidade Técnica da Dinamarca, pela nossa colaboração e a confiança em nosso trabalho.

Obrigada a todos os meus professores que me ensinaram muito e contribuíram não somente com meu crescimento como profissional, mas também com meu desenvolvimento como pessoa.

Obrigada à Secretaria do Programa de Pós-Graduação da Bioinformática pelas informações e disponibilidade, em especial, à Sheila e ao Tiago.

Obrigada ao Departamento de Física (ICEx) e ao Instituto de Ciências Biológicas (ICB) por serem mais um teto para mim.

Obrigada à FAPEMIG pela bolsa concedida e ajuda de custos, assim como também à Capes pelo apoio financeiro.

E obrigada a todos aqueles que acreditaram em seus sonhos e não desistiram. Suas conquistas são os pilares das nossas.

## 2 Resumo

Marcadores fluorescentes atados à moléculas de DNA são muito usados nas áreas de biomedicina e biotecnologia, por exemplo, e podem ser estudados aplicando-se modelos mesoscópicos, como o modelo Peyrard-Bishop (PB). Em particular, os marcadores Cy3 e Cy5 têm mostrado grande eficiência e bons resultados em sua performance, além de oferecerem uma maior estabilidade à molécula de DNA. Em nosso trabalho, otimizamos o parâmetro de potencial de Morse representando a ligação de hidrogênio e o potencial harmônico representando a interação de empilhamento para duplexos de DNA atados à Cy3 e Cy5. A partir de um conjunto de temperaturas de *melting* experimentais, coletados da literatura, para essa combinação de DNA + Cy3/Cy5, aplicamos o modelo mesoscópico Peyrard-Bishop para modelar esses parâmetros. Este modelo é um modelo de física estatística que nos permite obter algumas propriedades da molécula de DNA, a partir do cálculo da função partição clássica. Nossos resultados para Cy3 e Cy5 mostraram valores similares ao de par de bases AT em DNA. Isto confirma que Cy3 e Cy5 se ligam ao DNA como um par de bases extra, assim como tem sido sugerido em medidas de ressonância magnética nuclear. Outro problema que tratamos é a questão se o modelo mesoscópico poderia servir de preditor da intensidade de fluorescência de Cy3 e Cy5 no DNA. Para este fim tratamos outro conjunto de dados de microarranjo relacionados à Cy3 e Cy5 acoplados ao terminal 5' da molécula de DNA. O conjunto consiste de 1024 sequências de DNA, sintetizadas de forma uniforme, com dados de intensidade de fluorescência. Calculamos o deslocamento dos pares de bases na direção das ligações de hidrogênio, chamado de abertura média, que caracteriza o processo de desnaturação da dupla hélice de DNA. Os valores obtidos para esse deslocamento foram correlacionados com as intensidades de fluorescência. Correlacionamos esses dados com as aberturas médias para variadas condições e diferentes parâmetros, e encontramos uma alta correlação. Em seguida, desenvolvemos um programa para maximizar a correlação variando os parâmetros do modelo mesoscópico. Os resultados obtidos para a parametrização das temperaturas de desnaturação e a alta correlação das intensidades de fluorescência mostram que o modelo mesoscópico é uma ferramenta adequada para otimização de sondas de DNA.

Palavras-chave: marcadores fluorescentes, DNA, modelos mesoscópicos, temperatura de desnaturação, microarranjo

### 3 Abstract

Fluorescent dyes attached to DNA molecules are used in biomedicine and biotechnology areas, for example, and they can be studied with mesoscopic models such as Peyrard-Bishop model (PB). In particular, Cy3 and Cy5 dyes showed a great efficiency and good results in their performance, also they provide a higher stability to DNA molecule. In our work, we optimized the parameter for Morse potentials that represents the hydrogen bonding, and a harmonic potential that represents the stacking interaction for DNA duplexes attached to Cy3 and Cy5. Using an experimental melting temperatures set, collected from the literature, for this combination of DNA + Cy3/Cy5, we applied the PB mesoscopic model to describe these parameters. The PB model is a statistical physics model that allows us to obtain some properties of the DNA molecule, by calculating the canonical partition function. Our results for Cy3 and Cy5 show values which are similar to AT base pair in DNA. This confirms that Cy3 and Cy5 links to DNA like an extra base pair, as was suggested from nuclear magnetic resonance measurements. Another problem that we dealt with was the question if the mesoscopic model could serve as predictor for fluorescence intensity of Cy3 and Cy5 in DNA. To this end, we used another dataset of microarray intensities related to Cy3 and Cy5 coupled at 5' terminal of the DNA molecule. The set consists of 1024 DNA sequences, synthesized in an uniform way, with fluorescence intensity data. We calculated the displacement of base pairs in the direction of hydrogen bonds, called average opening, that characterizes the denaturation process of DNA double helix. The values obtained for this displacement were correlated to fluorescence intensities. We correlated these data to average openings for varied conditions and different parameters, and we found a high correlation. Then we developed a program to maximize the correlation varying the mesoscopic model parameters. The results obtained for the denaturation temperatures parametrization and the high correlation for fluorescence intensities show that the mesoscopic model is an appropriate tool to optimize DNA probes.

Keywords: fluorescent dyes, DNA, mesoscopic models, melting temperature, microarray



# Sumário

<b>1</b>	<b>Agradecimentos</b>	<b>2</b>
<b>2</b>	<b>Resumo</b>	<b>3</b>
<b>3</b>	<b>Abstract</b>	<b>4</b>
<b>4</b>	<b>Introdução</b>	<b>8</b>
4.1	Fluorescência: uma visão geral . . . . .	8
4.2	DNA: aspectos importantes . . . . .	12
4.3	Desnaturação de DNA . . . . .	15
<b>5</b>	<b>Objetivos</b>	<b>17</b>
5.1	Objetivo Geral . . . . .	17
5.2	Objetivos Específicos . . . . .	17
<b>6</b>	<b>Marcadores Fluorescentes</b>	<b>18</b>
6.1	Marcadores Cianina . . . . .	18
6.2	Cianinas Metino: Cy3 e Cy5 . . . . .	20
6.3	Conclusão . . . . .	24
<b>7</b>	<b>Métodos</b>	<b>25</b>
7.1	Modelo de desnaturação em dois estados . . . . .	25
7.2	Modelo Mesoscópico Peyrard-Bishop (PB) . . . . .	27
7.3	Otimização de Parâmetros . . . . .	31
7.4	Coeficiente de Correlação de Pearson . . . . .	34
<b>8</b>	<b>Parametrização por Temperatura de <i>Melting</i></b>	<b>36</b>
8.1	Dados Experimentais . . . . .	36
8.2	Adaptação do modelo PB . . . . .	37
8.3	Procedimentos da otimização . . . . .	37
8.4	Resultados e Discussão . . . . .	39
8.5	Conclusão . . . . .	45
<b>9</b>	<b>Correlação da Intensidade de Fluorescência</b>	<b>46</b>
9.1	Dados Experimentais . . . . .	46
9.2	Procedimentos . . . . .	46
9.3	Resultados . . . . .	49
9.4	Discussão . . . . .	55
9.5	Conclusão . . . . .	58

<b>10 Conclusões</b>	<b>59</b>
<b>11 Perspectivas</b>	<b>62</b>
<b>A Dados referentes aos procedimentos do capítulo 8</b>	<b>63</b>
A.1 Sequências modificadas de DNA e Temperaturas de <i>melting</i> . . . . .	63
A.2 Perfis de abertura média para Cy3 e Cy5 . . . . .	65
<b>B Dados referentes aos procedimentos do capítulo 9</b>	<b>66</b>
B.1 Script para montagem das sequências . . . . .	66
B.2 Scripts para gerar sequências complementares . . . . .	67
B.3 Script para calcular aberturas médias . . . . .	69
B.4 Script para extrair aberturas médias dos marcadores . . . . .	69
B.5 Script para calcular coeficiente de correlação de Pearson . . . . .	70
B.6 Coeficientes de correlação $\times$ Temperaturas de cálculo . . . . .	74
B.7 Script para otimização das correlações de Pearson . . . . .	75
B.8 Script para cálculo do $p$ -valor . . . . .	79
<b>Referências</b>	<b>84</b>

## Abreviações usadas

**DNA** ácido desoxirribonucleico (*deoxyribonucleic acid*)

**FCS** espectroscopia de correlação de fluorescência (*fluorescence correlation spectroscopy*)

**FISH** hibridização fluorescente *in situ* (*fluorescence in situ hybridization*)

**FRET** fluorescência por transferência ressonante de energia (*Fluorescence resonance energy transfer*)

**NMR** ressonância magnética nuclear (*nuclear magnetic resonance*)

**PB** Peyrard-Bishop

**PCR** reação em cadeia da polimerase (*polymerase chain reaction*)

**RNA** ácido ribonucleico (*ribonucleic acid*)

**SNP** polimorfismo de nucleotídeo único (*single nucleotide polymorphism*)

**UV** ultra-violeta

## 4 Introdução

A busca por uma compreensão cada vez maior sobre a molécula de ácido desoxirribonucleico (*deoxyribonucleic acid*, **DNA**) e suas propriedades tem levado, ano após ano, muitos pesquisadores a encontrarem novas informações que possam aprimorar técnicas e impulsionar novos meios de análise e aplicação.

O aprofundamento no funcionamento do **DNA** e a relação deste com o meio tem promovido o desenvolvimento de técnicas aplicadas em tratamentos, meios de diagnóstico e auxílio nos estudos de diversas doenças. A hibridização fluorescente *in situ* (*fluorescence in situ hybridization*, **FISH**), microarranjo de **DNA**, reação em cadeia da polimerase (*polymerase chain reaction*, **PCR**), fluorescência por transferência ressonante de energia (*Fluorescence resonance energy transfer*, **FRET**), eletroforese em gel, entre outros, são alguns métodos que analisam e estudam sequências de **DNA**.

Neste trabalho, apresentamos dois estudos sobre os marcadores fluorescentes Cy3 e Cy5 atados à molécula de **DNA**. No primeiro, modelamos Cy3 e Cy5 aplicando o modelo mesoscópico Peyrard-Bishop, que será discutido em maior detalhe na seção 7.2, a partir de um conjunto de temperaturas de *melting*. No segundo, mostramos se modelos mesoscópicos podem ser bons preditores para as intensidades de fluorescência de Cy3 e Cy5 a partir de resultados de correlação, o que será visto com mais detalhes na seção 9.

Este trabalho está estruturado com introdução; objetivos; informações sobre marcadores cianinas, mais detalhadamente, Cy3 e Cy5; os métodos que utilizamos; os procedimentos para a parametrização dos marcadores; os procedimentos para a correlação das intensidades de fluorescência de Cy3 e Cy5; conclusões; perspectivas e apêndice com informações complementares ao trabalho.

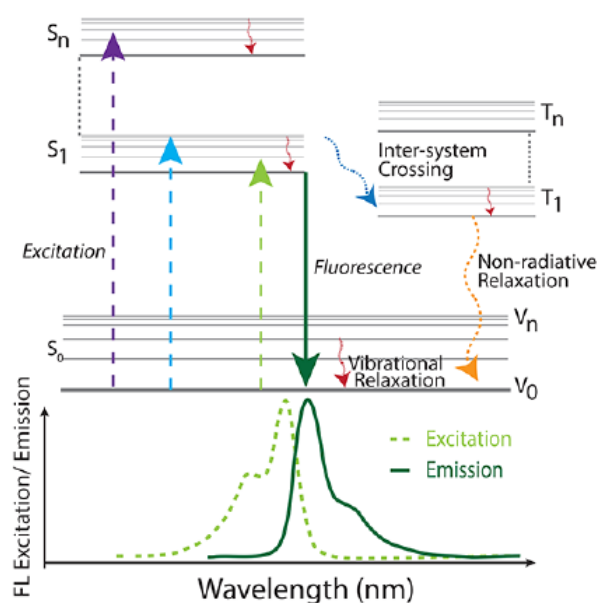
Neste capítulo, veremos alguns conceitos relevantes para este trabalho. Como funciona o processo de fluorescência; alguns aspectos sobre a molécula de **DNA** e, por fim, o que é e como ocorre o processo de desnaturação do duplexo de **DNA** — separação das fitas da dupla hélice.

### 4.1 Fluorescência: uma visão geral

Fluorescência é um tipo de luminescência, mais especificamente de fotoluminescência [1], em que luz é emitida devido à influência de uma fonte de energia [1, 2]. A excitação — emissão de luz — ocorre quando um ou mais fótons são absorvidos por uma molécula, a qual alcança um determinado estado eletrônico excitado e retorna, logo a seguir, ao estado eletrônico fundamental (estado não excitado) [1–4]. A energia dos fótons absorvidos, ou da luz incidente, deve ser aproximadamente a diferença de energia entre um estado eletrônico excitado e o estado fundamental [3, 4], sendo que a energia de um fóton é dada por:

$$E = h\nu \tag{1}$$

onde  $h$  é a constante de Planck e  $\nu$  é a frequência da luz. O retorno da molécula ao estado fundamental ocorre em etapas sucessivas, durante as quais a molécula emite dois ou mais fótons com energias menores e também comprimentos de onda mais longos que os incidentes [1, 4, 5]. O estado fundamental e alguns estados excitados são também chamados de estado singlete e cada estado eletrônico possui vários níveis de energia vibracionais e rotacionais [2, 4, 6]. À determinado comprimento de onda, uma molécula, que está no estado fundamental, absorve luz. Ao absorver fótons, a molécula passará do estado fundamental ( $S_0$ ) para um estado eletrônico excitado ( $S_1$  a  $S_n$ ) [1–4], correspondente às setas apontando para cima na Figura 1. Após determinado tempo, essa molécula irá emitir luz em um comprimento de onda diferente da do momento de absorção [1, 4, 5]. Nessa etapa, a molécula sofre relaxação dos estados vibracionais transitando dos estados excitados para o estado fundamental — momento da emissão de fótons — quando ocorre o fenômeno da fluorescência [1, 3, 4, 6], veja a seta contínua apontando para baixo na Figura 1.



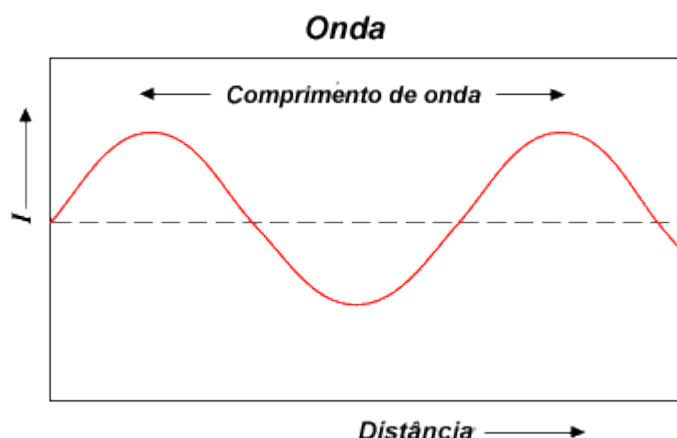
**Figura 1**

Esquema das transições quânticas da fluorescência.  $S_0$  representa o estado fundamental e seus níveis de energia vibracionais. A excitação para estados eletrônicos de maiores energias (estados excitados —  $S_1$  a  $S_n$ ) está representada pelas três setas apontando para cima. A ocorrência da fluorescência está representada pela seta contínua e verde apontando para baixo — o decaimento dos estados excitados para o estado fundamental. Ao lado, a relaxação dos estados vibracionais ( $V_0$  a  $V_n$ ). Os estados  $T_0$  a  $T_n$  são os estados tripletos relacionados ao fenômeno de fosforescência, que não é tratado em nosso trabalho. Abaixo do esquema de estados quânticos, o espectro de emissão por comprimentos de onda. Figura retirada da referência [3].

A fluorescência é um processo de ocorrência rápida, tipicamente a emissão de luz ocorre poucos nanossegundos após a absorção [4, 7]. Duas características são essenciais aos estudos de fluorescência: a eficiência quântica e o tempo de vida. A primeira é a relação entre o número de fótons emitidos pelo número de fótons absorvidos, enquanto a segunda é o tempo médio

entre o início da emissão até o cessar da fluorescência [2].

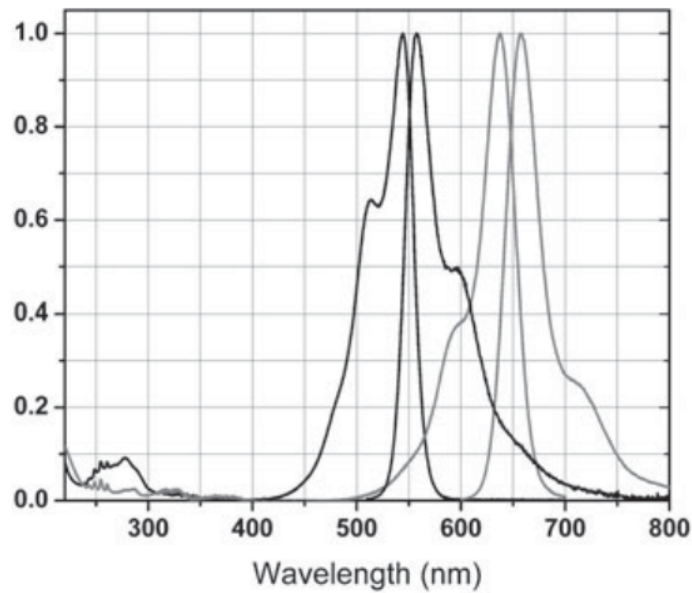
Em geral, a fluorescência ocorre a energias baixas, ou seja, comprimentos de onda longos [2]. Comprimento de onda, representado por  $\lambda$ , é a distância entre dois pontos da onda, que formam repetições, ao longo de seu deslocamento [8], por exemplo, de uma crista (ponto máximo da onda) à outra (Figura 2). O comprimento de onda é uma grandeza inversamente proporcional à frequência. Se a frequência aumenta, o comprimento de onda diminui e vice-versa. Proporcional à frequência e inversamente proporcional ao comprimento de onda, tem-se o número de onda, representado pela letra  $k$ , o qual é definido como a quantidade de comprimentos de onda em uma circunferência ( $2\pi$ ) [8]. Comprimento de onda (em nanômetros) e número de onda (em  $\text{cm}^{-1}$ ) são parâmetros usualmente relacionados à intensidade de fluorescência em espectros de emissão. Esse espectro é amplamente variável, além de ser dependente da estrutura química do fluoróforo e do solvente em que este está presente [2]. Na Figura 3, tem-se o exemplo de espectro de emissão e espectro de absorção, ambos sobrepostos. Esses espectros se referem aos marcadores fluorescentes Cy3 e Cy5, os quais serão tratados mais adiante neste trabalho. Os dois espectros da esquerda (mais escuros) são de Cy3 e os da direita (mais claros) são de Cy5.



**Figura 2**

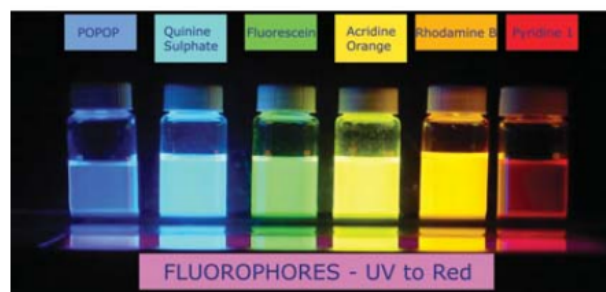
Gráfico de propagação de uma onda. O comprimento de onda ( $\lambda$ ) indica a distância entre dois pontos — uma forma da onda que se repete — em direção à propagação da onda. Figura retirada da referência [9].

A fluorescência é um processo empregado para estudos relacionados à interação da luz com a matéria. Ocorre, com maior frequência, em moléculas aromáticas e as taxas de emissão de luz são, em geral, da ordem de  $10^8 \text{ s}^{-1}$  [2]. Esse fenômeno tem sido aplicado em técnicas para investigação de moléculas como DNA, ácido ribonucleico (*ribonucleic acid*, RNA) e proteínas [11]. Em geral, a fluorescência pode ser obtida por meio de marcadores fluorescentes, naturais ou sintéticos, como se pode ver na Figura 4. Alguns parâmetros físicos e químicos podem influenciar a emissão de fluorescência de moléculas, tais como polaridade, pressão, pH, as ligações de hidrogênio e temperatura [1].



**Figura 3**

Espectro de absorção e emissão para Cy3 e Cy5. As curvas mais escuras à esquerda são a absorção (pico esquerdo) e a emissão (pico direito) referentes a Cy3 e as mais claras à direita, mostram a absorção e a emissão de Cy5. No eixo horizontal, os comprimentos de onda (em nanômetros). No eixo vertical, as intensidades de fluorescência. Figura retirada da referência [10].



**Figura 4**

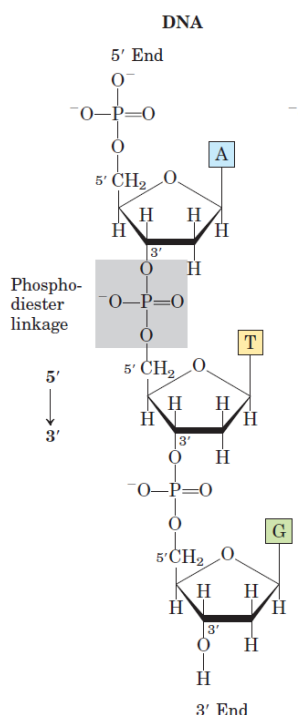
Alguns exemplos de marcadores fluorescentes (ou fluoróforos). Figura retirada da referência [2].

Uma das formas em que o processo de fluorescência é empregado em algumas técnicas é por meio de sondas fluorescentes. Elas são utilizadas em métodos de detecção de ácidos nucleicos e possuem tamanho inferior às sequências de **DNA** a serem analisadas [11]. Essas sondas são comumente constituídas de oligonucleotídeos contendo um marcador fluorescente atado em uma das suas pontas. Os marcadores fluorescentes interagem com o duplexo e podem aumentar a estabilidade dele, a qual varia com a sequência de bases do **DNA**, o tamanho do duplexo e a natureza do fluoróforo presente [11].

Além desse método, o fenômeno da fluorescência pode ser empregado em diagnósticos médicos, análise forense, sequenciamento de **DNA** e análise genética. Também em técnicas como excitação multifóton, citometria de fluxo, detecção de molécula única e espectroscopia de correlação de fluorescência (*fluorescence correlation spectroscopy*, **FCS**) [2].

## 4.2 DNA: aspectos importantes

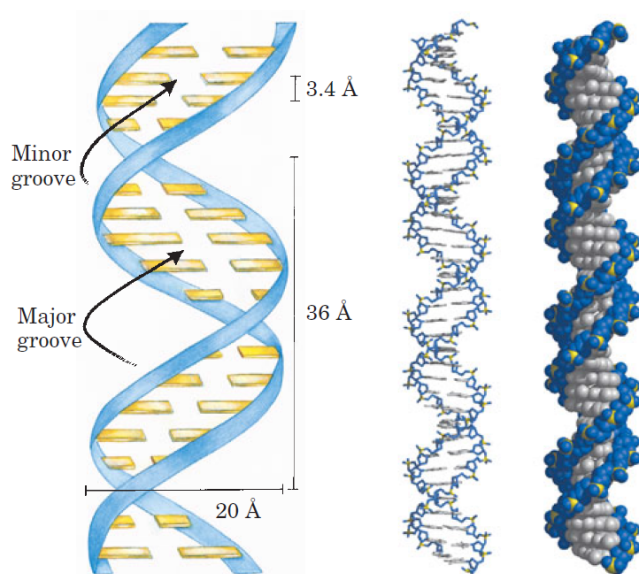
As informações essenciais para a constituição de um organismo se encontram presentes nos ácidos nucleicos: **DNA** e **RNA**. Eles são polímeros de nucleotídeos (grupo fosfato + pentose + base nitrogenada), os quais são unidos por ligações fosfodiéster entre o grupo 5'-hidroxila da pentose de um nucleotídeo e o grupo 3'-hidroxila da pentose do nucleotídeo seguinte [6, 12–14], como pode ser observado na Figura 5. Essa sequência de nucleotídeos forma a fita (ou cadeia) do ácido nucleico. A fita é orientada na direção do terminal 5' (fosfato livre) para o terminal 3' (hidroxila livre). O **DNA** é caracterizado pela presença da desoxirribose e é formado por duas cadeias antiparalelas (uma no sentido 5' → 3' e a outra 3' → 5'), complementares e enroladas, criando a aparência helicoidal característica do **DNA** [6, 12–14], como pode ser visto na Figura 6.



**Figura 5**

Fita simples de DNA. A área destacada ao centro (em cinza) mostra a ligação fosfodiéster entre os açúcares, os quais estão conectados às bases nitrogenadas, representadas pelos quadrados coloridos com as letras A (adenina), T (timina) e G (guanina). Fita orientada na direção 5' → 3'. Figura retirada da referência [12].

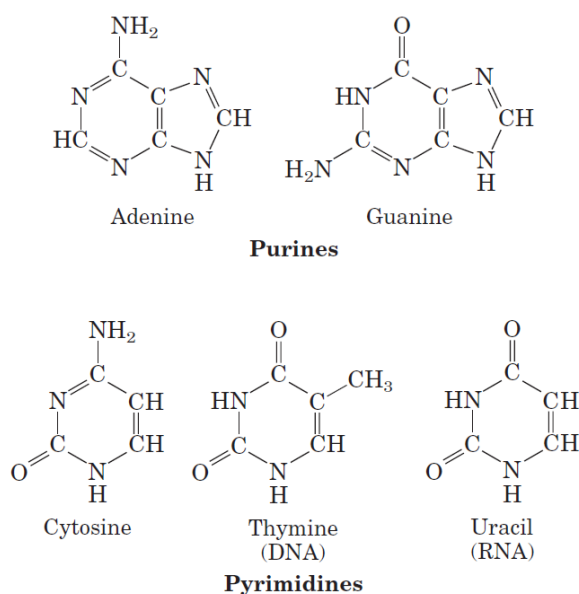




**Figura 6**

A molécula de DNA, com as ligações entre as bases nitrogenadas, possui uma aparência helicoidal. À esquerda, vê-se um esquema da fita dupla indicando a largura da hélice (20 Å), a distância entre cada volta (36 Å) e a distância entre cada base (3.4 Å). À direita, duas representações em 3D da molécula de DNA. Figura retirada da referência [12].

As duas fitas de nucleotídeos formam a dupla hélice do **DNA** e são constituídas por quatro bases nitrogenadas: adenina (A), timina (T), citosina (C) e guanina (G) [6, 12–16]. Elas são divididas em bases púricas – adenina e guanina – e bases pirimídicas – citosina e timina, veja Figura 7.

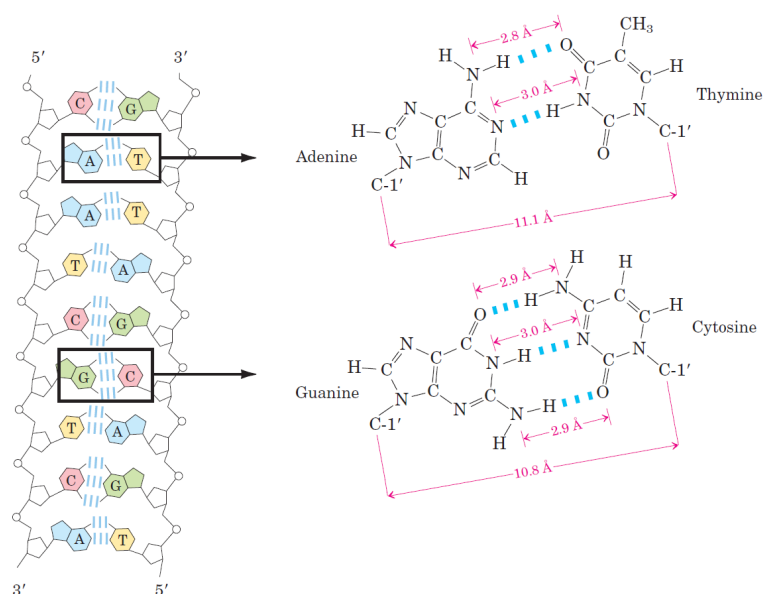


**Figura 7**

Acima, as estruturas químicas de adenina (A) e guanina (G) — as bases púricas. Abaixo, citosina (C), timina (T) e uracila (U) — as bases pirimídicas. A uracila, em geral, é encontrada em RNA no lugar da timina. Figura retirada da referência [12].

Próximo a pH neutro, as quatro bases são hidrofóbicas e relativamente insolúveis em água; em pH ácido, elas se tornam carregadas e, em água, sua solubilidade se eleva [12]. Devido aos seus anéis heterocíclicos, as bases absorvem luz ultra-violeta (UV) em comprimentos de onda próximos a 260 nm, contribuindo com a maior absorção da molécula de **DNA** [12, 17]. A relação entre as bases nitrogenadas é extremamente importante para a estabilidade do **DNA**. As bases interagem entre si ao longo de cada fita (bases vizinhas) e entre as fitas (pares de bases). As interações das bases entre as fitas são determinadas pelas ligações de hidrogênio e definidas

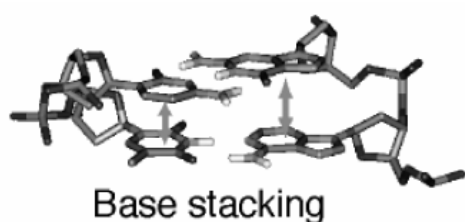
por dois pares estabelecidos: adenina interage com timina; citosina com guanina [6, 12, 16]. Entre citosina e guanina ocorrem três ligações de hidrogênio, enquanto entre adenina e timina ocorrem somente duas [6, 12–15], como se vê na Figura 8. A interação dos pares promove a complementaridade entre as bases, conhecida como complementaridade de Watson-Crick (ou pares de bases Watson-Crick), a qual é de grande importância uma vez que contribui para a estabilidade da dupla hélice. Em consequência a essa complementaridade, como definido pelas Leis de Chargaff, o número de resíduos de adenina encontrado em uma amostra de DNA é igual ao número de resíduos de timina, da mesma forma que o número de resíduos de guanina é igual ao de citosina [12–15].



**Figura 8**

À esquerda, a dupla fita de DNA mostrando as ligações entre as bases nitrogenadas: timina (T) com adenina (A) e guanina (G) com citosina (C). À direita, ampliação das ligações, mostrando como as bases se ligam e as três ligações de hidrogênio para G e C e as duas ligações para A e T. Figura retirada da referência [12].

As bases nitrogenadas, além de interagirem aos pares por meio das ligações de hidrogênio, também interagem com as bases vizinhas que juntas constituem uma fita de DNA. Como as bases se encontram em uma conformação planar na fita, ocorre interações eletrostáticas entre os anéis aromáticos das bases, o que é conhecido como interação de empilhamento (Figura 9). Tanto as ligações de hidrogênio, quanto as interações de empilhamento contribuem para a estabilidade da dupla hélice de DNA, sendo que a maior contribuição provém da segunda [12, 14]. A interação entre os pares de bases contribui para a aparência helicoidal do DNA formando um padrão a cada volta da hélice, (Figura 6). Esse padrão é uma quantidade determinada de pares de bases a cada volta e foi identificado por Watson e Crick [18]. Eles encontraram que a cada volta da hélice de DNA, há aproximadamente 10 pares de bases ou 36 Å (3.6 nm) [12].

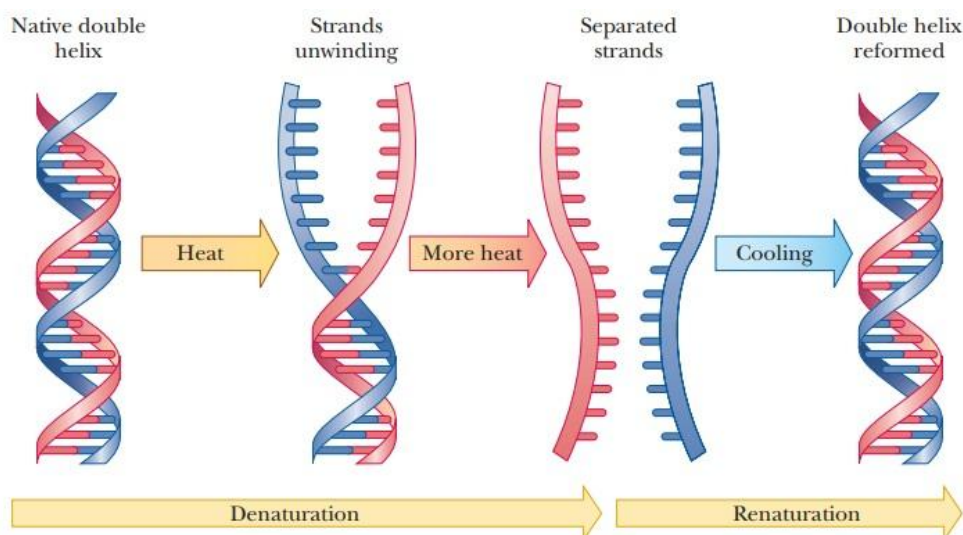


**Figura 9**

Interação de empilhamento (*base stacking*) entre as bases nitrogenadas, representada pelas setas. Observa-se como uma base está sobre a outra em uma conformação planar. Figura retirada da referência [19].

### 4.3 Desnaturação de DNA

As fitas de **DNA** estão ligadas entre si pela interação dos pares de bases — as ligações de hidrogênio. Mantendo a molécula de **DNA** a uma temperatura apropriada, a dupla hélice se mantém estável. Quando a temperatura sofre alguma elevação, a molécula de **DNA** começa a se desestabilizar. O aumento na temperatura influi nas ligações de hidrogênio entre os pares de bases. Essa desestabilização faz com que as fitas de **DNA** comecem a se separar, podendo sofrer uma separação total [13, 14, 20, 21], como pode ser observado na Figura 10. O processo em que as duas fitas de **DNA** desestabilizam resultando em sua separação é conhecido como “desnaturação”. A temperatura em que ocorre a dissociação das fitas de **DNA** é dependente das bases nitrogenadas presentes na molécula. Como a interação CG é mais forte que AT, devido à ligação de hidrogênio extra, o conteúdo de CG em um duplexo de **DNA** se relaciona diretamente ao valor da temperatura para a desnaturação da molécula. Quanto maior o conteúdo de CG no duplexo, maior será a dificuldade para o rompimento das ligações entre as fitas de **DNA** ocasionando a desnaturação do duplexo [12, 20, 22].

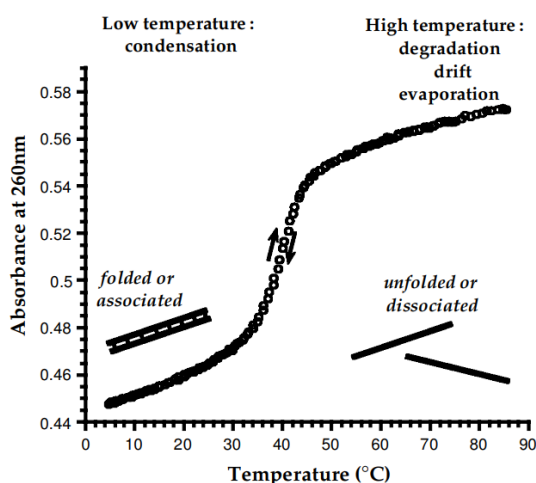


**Figura 10**

Desnaturação do DNA — Elevando a temperatura do duplexo de DNA, a dupla fita começará a se dissociar. À determinada temperatura, o duplexo se dissociará completamente. Renaturação do DNA — Esfriando as duas fitas dissociadas, após certo tempo, elas se associarão formando a dupla fita de DNA novamente. Figura retirada de [23].

A desnaturação do **DNA** pode ser estudada através de perfis que mostram a absorção pela temperatura [22, 24]. Como mencionado anteriormente, as bases nitrogenadas absorvem luz **UV** ao redor de 260 nm. Quando o duplexo de **DNA** é desnaturado à absorbância de 260 nm, a absorção de luz **UV** aumenta à medida que a temperatura também aumenta. Ao atingir a separação total das fitas, a absorção máxima pelo duplexo também é atingido. Isso é devido às bases nitrogenadas estarem expostas ao meio podendo absorver o máximo de luz possível. Nessa situação, a luz **UV** é absorvida 37% mais do que quando o **DNA** está no estado de dupla

hélice [17]. O ponto de inflexão da curva desses perfis sinaliza a temperatura de *melting* ( $T_m$ ) — ou temperatura de desnaturação —, na qual 50% das fitas estão ligadas e 50% se encontram separadas [20, 22, 25–28], veja a Figura 11. Em um ambiente específico, a temperatura de *melting* é uma propriedade física dos ácidos nucleicos que dá informações sobre a estabilidade do duplexo de DNA [25, 29]. Valores de temperatura de *melting* podem ser utilizados para o aprimoramento de modelos computacionais relacionados à estabilidade e energias de hibridização de ácidos nucleicos [29]. São também parâmetros importantes para o design experimental de aplicações como sondas de nucleotídeo [25], RNA silenciado [25], polimorfismo de nucleotídeo único (*single nucleotide polymorphism*, SNP) [25, 26] e primers para PCR [25, 30].



**Figura 11**

Perfil de temperatura de *melting* (em °C) pela absorção de luz UV a 260 nm de uma fita dupla. O princípio da curva mostra a fita ainda ligada. Ao elevar a temperatura, no final da curva, a fita se dissocia e separa totalmente. As duas setas no meio da curva mostram o processo de dissociação das fitas e indicam que ele é reversível. Figura retirada da referência [30].

A temperatura de *melting* é um parâmetro utilizado em experimentos à base de desnaturação. Alterando termicamente o ambiente de uma amostra, observa-se mudança nas propriedades de absorbância refletindo em mudança conformacional da molécula em solução e também possibilita a determinação da estabilidade estrutural de DNA ou RNA [30]. Essa alteração térmica propicia o estudo de hibridização entre oligonucleotídeo e alvo, o que pode ser empregado em design de primer para PCR, sequenciamento e hibridização [30].

Assim como as fitas de DNA podem ser separadas elevando-se a temperatura do ambiente onde se encontra o DNA, elas também podem ser conectadas novamente por meio do processo conhecido como "renaturação" (Figura 10). Nesse processo, a temperatura é diminuída paulatinamente promovendo um ambiente que propicie a interação entre os pares de bases resultando no reanelamento das fitas [22, 30]. Também é um processo utilizado em técnicas analíticas contribuindo com informações para estudos genéticos.

## 5 Objetivos

### 5.1 Objetivo Geral

Modelar os marcadores fluorescentes Cy3 e Cy5 atados à molécula de **DNA** para modelos mesoscópicos e verificar se modelos mesoscópicos podem ser bons preditores para intensidades de fluorescência.

### 5.2 Objetivos Específicos

1. Modelar os marcadores cianina Cy3 e Cy5 ligados à molécula de **DNA** a partir de dados de temperatura de *melting*.
2. Parametrizar os marcadores fluorescentes Cy3 e Cy5 para modelos mesoscópicos.
3. Correlacionar intensidades de fluorescência de dados de microarranjo com aberturas médias dos pares de bases.
4. Verificar se o modelo mesoscópico Peyrard-Bishop (**PB**) é um bom preditor para intensidades de fluorescência.

## 6 Marcadores Fluorescentes

Marcadores fluorescentes (ou fluoróforos) têm sido amplamente utilizados em diversas técnicas aplicadas nas áreas de biotecnologia, biomedicina, farmacologia e biofísica [31–33]. Experimentos de microarranjo de DNA [32, 34], hibridização fluorescente *in situ* (FISH) [34], transferência de energia por ressonância Förster (FRET) [32, 33, 35–39], reação em cadeia da polimerase (PCR) [33–35] e PIFE (*protein-induced fluorescence enhancement*) [40–43] são alguns dos métodos que empregam esses marcadores. Esse tipo de marcador responde à pH, viscosidade, polaridade, elasticidade e polarizabilidade [44]. Dentre diversos tipos de marcadores fluorescentes, os mais comumente usados são os marcadores cianina [35], conhecidos também como “*cy dyes*”.

### 6.1 Marcadores Cianina

#### 6.1.1 Características

Os marcadores cianina são corantes sintéticos [10] (Figura 12) que se tornaram relevantes por serem fotossensíveis [33], pela alta afinidade de ligação com ácidos nucleicos [33], possuírem bandas de absorção estreita e coeficientes de extinção altos [33, 45–47], serem resistentes ao fotobranqueamento (*photobleaching*) [45] e tornarem um oligômero de DNA altamente fluorescente [45]. Eles aumentam a temperatura de desnaturação independentemente da sequência do oligômero ao qual estejam ligados [45].



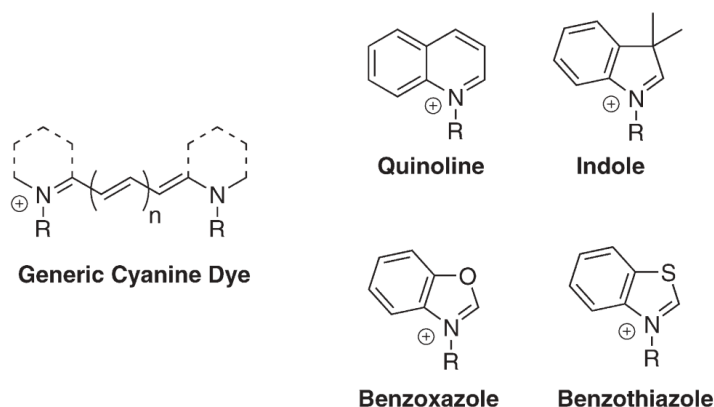
**Figura 12**

Há uma grande variedade de marcadores cianina, como Cy2, Cy3 e Cy7. Cada cianina fluoresce em um trecho do espectro resultando em diferentes cores. Figura retirada de [48].

A estrutura química básica desses marcadores é formada por dois átomos de nitrogênio conectados por uma cadeia polimetina, a qual é constituída de um número ímpar de átomos de carbono [10, 31, 39, 43, 46, 49, 50]. Para que a cianina seja estável, é necessário que seja ligado a cada um dos seus nitrogênios um grupo heterocíclico, como indol, quinolina, benzoxazol ou benzotiazol [10, 50], veja a Figura 13. A presença desses grupos heterocíclicos define a classificação da cianina nas seguintes categorias: marcadores cianina metino, marcadores merocianina, marcadores apocianina e marcadores cianina squarylium [31].

As cianinas são conhecidas também por agregarem em solução aquosa e são fracamente fluorescentes nesse tipo de solução [10, 39]. O tamanho da cadeia polimetina influi na variação dos espectros de absorção e de emissão [36] e elas também se tornam mais estáveis quando





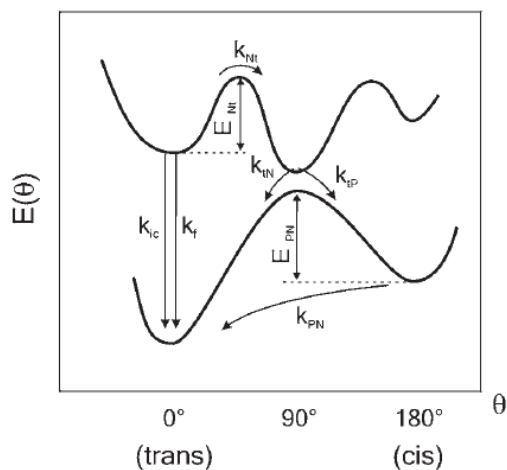
**Figura 13**

Estrutura da cianina. À esquerda, a estrutura genérica dos marcadores cianina com os dois átomos de nitrogênio e a cadeia polimetina. À direita, alguns grupos heterocíclicos que podem ser ligados aos nitrogênios das cianinas. Figura retirada da referência [39].

ligadas a um terminal de DNA [36]. Há três configurações possíveis das cianinas se ligarem a uma sequência de DNA: em uma das pontas, em ambas as pontas, e no meio da sequência [10, 44]. As cianinas se ligam à molécula de DNA por meio de um *linker*, por exemplo, de etinil, amida ou alceno [51]. Na Figura 13, os R's na estrutura genérica das cianinas representam as posições possíveis para acoplar o *linker*. Ele pode ser de vários tamanhos, rígido ou flexível, o que influencia na interação do marcador com a molécula e, conseqüentemente, os resultados obtidos em experimentos como microarranjo, por exemplo. Estudos de ressonância magnética nuclear (*nuclear magnetic resonance*, NMR) mostraram também que a cianina acoplada ao terminal de uma molécula de DNA se comporta como um par de bases adicional [37, 52–54].

Os marcadores cianina podem assumir outras conformações — isômeros — rotacionando ao redor das ligações carbono-carbono da cadeia polimetina [10]. A flexibilidade da cadeia proporciona à molécula transitar da conformação *trans* para a *cis* levando à competição da isomerização com a fluorescência [35], o que torna as cianinas vulneráveis à isomerização *cis-trans*. Quando a molécula é excitada e ocorre a isomerização, há perda da fluorescência, em particular, à altas temperaturas [55]. A isomerização das cianinas inicia-se no primeiro estado excitado, em que se encontram na conformação *trans*, e passam para o estado fundamental, para a conformação *cis*, ao receber energia, chamada de energia de isomerização (ou de ativação)  $\Delta E_{iso}$  [10, 56]. Entre essas duas conformações, a molécula excitada passa rapidamente por uma conformação *twisted* que retorna para o estado fundamental para a conformação *cis* ou para a conformação *trans* [10, 40] (Figura 14).

Alguns fatores influenciam na isomerização das cianinas, tais como viscosidade do meio, temperatura e substituintes (grupo heterocíclico, ácidos nucleicos, proteínas, etc) [10]. Em ambientes mais viscosos, a isomerização se torna mais limitada e a fluorescência aumenta pela liberação de mais energia devido aos fótons absorvidos. Ao se ligarem à DNA ou RNA, a fluorescência das cianinas cresce significativamente [57], principalmente, devido à sequência de bases a qual se liga [35]. Além disso, para as cianinas, o ambiente molecular em que elas se encontram influencia fortemente a sua fluorescência [42].

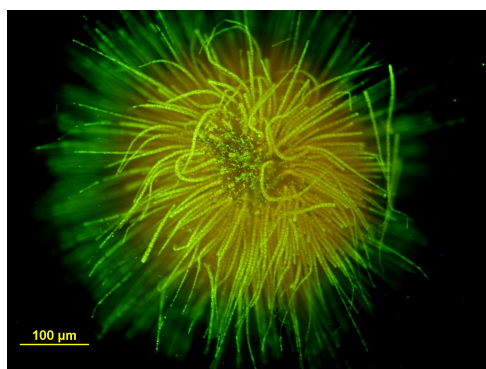


**Figura 14**

Energias do estado fundamental e do primeiro estado excitado singlete pelo ângulo de torção ( $\theta$ ). A fluorescência ( $k_f$ ) ocorre na conformação *trans*. A conformação *twisted* pode retornar para a *cis* ou para a *trans*. A cianina pode retornar à conformação *trans* no estado fundamental direto da conformação *cis* também no estado fundamental. Figura retirada da referência [10].

### 6.1.2 Aplicações

As cianinas têm sido muito usadas como sondas fluorescentes aplicadas em biologia molecular, bioquímica e biofísica [33, 58]. Além de poderem ser aplicadas em variadas áreas da ciência, como engenharia tecnológica, medicina, farmacologia, entre outras. Na área química, são usadas como indicadores ácido-base em química analítica. Em medicina, como anti-tumor e/ou anti-câncer. Em farmacologia, como agentes fungicidas e bactericidas (Figura 15). Em biologia, são utilizadas como reagentes fluorescentes para a detecção de ácidos nucleicos e proteínas [32, 33] e marcação de biomoléculas [31, 34, 59]. Também são empregadas como reagentes em biomedicina. Em tecnologia, são empregadas em células solares orgânicas, discos óticos como CD e DVD e tintas de impressão [31].



**Figura 15**

Microorganismos também podem ser marcados com corantes fluorescentes. Bactéria marcada com o fluoróforo cianina. Figura retirada de [60].

## 6.2 Cianinas Metino: Cy3 e Cy5

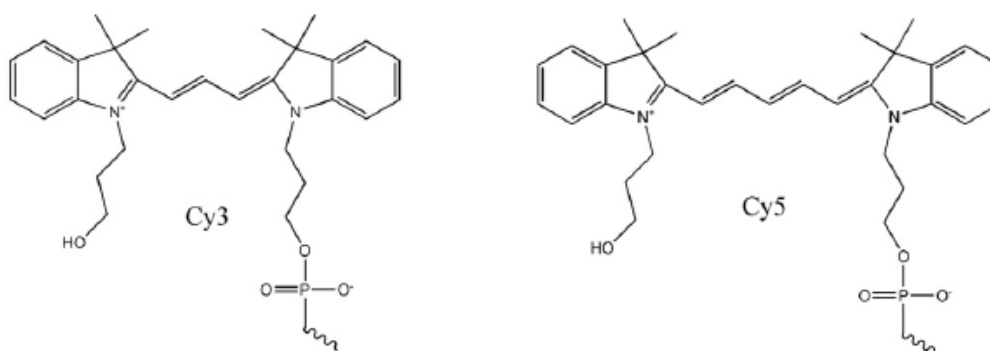
### 6.2.1 Características

As cianinas metino são um grupo de marcadores caracterizados pela variação da cadeia polimetina e a presença do grupo heterocíclico indol em sua estrutura. O número de carbonos dessa cadeia diferencia as cianinas desse grupo. Em técnicas de estudos de biomoléculas, as cianinas metino mais comumente aplicadas são as cianinas Cy3 [1,1'-bis(3-hydroxypropyl)-3,3,3',3'-



tetramethylindocarbocyanine] e Cy5 [1,1'-bis(3-hydroxypropyl)-3,3,3',3'-tetramethylindodicarbocyanine] [35, 58]. A estrutura química desses marcadores se constitui da cadeia polimetina com seus átomos de nitrogênio ligados a dois anéis indol [44, 45, 61], como na Figura 16. No primeiro fluoróforo, encontram-se três átomos de carbono na cadeia polimetina, e no segundo, cinco átomos.

Cy3 e Cy5 são os marcadores fluorescentes mais usados nos estudos da estrutura e dinâmica de biomoléculas através de transferência de energia por ressonância Förster (FRET), sendo aplicado, mais comumente, Cy3 como o doador e Cy5 como o acceptor [36, 37, 39, 43, 52, 53, 56, 62–64]. A absorção e a emissão máximas de Cy3 são 550 e 570 nm, e para Cy5, 650 e 670 nm, respectivamente [36]. Além de serem os marcadores fluorescentes mais brilhantes, ambos possuem maior fotoestabilidade [10, 36, 42, 42, 61], baixa sensibilidade à pH [36], eficiência quântica mais alta quando ligados à DNA [35, 56] e são sensíveis ao ambiente sequencial em que se encontram [34, 35, 41, 44]. As cianinas podem ser atadas ao terminal 5' de fita simples ou fita dupla de DNA [34].



**Figura 16**

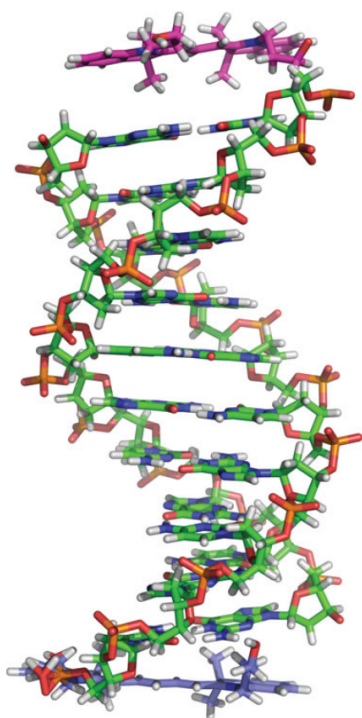
Estruturas químicas de Cy3 (à esquerda) e Cy5 (à direita). A cadeia polimetina com três átomos de carbono para Cy3 e cinco átomos para Cy5, dois anéis indol, dois átomos de nitrogênio e um grupo fosfato representando o início de um *linker*. Figura retirada da referência [45].

Informações sobre Cy5 na literatura ainda são restritas, porém Cy3 tem tido uma grande atenção contribuindo com mais detalhes. Por exemplo, é conhecido que Cy3 possui uma estabilidade marcante contra *photobleaching* [58, 59, 65], insensibilidade à pH [65], compatibilidade com lasers verdes detectores de fotodiodo de avalanche e é disponibilizado comercialmente [66]. Cy3 é o marcador fluorescente mais empregado em hibridização fluorescente *in situ* (FISH), possuindo um coeficiente de absorção alto e uma eficiência quântica também alta [42, 65].

### 6.2.2 Interação de Cy3 e Cy5 com DNA

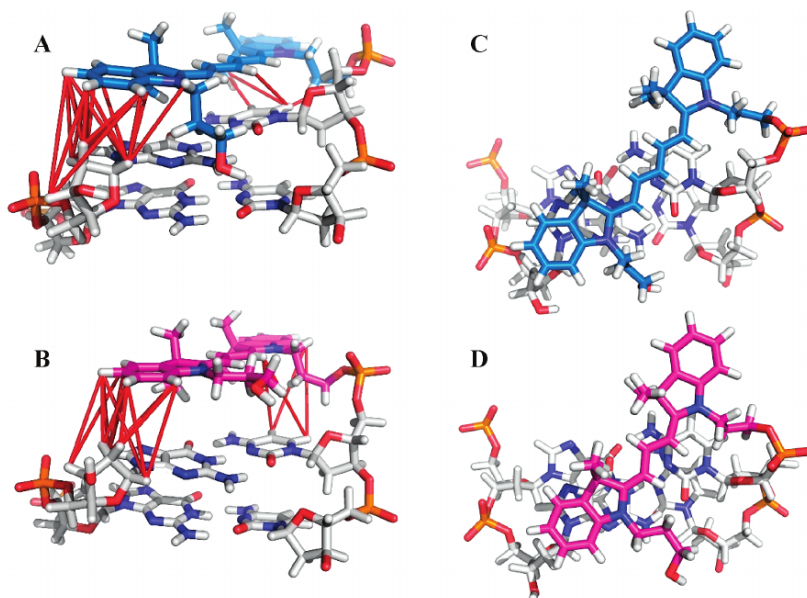
Estudos de NMR realizados por Iqbal e colaboradores [53] mostraram que tanto Cy3 quanto Cy5, quando acoplados ao terminal 5' do duplexo de DNA, empilham ao final da hélice como um par de bases adicional [34, 44, 53, 54, 56, 62], como representado na Figura 17. Esse mesmo

comportamento foi observado por Urnavicius et al. [62], através de NMR e FRET, nos casos em que os marcadores foram atados à molécula por um *linker* curto e por um *linker* longo. O comportamento similar a um par de bases se deve à cianina assumir uma configuração planar, assim como um par de bases, quando ligada ao DNA nessa posição [44], veja a Figura 18.



**Figura 17**

Imagem de NMR de uma fita dupla de DNA com Cy3 (roxo) acoplado no topo e Cy5 (azul) acoplado embaixo. Observa-se uma configuração extremamente similar a um par de bases adicional. Figura retirada da referência [10].



**Figura 18**

Cy5 (A e C) e Cy3 (B e D) acoplados ao terminal de um duplexo de DNA. Em A e B, as cianinas em uma configuração planar empilhadas sobre os pares de bases canônicos. As linhas vermelhas mostram interações de NOE. Em C e D, uma visão superior desse empilhamento, onde se vê o comportamento semelhante a um par de bases. Figura retirada da referência [53].

Quando Cy3 encontra-se ligado ao terminal 5' de uma molécula de DNA, ele forma com o par de bases terminal um complexo de empilhamento- $\pi$  [34, 39, 49]. A hidrofobicidade e a polarizabilidade das cianinas favorecem para a ocorrência desse complexo [39]. As interações de empilhamento- $\pi$  tendem a restringir a isomerização *cis-trans* elevando a fluorescência [34,

[44]. Essas interações dificultam a rotação da cadeia polimetina, o que levaria à mudança de conformação [34]. Por exemplo, a interação de empilhamento- $\pi$  de Cy3 com purinas aumenta a fluorescência da molécula [35]. Além disso, quando Cy3 e Cy5 se ligam a uma biomolécula, como uma proteína ou uma molécula de DNA, há a inibição da isomerização levando ao aumento do tempo de vida e da eficiência quântica da fluorescência [10, 44]. No caso do DNA, isso se deve ao comportamento das cianinas acoplarem ao duplexo como um par de bases [35, 44], Figura 18.

Em fitas simples de DNA, foi observado também que a intensidade de fluorescência está correlacionada à presença de purinas na sequência [44]. Quando há um grande número de purinas, a fluorescência se eleva, enquanto na presença de pirimidinas, a intensidade diminui, principalmente, devido ao número de citosinas [35, 44]. Estudos de Agbavwe e Somoza [35] encontraram que Cy3 mostra uma maior intensidade de fluorescência quando ligado à guanina, enquanto que ligado à timina a intensidade é menor. O mesmo estudo mostrou que, quando ligado à guanina seguida de várias adeninas, Cy5 fluoresce com maior intensidade. Quando o marcador cianina está ligado a uma pirimidina, o duplexo de DNA tende a estabilizar mais do que quando a cianina está ligada a uma purina [45]. A exceção existe para Cy3 atado à timina que estabiliza o duplexo menos do que o esperado [45]. Spiriti et al. [56] mostraram que a interação entre Cy3 e o par de bases AT é mais estável que Cy3 interagindo com o par de bases TA.

Kretschy et al. [44] mostraram que a fluorescência de Cy3 e Cy5 é dependente do ambiente sequencial e que a intensidade de fluorescência em relação à qual base nitrogenada os marcadores estejam ligados segue a ordem  $dG > dA > dT > dC$  [35, 44, 58]. A fluorescência de Cy3 e Cy5 é realçada quando eles se encontram ligados a uma base nitrogenada do que quando estão livres em solução e é maior quando os marcadores estão atados a uma fita simples de DNA do que uma fita dupla [10, 49, 66].

A literatura tem sugerido que Cy3 e Cy5 tendem a estabilizar a molécula de DNA [10, 45]. Quando estão acoplados por um *linker* rígido não perturbam o duplexo de DNA [10, 51, 67]. Se o *linker* for flexível, a estabilidade tende a ocorrer [51]. A estabilidade também é influenciada pela interação entre o marcador e a base à qual está ligado [45]. Além disso, Cy3 e Cy5 dependem fortemente da sequência de DNA ao qual se encontram atados [44, 56, 58]. Segundo Morozov et al. [67], o duplexo de DNA tende a desestabilizar quando as cianinas estão presentes no interior da molécula de DNA.

A interação de Cy3 com o duplexo de DNA é dependente da presença de purinas e a flexibilidade do oligômero [58]. Possui menor mobilidade quando ligado a um ácido nucleico e se encontra num estado excitado mais alto e com maior tempo de vida [56]. Quando ligado covalentemente ao DNA, as propriedades fotofísicas de Cy3 dependem do tipo do *linker* usado, da sequência a qual o marcador está ligado e da estrutura secundária do DNA [10]. O tipo do *linker* também influencia o tempo de vida e a eficiência quântica para Cy3 quando atado à molécula de DNA [10]. A eficiência de fluorescência de Cy3 é influenciada pela temperatura e

pela sequência do oligômero ao qual está acoplado [10, 40, 44, 58].

### 6.2.3 Aplicações de Cy3 e Cy5

As características conhecidas sobre Cy3 e Cy5 levaram esses marcadores a serem empregados em lasers [10], espectroscopia de molécula única, ótica não linear [36], reagentes fluorescentes para proteínas e ácidos nucleicos [59] e outras aplicações. Ambos são os mais usados em técnicas como microarranjo de DNA e RNA [44], hibridização fluorescente *in situ* (FISH) [44, 65], transferência de energia por ressonância Förster (FRET) [10, 35–38, 44], reação em cadeia da polimerase (PCR) [34, 35, 44] e PIFE (*protein-induced fluorescence enhancement*) [40, 41, 41–43]. Além disso, Cy3 também tem sido usado em técnicas como imagem de tráfico intracelular em células vivas, interações proteína-DNA e proteína-proteína e estrutura de DNA [66].

## 6.3 Conclusão

Neste capítulo, revisamos as principais características dos marcadores fluorescentes, em especial, as cianinas Cy3 e Cy5. Discutimos também como estes marcadores interagem com DNA, o que será importante para entender o resultados que apresentaremos nos capítulos 8 e 9.

## 7 Métodos

Neste capítulo, descrevemos os principais modelos teóricos usados nesta dissertação. Na seção que segue, apresentamos o modelo de desnaturação de dois estados que é usado nos trabalhos experimentais para extrair temperaturas de desnaturação. Na seção 7.2, falamos do modelo mesoscópico usado para descrever as aberturas de cianinas ligadas à DNA. O método de otimização de parâmetros é descrito na seção 7.3 e é usado nas análises dos capítulos 8 e 9. Por fim, discutimos os principais aspectos da correlação de Pearson na seção 7.4 que será usada no capítulo 9.

### 7.1 Modelo de desnaturação em dois estados

Sabe-se que o duplexo de DNA pode-se dissociar e hibridizar novamente, veja a Figura 10 na página 15. O modelo de desnaturação em dois estados considera que ou as fitas em solução se encontram completamente separadas ou encontram-se totalmente hibridizadas. Isto é justificado considerando que a transição entre a hibridização completa e separação completa se dá em um intervalo de temperatura muito pequeno. Considere então uma fita dupla de DNA em que uma fita é  $X$  e a outra  $Y$ . Dessa forma pode-se escrever a transição de desnaturação na solução como uma reação química:



Durante qualquer um dos dois processos — dissociação e hibridização — a solução de DNA está com fitas dissociadas e fitas ligadas. Essa concentração é chamada de concentração de espécie e pode ser representada como:

$$C_t = [X] + [Y] = [X] + [Y] + 2[X \cdot Y] \quad , \quad (3)$$

em que  $[X]$  e  $[Y]$  representam as concentrações de fitas dissociadas e  $[X \cdot Y]$  a concentração de fitas ligadas. Para caracterizar a concentração de espécie, tem-se a constante de equilíbrio  $K_{eq}$ :

$$K_{eq} = \frac{[X][Y]}{[X \cdot Y]} \quad (4)$$

Tomando a temperatura de *melting* em relação às concentrações dos estados da molécula de DNA, tem-se:

$$[X \cdot Y] = [X]_m = [Y]_m \quad (5)$$

Da relação (4), tem-se que  $K_{eq} = \frac{C_t}{4}$ , quando todas as sequências de DNA do conjunto tomado para o cálculo são sequências não-auto-complementares (as fitas, lidas de  $5' \rightarrow 3'$ , são distintas). Aplicando a constante de equilíbrio na equação para a energia livre de Gibbs ( $\Delta G$ ), tem-se:

$$\Delta G = -RT_m \ln K_{eq} = \Delta H - T_m \Delta S \quad , \quad (6)$$

onde  $R$  é a constante universal dos gases. Vê-se na equação (6) como a energia livre de Gibbs ( $\Delta G$ ), a variação de entalpia ( $\Delta H$ ), a variação de entropia ( $\Delta S$ ) e a temperatura de *melting* se relacionam. Rearranjando essa equação, pode-se encontrar a relação direta para a temperatura de *melting*:

$$T_m = \frac{\Delta H}{\Delta S - R \ln C_t/4} \quad (7)$$

Para sequências canônicas (pareamentos AT e CG somente), geralmente, tem-se  $\Delta H < 0$  e  $\Delta S < 0$ . A relação (7) é válida para quando tem-se sequências não-auto-complementares, como mencionado anteriormente. No caso de sequências auto-complementares — as duas fitas, lidas de  $5' \rightarrow 3'$ , são iguais —, tem-se  $K_{eq} = C_t$  e a relação de concentrações de fitas dissociadas e fitas ligadas:

$$[X \cdot X]_m = [X_m] \quad (8)$$

Dessa forma, a equação para temperatura de *melting* relacionando entalpia e entropia se modifica para:

$$T_m = \frac{\Delta H}{\Delta S - R \ln C_t} \quad (9)$$

Tratamos aqui o caso das sequências de **DNA** no processo de desnaturação. Quando trata-se do processo inverso — a renaturação (ou hibridização) — a constante de equilíbrio ( $K_{eq}$ ) é o inverso:

$$K_{desn} = \frac{1}{K_{hibr}} \quad , \quad (10)$$

onde  $K_{desn}$  é a constante de equilíbrio para o processo de desnaturação e  $K_{hibr}$  a constante de equilíbrio para o processo de hibridização. Com essa relação, a equação para a temperatura de *melting* (9) também se modifica levemente:

$$T_m = \frac{\Delta H}{\Delta S + R \ln C_t/4} \quad (11)$$

Para sequências canônicas, geralmente, tem-se  $\Delta H > 0$  e  $\Delta S > 0$ . A equação (11) é referente às sequências não-auto-complementares. A diferença para as sequências auto-complementares é semelhante à do processo de desnaturação resultando em:

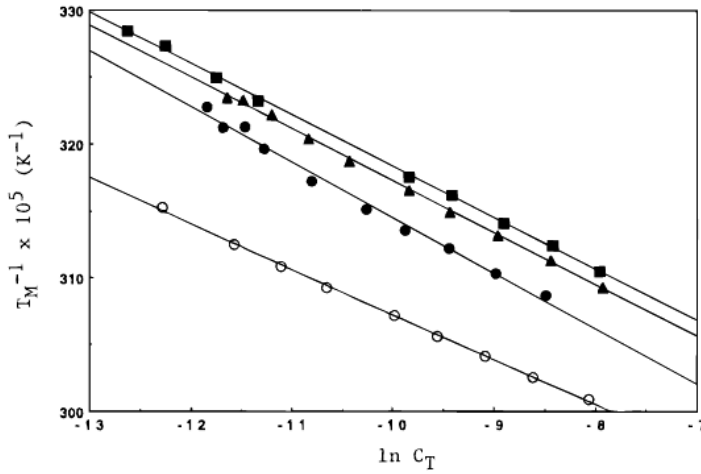
$$T_m = \frac{\Delta H}{\Delta S + R \ln C_t} \quad (12)$$

A temperatura de *melting* pode ser medida a partir de diferentes concentrações. Tomando a equação (11), pode-se reescrevê-la de forma a se assemelhar à equação da reta ( $x = a + by$ ):

$$T_m^{-1} = \frac{\Delta S}{\Delta H} - \frac{R}{\Delta H} \ln C_t/4 \quad , \quad (13)$$

onde, pela equação da reta,  $T_m^{-1}$  representa x e  $\ln C_t/4$  representa y. Na Figura 19, vê-se um gráfico para temperatura de *melting* ( $T_m^{-1}$ ) versus concentração ( $C_t$ ). Além de ser possível me-

dir a temperatura de *melting* tomando várias concentrações, ajustando os dados na equação (13), pode-se obter também  $\Delta S$  e  $\Delta H$  [68].

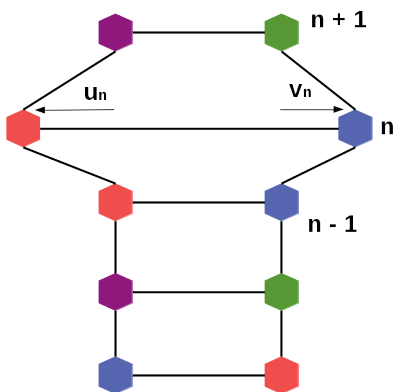


**Figura 19**

A temperatura de *melting*  $T_m^{-1}$  pode ser medida a partir de várias concentrações  $C_t$ . No gráfico, cada reta representa uma sequência distinta. Figura retirada de [68].

## 7.2 Modelo Mesoscópico Peyrard-Bishop (PB)

Em termos simples, mesoscópico é uma escala de comprimento entre o microscópico e o macroscópico. Em 1989, Peyrard e Bishop propuseram um modelo estatístico simplificado, que eleva a eficiência computacional, para descrever as propriedades físicas da dupla hélice de **DNA**, mais especificamente o processo de desnaturação [69, 70]. Outros modelos derivaram deste tais como o modelo Peyrard-Bishop-Dauxois e o Peyrard-Bishop-Holstein. O modelo **PB** toma a molécula de **DNA** como homogênea (pares de bases iguais) e a representa como uma estrutura bidimensional plana e cada base nitrogenada é considerada como uma partícula. Para cada base, é tomada uma massa  $m$  [69, 70]. O modelo **PB** também considera independentemente as ligações de hidrogênio entre as bases dos pares e a interação de empilhamento entre as bases vizinhas [69, 70], e permite prever temperaturas de desnaturação a partir das sequências dadas.

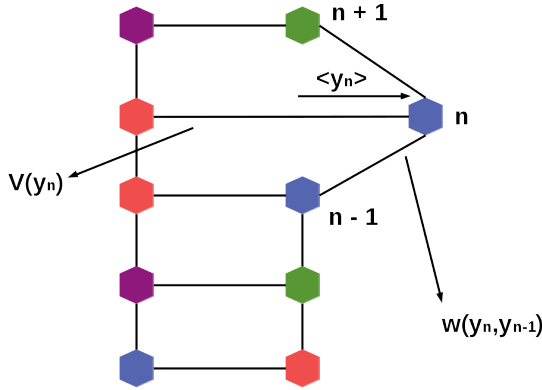


**Figura 20**

Representação dos graus de liberdade,  $u_n$  e  $v_n$ , pelo modelo Peyrard-Bishop (PB). No segundo par de bases da figura, observa-se os deslocamentos das bases ao longo da direção das ligações de hidrogênio a partir do ponto de equilíbrio.

Nesse modelo, os autores aplicaram o método de integral de transferência para determinar a separação entre as fitas da molécula de **DNA** [69, 70]. Eles consideram dois graus de liberdade





**Figura 21**

Representação do modelo PB. Vê-se a posição de uma das bases representada por  $n$ , o potencial de Morse  $V(y)$  referente às ligações de hidrogênio, o empilhamento das bases ao longo das fitas representado por  $w(y_n, y_{n-1})$  e o deslocamento  $\langle y_n \rangle$  das bases na direção das ligações de hidrogênio.

( $u_n$  e  $v_n$ ) para o par de bases, os quais representam os deslocamentos das bases, a partir do ponto de equilíbrio, ao longo da direção em que está a ligação de hidrogênio do par [69]. A representação dos graus de liberdade pode ser vista na Figura 20 e eles são representados por:

$$u_n = \frac{x_n + y_n}{\sqrt{2}}, \quad v_n = \frac{x_n - y_n}{\sqrt{2}} \quad (14)$$

Para o potencial das ligações de hidrogênio entre as bases, o modelo PB o aproxima ao potencial de Morse, representando duas ou três ligações [69], veja Figura 21. O potencial de Morse  $V(y)$  é dado por [71]:

$$V(y) = D \left( e^{-y/\lambda_a} - 1 \right)^2, \quad (15)$$

onde  $D$  é a energia de dissociação do par de bases e  $\lambda_a$  é o parâmetro que define o alcance do potencial [72]. Além da interação entre as bases dos pares, tem-se a interação de empilhamento resultante da relação entre os pares de bases vizinhas (Figura 21). Para esse empilhamento, considera-se um acoplamento harmônico. A constante elástica  $k$  é tomada como a mesma para todas as interações entre bases ao longo de toda a fita [69]:

$$w(y_n, y_{n-1}) = \frac{k}{2} (y_n - y_{n-1})^2 \quad (16)$$

Reunindo o potencial para as ligações de hidrogênio entre os pares de bases complementares e o acoplamento harmônico para a interação entre as bases vizinhas, tem-se a Hamiltoniana:

$$H = \sum_n \frac{1}{2} m [\dot{u}_n^2 + \dot{v}_n^2] + \frac{1}{2} k_u (u_n - u_{n-1})^2 + \frac{1}{2} k_v (v_n - v_{n-1})^2 + D \left[ e^{-\frac{a}{\sqrt{2}}(u_n - v_n)} - 1 \right]^2 \quad (17)$$

O primeiro termo da hamiltoniana representa a energia cinética da molécula de DNA (seu aspecto dinâmico), o segundo e terceiro termos descrevem a interação de empilhamento entre as bases vizinhas como osciladores harmônicos acoplados e, por fim, o quarto e último termo mostra o potencial de Morse que descreve a interação das ligações de hidrogênio entre as bases dos pares. Realizando uma transformação de coordenadas, pode-se descrever o movimento das



duas fitas do duplexo de **DNA** através das variáveis  $x_n$  e  $y_n$ , representadas pelas equações (18):

$$x_n = \frac{u_n + v_n}{\sqrt{2}}, \quad y_n = \frac{u_n - v_n}{\sqrt{2}}, \quad (18)$$

onde  $y_n$  representa apenas o deslocamento na direção das ligações de hidrogênio [69]. Com essas variáveis, obtém-se agora a hamiltoniana:

$$H = \sum_n \left[ \frac{1}{2m} p_n^2 + \frac{1}{2m} q_n^2 + w(x_n, x_{n-1}) + w(y_n, y_{n-1}) + V(y_n) \right], \quad (19)$$

onde  $p_n = m\dot{x}_n$  e  $q_n = m\dot{y}_n$  são os momentos para  $x_n$  e  $y_n$ . A partir dessa hamiltoniana (19), é possível obter a função clássica de partição, como na equação (20), para uma molécula de **DNA** contendo  $N$  pares de bases e calculada a uma temperatura fixa  $T$ :

$$Z = \int \prod_n e^{-\beta H} dp_n dq_n dx_n dy_n = Z_p Z_q Z_x Z_y, \quad (20)$$

onde  $\beta = 1/(k_B T)$ , sendo  $k_B$  a constante de Boltzmann. As integrais dos momentos,  $p_n$  e  $q_n$ , resultam em:

$$Z_p = Z_q = (2\pi m k_B T)^{N/2} \quad (21)$$

Caracterizando a contribuição da energia potencial em uma "fita harmônica",  $Z_x$ :

$$Z_x = \left( \frac{2\pi}{\beta k} \right)^{N/2} \quad (22)$$

Como o empilhamento leva em consideração apenas a interação entre bases vizinhas [69, 72, 73], tem-se para  $Z_y$ :

$$(23)$$

Do modelo **PB** também é possível obter a abertura média das fitas de **DNA** [69, 74] — distância média entre as bases — representada por  $\langle y_m \rangle$ . O valor da abertura média indica a desnaturação ou não da molécula de **DNA**. A partir dos cálculos de potencial de Morse e acoplamento harmônico, a abertura média das ligações de hidrogênio é encontrada:

$$\langle y_m \rangle = \frac{1}{Z} \int \prod_n y_m e^{-\beta H} dy_n = \frac{1}{Z_y} \int dy_1 \int dy_2 \dots \int dy_N y_m \prod_n e^{-\beta w(y_n, y_{n-1})} e^{-\beta V(y_n)} \quad (24)$$

### 7.2.1 Cálculo de temperaturas de *melting* com o modelo PB

No modelo original de 1989, os autores consideram a molécula de **DNA** homogênea — somente composta por pares de bases iguais [69]. Ou seja, os parâmetros de potencial de Morse e de empilhamento das bases são similares para todas as bases de ambas as fitas. Mais tarde, revisões desse modelo levaram a reformulações que consideram também a heterogeneidade da

molécula de DNA [75]. Em nosso trabalho, tomamos a versão do modelo para a heterogeneidade do DNA, em que são considerados pares de bases distintos (AT e CG).

O modelo PB considera a equivalência térmica das sequências de DNA utilizando um índice  $\tau$  que agrupa as sequências por temperaturas de *melting* similares [76]. Índices similares indicam temperaturas de *melting* similares. Ajustando os índices para as temperaturas de desnaturação experimentais, pode-se calcular novas temperaturas por meio de uma regressão linear dos índices. Dessa forma, o modelo refina a parametrização e pode-se prever temperaturas de desnaturação a partir de dados experimentais.

A função de partição clássica  $Z$  é calculada uma única vez no modelo PB e então expandida em termos não-diagonais da matriz integral de transferência. O maior termo dessa expansão é o índice  $\tau_i$  para cada sequência  $i$ . A expansão da função partição pode ser descrita da seguinte forma:

$$Z = Tr(C^{(1,2)}C^{(2,3)}...C^{(N,1)}) \quad , \quad (25)$$

onde  $C^{(n,n+1)}$  representa a matriz de interação entre os pares de bases vizinhas e  $C^{(N,1)}$  a condição de contorno (o primeiro e o último par de bases ligados) [76]. A condição de contorno considera duas situações: uma em que a sequência de DNA é tomada como um anel, pois considera-se que o primeiro e o último par estão ligados, e outra em que não considera-se essa última ligação.

Tomando as interações (ligações de hidrogênio) dos pares de bases AT e CG como fraca (w) e forte (s), respectivamente, tem-se quatro tipos de vizinhos e as matrizes correspondentes  $C^{(w,w)}$ ,  $C^{(w,s)}$ ,  $C^{(s,w)}$  e  $C^{(s,s)}$  [76]. Usando como base ortonormal as matrizes referentes ao par CG, a matriz  $C^{(s,s)}$  converte para uma matriz diagonal  $\Lambda$  contendo os autovalores  $\lambda_i$  referentes à sequência homogênea de CG. Dessa forma, é definida a matriz  $\Delta^{(a,b)}$  que representa a diferença da interação entre bases vizinhas do tipo (a,b) e bases vizinhas do tipo (s,s):

$$C^{(a,b)} = \Lambda + \Delta^{(a,b)} \quad (26)$$

A partir dessa relação, a função de partição pode ser reescrita da seguinte forma:

$$Z = Tr[(\Lambda + \Delta^{(1,2)})(\Lambda + \Delta^{(2,3)})...(\Lambda + \Delta^{(N,1)})] \quad (27)$$

Utilizando propriedades de traço, a função partição se torna:

$$Z = \sum_{\omega=0}^N Z_{\omega}(\Lambda) = \sum_{\omega=0}^N Tr[M(\Lambda^{\omega})] \quad , \quad (28)$$

onde  $M(\Lambda^{\omega})$  são todos os termos que contêm  $\omega$  multiplicações da matriz  $\Lambda$ . Por tomarmos a sequência homogênea CG, tem-se  $\Delta^{(s,s)} = 0$  e, assim, somente os termos de  $\Lambda^N$ .

Na Figura 22, estão representadas quatro sequências formadas por 10 pares de bases, com conteúdo CG entre 40% e 60%, e distribuição normal para cada sequência, calculadas a 370

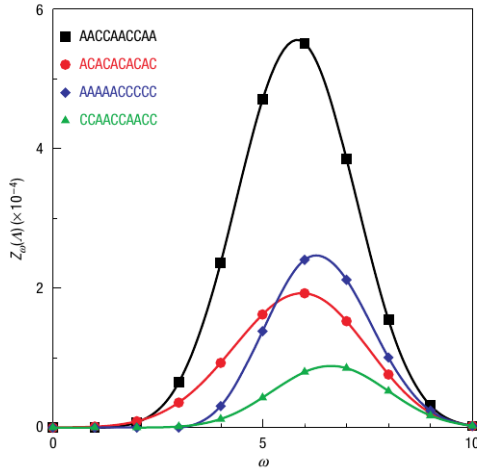
K. O ponto máximo de cada curva,  $\omega_{max}$ , é dependente da composição da sequência e não possui uma dependência forte em relação à temperatura (Figura 23). Na Figura 24, temos o exemplo de sequências de diferentes tamanhos ajustadas por regressão linear para cada grupo de mesmo tamanho mostrando a relação entre o parâmetro  $(\omega_{max})^{1/2}$  e a temperatura de *melting* experimental. Dessa relação, tem-se que  $(\omega_{max})^{1/2}$  pode ser usado como um valor adimensional que represente a equivalência térmica — índice  $\tau$ . Além disso, observa-se que a regressão é dependente de  $N^{1/2}$ .

Como os índices  $\tau_i$  são correlacionados às temperaturas de desnaturação experimentais, como visto na Figura 23, a regressão linear é executada para cada grupo de sequências de mesmo tamanho:

$$T'_i = a_0(N) + a_1(N)\tau_i \quad , \quad (29)$$

onde  $T'_i$  é a temperatura predita para a sequência  $i$ ,  $N$  é o tamanho da sequência e  $a_0(N)$  e  $a_1(N)$  são os coeficientes da regressão linear para cada grupo  $N$ . Como há a dependência com  $N^{1/2}$ , quando se procura a temperatura de uma sequência de tamanho diferente, o qual não tem referência no conjunto de dados utilizado, pode-se realizar uma nova regressão linear:

$$a_k(N) = b_{0,k} + b_{1,k}N^{1/2} \quad (30)$$



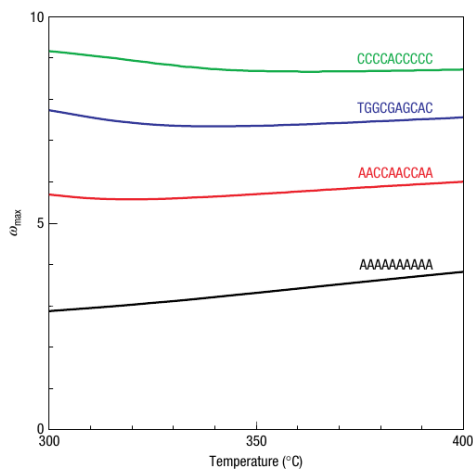
**Figura 22**

Distribuição normal para quatro diferentes sequências compostas por 10 pares de bases. O conteúdo CG de cada uma está entre 40% e 60% e foram calculadas à temperatura de 370 K. Figura retirada da referência [76].

### 7.3 Otimização de Parâmetros

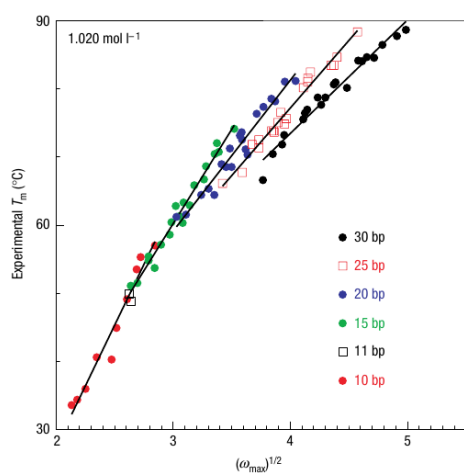
Os parâmetros de potencial de Morse e potencial harmônico são variados até serem minimizados. Para obtermos resultados mais apurados, a minimização é realizada várias vezes. A minimização é calculada pela diferença entre as temperaturas de *melting* obtidas experimentalmente e as temperaturas preditas aplicando-se [77–83]:

$$\chi^2 = \sum_{i=1}^N (T'_i - T_i)^2 \quad , \quad (31)$$



**Figura 23**

Relação do parâmetro  $\omega_{max}$  pela temperatura. Quatro sequências distintas compostas por 10 pares de bases. Figura retirada da referência [76].



**Figura 24**

Parâmetro  $(\omega_{max})^{1/2}$  pela temperatura experimental  $T_m$ . As sequências possuem de 10-30 pares de bases. As linhas representam a regressão linear para cada grupo de sequências de mesmo tamanho, exceto para os dois pontos de dados para sequências de 11 pares de bases. Figura retirada da referência [76].

onde  $T_i$  são as temperaturas de *melting* experimentais,  $T'_i$  as temperaturas de *melting* preditas e  $N$  é o total de sequências minimizadas. O desvio médio é representado por (32). Dessa minimização, quanto menor o valor encontrado, melhor a relação entre as temperaturas preditas e as temperaturas experimentais.

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T'_i - T_i| \quad (32)$$

A otimização dos parâmetros é realizada pelo método *downhill* simplex, também conhecido como método Amoeba ou método Nelder-Mead [84]. Esse método foi descrito por John Nelder e Roger Mead em 1965 [85]. Ele é utilizado para encontrar, em um espaço multidimensional, o valor mínimo ou máximo de uma função e consiste em realizar a minimização (ou maximização) dos parâmetros apenas avaliando a função requerida, sem considerar suas derivadas [84, 85].

O cálculo inicia-se por um simplex. Um simplex é uma figura geométrica formada por  $N$  dimensões e  $N + 1$  vértices [84, 86]. Sua forma é nomeada de acordo com a quantidade de dimensões. Por exemplo, em duas dimensões, um simplex é conhecido como um triângulo; em três dimensões, é um tetraedro. Para esse simplex inicial, é necessário fornecer um vetor de  $N$  variáveis independentes - "*starting guess*" - para o algoritmo [84, 86]. O algoritmo irá considerar esse "*starting guess*" como o primeiro ponto para o cálculo. A partir desse ponto, o simplex inicial é construído por:

$$P_i = P_0 + \lambda e_i \quad , \quad (33)$$

onde  $e_i$  são os vetores de tamanho  $N$  e  $\lambda$  é uma constante. Com o simplex inicial estabelecido, segue-se uma série de etapas para o cálculo da minimização, apresentadas na Figura 25. Primeiramente, o algoritmo reconhece qual dos três pontos do simplex é o melhor (o de menor valor), o intermediário e o pior (o de maior valor). Definida essa ordenação, calcula-se o centróide ( $x_0$ ) entre os pontos melhor e intermediário. A próxima etapa é chamada de reflexão. Nessa etapa, o algoritmo reflete o pior ponto, através do centróide, em prol de se obter um simplex de menor volume:

$$x_r = x_0 + \alpha(x_0 - x_{n-1}) \quad , \quad (34)$$

onde  $\alpha$  é uma constante positiva, chamada de coeficiente de reflexão. Se o ponto refletido estiver entre o melhor e o intermediário, então substitui-se o pior ponto pelo ponto refletido e reinicia-se o cálculo para um novo simplex. Caso contrário, segue-se para a próxima etapa chamada de expansão [84–86]. A expansão é uma forma de acelerar a redução do simplex em direção a um simplex de menor tamanho. Quando o ponto refletido é menor que o melhor ponto, calcula-se a expansão:

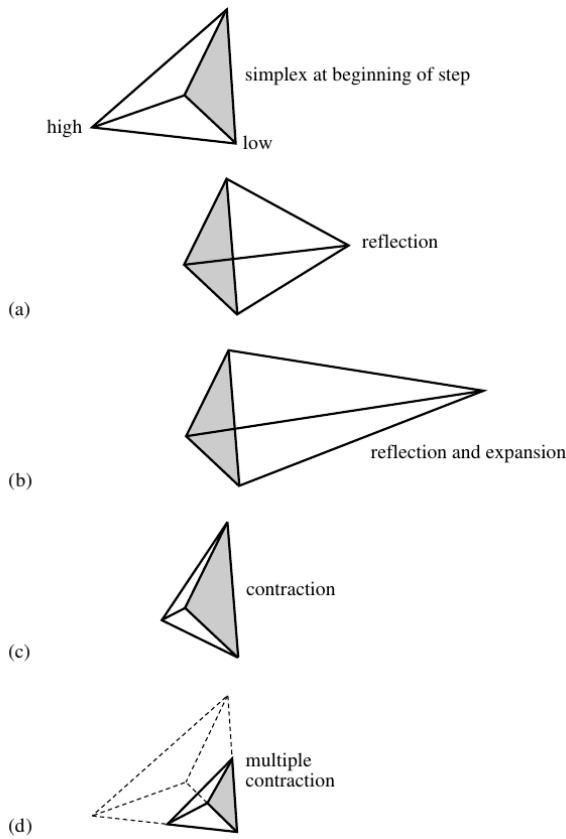
$$x_e = x_0 + \gamma(x_r - x_0) \quad , \quad (35)$$

onde  $\gamma$  é o coeficiente de expansão. Se o ponto de expansão for menor que o ponto refle-

tido, substitui-se o pior ponto pelo ponto de expansão e inicia-se novamente o cálculo. Caso contrário, o pior ponto é substituído pelo ponto refletido. Quando o valor do ponto refletido é maior ou igual ao ponto intermediário, realiza-se a etapa chamada de contração, para manter o simplex pequeno:

$$x_c = x_0 + \beta(x_{pior} - x_0) \quad , \quad (36)$$

onde  $\beta$  é o coeficiente de contração entre 0 e 1. Na contração, o simplex é contraído de forma a obter um volume menor. Se o ponto de contração for menor que o pior ponto, este é substituído pelo ponto de contração e inicia-se, novamente, o cálculo do simplex. Se nenhuma dessas etapas resultarem no menor simplex possível, mantém-se o melhor ponto e substitui-se os outros pontos reiniciando o cálculo. Os coeficientes  $\alpha$ ,  $\gamma$  e  $\beta$ , referentes a reflexão, expansão e contração, respectivamente, são usados como parâmetros para a mudança no volume do simplex no cálculo de cada etapa [85, 86].



**Figura 25**

No topo da figura, observa-se o simplex inicial e os pontos melhor e pior representados por *low* e *high*, respectivamente. Em (a), a etapa de reflexão em que o pior ponto é refletido através do centróide. Em (b), vê-se o simplex refletido e expandido (etapa de expansão). Em (c), o simplex sofreu a contração reduzido a um volume menor. Em (d), a etapa de compressão em que o melhor ponto é conservado e os pontos intermediário e pior substituídos. Figura retirada da referência [84].

## 7.4 Coeficiente de Correlação de Pearson

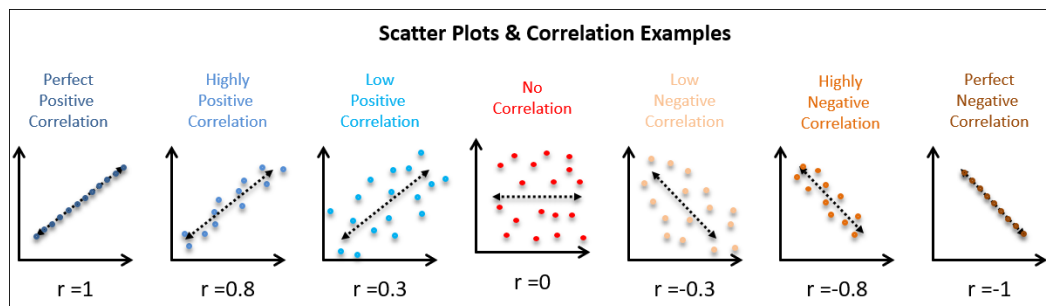
Correlação entre duas variáveis é definida como a quantificação do grau de relação entre essas variáveis. Essa correlação não define uma causalidade ou dependência entre elas. Considerando dois conjuntos de variáveis aleatórias  $X$  e  $Y$ , pode-se encontrar uma relação linear, ou não, entre ambos por meio de um coeficiente de correlação. Dos diversos coeficientes de

correlação existentes, o mais comumente aplicado é o coeficiente de correlação de Pearson ( $r$ ):

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad , \quad (37)$$

onde  $x_i$  e  $y_i$  são os valores provenientes dos dois conjuntos de variáveis aleatórias  $X$  e  $Y$  e os valores  $\bar{x}$  e  $\bar{y}$  são as médias desses valores. O cálculo da correlação toma o somatório desses dados obtendo o coeficiente de Pearson  $r$  que indica como as duas variáveis se relacionam linearmente, ou não.

O resultado da correlação indica a direção em que a relação entre as variáveis se define. O coeficiente de correlação é um valor restrito ao intervalo de -1 a +1 e sem unidade [87, 88]. Na Figura 26, tem-se as representações gráficas para o intervalo da correlação. Quando se obtém o coeficiente igual a +1, tem-se uma correlação positivamente perfeita que indica uma relação linear exata entre as variáveis. A mesma interpretação é válida para o valor -1, uma correlação negativamente perfeita para uma relação linear exata [87, 88]. Quando essa relação não é exata, há um deslocamento do coeficiente de correlação em direção a 0. Quanto mais próximo de 0, de ambos os lados, o deslocamento for, menor é a relação linear entre as variáveis indicando uma não-correlação entre elas. Porém não significa que não haja uma relação não-linear [88]. Obtendo valores entre 0 e +1, tem-se que as variáveis são positivamente correlacionadas. Assim como, valores entre -1 e 0 indicam variáveis negativamente correlacionadas.



**Figura 26**

Exemplos de coeficientes de correlação de Pearson desde a correlação positivamente perfeita ( $r = 1$ ), passando por correlações positivas alta e baixa, não correlação, correlações negativas alta e baixa, até a correlação negativamente perfeita ( $r = -1$ ). Figura retirada de [89].

## 8 Parametrização por Temperatura de *Melting*

Neste capítulo, será abordado o estudo sobre a modelagem dos marcadores Cy3 e Cy5. Apresentamos os dados experimentais que usamos para a parametrização; como adaptamos o modelo mesoscópico **PB** para podermos incluir os marcadores cianina; os procedimentos realizados para a otimização dos dados e concluímos com a discussão dos resultados obtidos.

### 8.1 Dados Experimentais

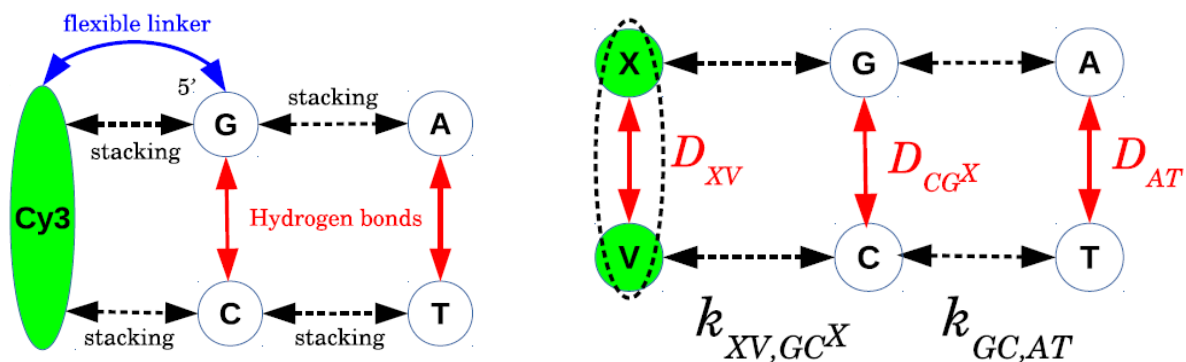
Os dados usados foram de Moreira et al. [45], um conjunto de temperaturas de *melting* obtidas por medidas de absorção **UV**, este tipo de medida foi revisado na seção 7.1, na página 25. Esse conjunto consiste de valores experimentais referentes à temperatura de *melting* ( $T_m$ ) de 35 sequências de **DNA** (tabelas 4 e 5 em A.1, página 63), modificadas com a presença de dois marcadores fluorescentes, Cy3 e Cy5. O conjunto consta de 19 sequências com Cy3 e 16 sequências com Cy5, sendo cada uma formada de 8 a 12 pares de bases. Os marcadores estão ligados ao terminal 5' da fita de **DNA** através de um *linker* flexível e, em alguns casos, nas duas fitas (nas pontas 5' ou 3'). Essas sequências de **DNA** são iguais tanto para Cy3 quanto para Cy5, diferindo apenas em três sequências extras para Cy3. Para ambos os marcadores, quatro sequências são auto-complementares. O conjunto de sequências satisfaz as dezesseis possibilidades de próximos vizinhos para os dois marcadores. A sequência a seguir faz parte desse conjunto de dados e ilustra o aspecto de todas elas:

Cy3 - ACGATCGT/TGCTAGCA - Cy3

O conjunto de dados de Cy3 possui quatro sequências auto-complementares (quatro primeiras sequências da Tabela 4). Para elas, em nosso trabalho, recalculamos as temperaturas de *melting* e atribuímos a concentração de espécie ( $C_t$ ) de  $1\mu\text{M}$ . Para todas as outras sequências, tanto de Cy3 quanto de Cy5, a concentração foi de  $2\mu\text{M}$ . A razão para esta diferença de concentração está na simetria natural das sequências auto-complementares que altera a entropia por um fator de  $R \ln(2)$  na equação (7). Então para poder comparar as propriedades térmicas de sequências auto-complementares e não-auto-complementares introduz-se um fator de dois entre as concentrações [90]. O erro experimental para as temperaturas de *melting*, reportado pelos autores, foi de  $0.5^\circ\text{C}$  e, para nossas análises descritas a seguir, usamos parâmetros à concentração salina de  $1020\text{ mM}$ .

Nós separamos as medidas de Moreira et al. [45] em dois conjuntos: um contendo as sequências de **DNA** com Cy3 e o outro com as sequências com Cy5. Ou seja, todos os cálculos e otimizações foram realizados independentemente sobre esses dois conjuntos.





**Figura 27**

Representação dos marcadores Cy3 e Cy5 nas sequências de DNA. À esquerda, Cy3 ligado ao primeiro par de bases pelo terminal 5' por um *linker* flexível mostrado pela dupla seta curvada (em azul). As duplas setas tracejadas (em preto) mostram o empilhamento das bases e as duplas setas contínuas (em vermelha), ao centro, as ligações de hidrogênio. À direita, a representação para o nosso trabalho. Cy3, nesse caso, é o par XV, o empilhamento é o potencial harmônico  $k$  e as ligações de hidrogênio são os potenciais de Morse  $D$ .

## 8.2 Adaptação do modelo PB

Estritamente falando, o modelo **PB** foi desenvolvido somente para pares de bases de DNA e não para marcadores fluorescentes, veja também a discussão na seção 7.2, página 27. No entanto, há evidência experimental de que os marcadores quando atados aos terminais se comportam como se fossem pares de bases adicionais [37, 52–54], veja também a discussão na seção 6.2.2, página 21. Assim, nós resolvemos testar a possibilidade em tratar os marcadores como se fossem pares de bases, o que na prática exige a atribuição de um potencial de Morse  $D$  e de empilhamento  $k$  específicos. A maneira como isto foi feito está mostrada esquematicamente na Figura 27, no lugar da cianina Cy3 usamos a letra X 'pareando' com um sítio vazio V, de tal maneira que se tem um pseudo-par-de-bases XV com potencial de Morse  $D_{XV}$ . Similarmente, para Cy5 usamos a letra Y, as tabelas 4 e 5 mostram as sequências já com esta nomenclatura. Como a Figura 27 sugere, os pares de bases canônicos AT e CG vizinhos das cianinas devem sofrer alterações nas suas ligações de hidrogênio e de empilhamento. Para representar essa situação de maneira distinta, acrescentamos X ou Y como índice aos pares canônicos AT e CG flanqueados por cianinas. Por exemplo, CG vizinho de Cy3 será representado por  $CG^X$ . Desta maneira, o empilhamento médio entre o marcador e o terminal do DNA será indicado por exemplo como  $k_{XV,CG^X}$  no caso de CG terminal vizinho de Cy3, como mostrado na Figura 27.

## 8.3 Procedimentos da otimização

A otimização dos parâmetros, partindo das temperaturas medidas, foi realizado seguindo o procedimento detalhado na seção 7.3, procurando minimizar a equação (31). Nós partimos de valores iniciais genéricos para os potenciais de Morse e empilhamento das cianinas e realizamos diversas rodadas de otimização. Nós analisamos os conjuntos de dados separa-

damente, seguindo os mesmos passos de parametrização para ambos. Fizemos três rodadas de minimizações globais para cada um. Nessas minimizações somente os parâmetros — potenciais de Morse e harmônico — com a presença das cianinas foram deixados livres para variar.

### 8.3.1 Valores iniciais

Para o parâmetro de potencial de Morse referente à Cy3 e Cy5, usamos o mesmo valor do par de bases canônico AT (33.1 meV) e para os potenciais harmônicos, interação de empilhamento do marcador com qualquer uma das bases canônicas, usamos 2.5 meV/nm<sup>2</sup>. Os parâmetros foram calculados à temperatura fixa de 370 K conforme detalhado na seção 7.2, esta temperatura não tem relação com as temperaturas de *melting*. Para os pares de bases canônicos vizinhos de cianinas, AT<sup>X</sup>, AT<sup>Y</sup>, CG<sup>X</sup> e CG<sup>Y</sup>, usamos como valor inicial os parâmetros de pares canônicos em 1020 mM de sódio de um trabalho anterior [91].

### 8.3.2 Minimização 1

Na primeira minimização, nós escolhemos aleatoriamente um novo conjunto de valores iniciais, por volta de  $\pm 20\%$ , a partir do parâmetro inicial que propusemos para XV e YV. Mantivemos apenas os pares de bases com X ou Y livres para variar. A cada nova minimização local, valores iniciais diferentes eram escolhidos randomicamente dentro de  $\pm 20\%$  daqueles que nós propusemos em 8.3.1. A única condição para a minimização foi que o potencial de Morse ( $D$ ) e o potencial harmônico ( $k$ ) deveriam ser positivos. Dessa forma, para cada um dos marcadores, foram realizadas 500 rodadas dessa primeira minimização, sendo que cada rodada foi realizada com um conjunto inicial de parâmetros diferente.

### 8.3.3 Minimização 2

Para essa nova minimização, utilizamos os valores médios dos parâmetros encontrados na minimização anterior (8.3.2) — resultados respectivos a cada marcador — como os valores iniciais dos parâmetros. Seguimos o mesmo procedimento de 8.3.2, mantendo a variação apenas dos parâmetros relacionados a X e Y. A cada rodada, um novo conjunto de valores iniciais diferentes foi escolhido randomicamente mantendo  $\pm 20\%$  ao redor dos valores médios obtidos em 8.3.2. Novamente, para cada marcador, foram realizadas 500 rodadas de minimização.

### 8.3.4 Minimização 3 — determinação da incerteza

Para estimar a incerteza dos novos parâmetros devido ao erro experimental, nós procedemos a um método de estatística robusta [92]. Para cada rodada, nós alteramos os dados originais aleatoriamente de tal maneira que a diferença entre o original e o novo conjunto resulta em um desvio padrão próximo do erro experimental relatado de 0.5 °C [45]. Isto simula a situação em que as medidas tivessem sido refeitas um certo número de vezes. Os resultados médios obtidos em 8.3.3 foram usados como valores iniciais dos parâmetros para a terceira e última

	$\langle \Delta T \rangle$ (C°)	$\chi^2$ (C°) <sup>2</sup>
Cy3	0.69	15.5
Cy5	0.63	10.9

**Tabela 1**

Parâmetros de qualidade após a última minimização 8.3.4.

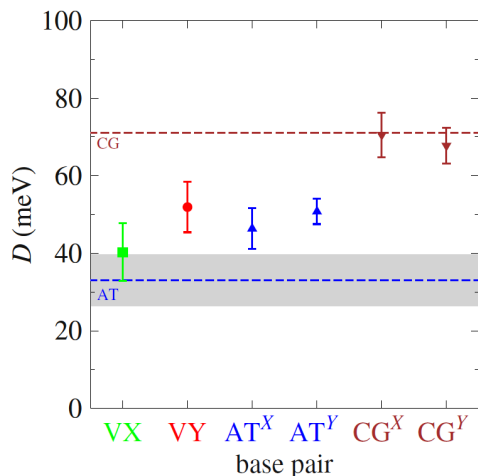
minimização. Continuamos a permitir a variação somente dos parâmetros relacionados a X e Y, e foram realizadas 500 rodadas, ou seja, 500 conjuntos diferentes de dados alterados como descrito acima. Os parâmetros de qualidade final estão mostrados na Tabela 1.

## 8.4 Resultados e Discussão

Para encontrarmos parâmetros de potencial de Morse e potencial harmônico para os marcadores Cy3 e Cy5, que pudessem ser inseridos em modelos mesoscópicos, otimizamos dois conjuntos de temperaturas de *melting* experimentais publicadas por Moreira et al. [45], conforme descrito na seção 8.1. A otimização foi realizada pelo processo de três minimizações descrito na seção 8.3. Os parâmetros de qualidade final são mostrados na Tabela 1 que mostram que a otimização dos parâmetros chegou a um valor que é aproximadamente o dobro do erro experimental, o que é similar ao que foi encontrado em outros tipos de parametrização pelo modelo PB [91, 93].

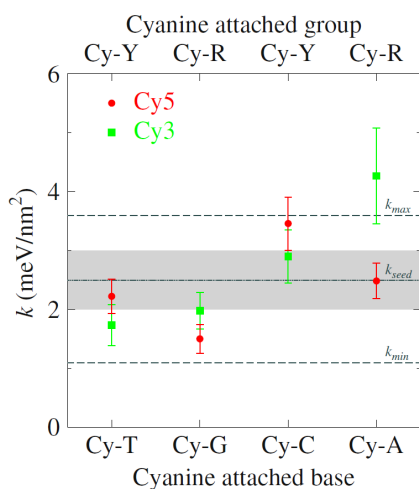
No gráfico da Figura 28, vemos a relação entre os valores de potencial de Morse para os dois marcadores, bem como a variação dos pares canônicos vizinhos de cianinas. As linhas tracejadas, indicadas por CG e AT, representam os valores de potencial de Morse para os pares canônicos, 70.7 meV e 33.1 meV, respectivamente [91]. A faixa sombreada mostra a variação dos parâmetros iniciais, de cada rodada, de  $\pm 20\%$  ao redor do valor de Morse de AT da minimização 1, como descrito na seção 8.3.2. Os valores de potencial de Morse para o par CG ligado tanto a Cy3 (CG<sup>X</sup>) quanto a Cy5 (CG<sup>Y</sup>) estão próximos do valor usado para o par canônico, indicando que os marcadores ligados a CG não perturbam a ligação de hidrogênio deste par de bases. Enquanto que para o par AT, para ambas as cianinas, há um leve aumento no potencial de Morse indicando um aumento na ligação de hidrogênio deste par quando da presença do marcador, seja Cy3 ou Cy5. A diferença mais distinta encontra-se nos valores relacionados às cianinas. Para Cy3 (VX), o potencial de Morse está bem próximo do inicial, mas ainda assim mostra um discreto aumento na estabilidade do duplexo. Para Cy5 (VY), observa-se uma estabilidade maior a partir do aumento no potencial. A similaridade dos potenciais de Morse para os marcadores com os potenciais canônicos parece indicar que é justificada a sua descrição, no modelo PB, dos marcadores como se fossem pares de bases comuns.

No gráfico da Figura 29, vemos a relação do potencial harmônico ( $k$ ) com a ligação do marcador a cada uma das quatro bases canônicas. As linhas tracejadas representam os valores mínimos e máximos para o empilhamento  $k$  para os próximos vizinhos das bases canônicas em DNA. Ou seja, valores de potencial de empilhamento que se localizam entre esses dois valores estão na mesma faixa que de bases canônicas. A faixa sombreada ao centro mostra o intervalo



**Figura 28**

Potencial de Morse, com as barras de erro, para cada marcador Cy3 e Cy5, respectivamente, VX e VY. Mostra também os potenciais de Morse para cada marcador ligado aos pares canônicos AT e CG. As linhas tracejadas indicam os valores dos pares canônicos. A faixa indica o intervalo de variação dos valores iniciais usados para Cy3 e Cy5.



**Figura 29**

O potencial de empilhamento, com as barras de erro, para cada marcador Cy3 e Cy5 quando atados a cada uma das bases canônicas. As linhas tracejadas indicam os valores mínimo e máximo de  $k$ . A faixa ao centro mostra o intervalo de variação dos valores iniciais para o potencial de empilhamento.

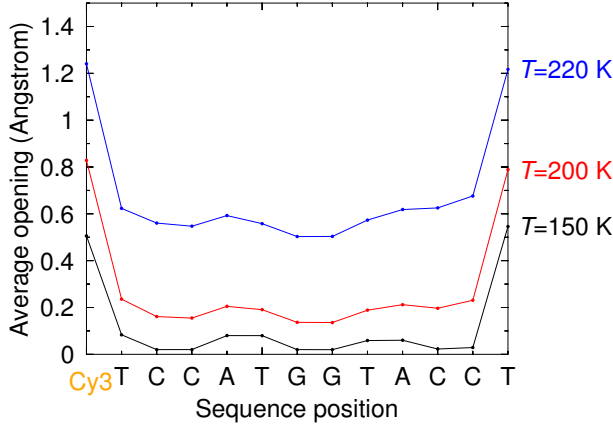
em que os parâmetros iniciais foram permitidos variar, como apresentado em 8.3, e as barras mostram, com os devidos desvios, os valores finais para Cy3 e Cy5. Observa-se que para Cy3 ligado à adenina há uma interação mais forte. Com Cy5 ligado à citosina, o potencial harmônico não se mostra tão elevado, mas se encontra bem próximo do valor máximo dos parâmetros de próximos vizinhos. Nesses dois casos, vemos uma maior interação indicando que Cy3-A e Cy5-C são interações que possivelmente estabilizam melhor o duplexo de DNA. É interessante notar que para três das quatro vizinhanças os valores de potencial de empilhamento de Cy3 e Cy5 são muito semelhantes. Somente entre Cy3-A e Cy5-A é que observamos uma diferença importante.

Diferente do potencial de Morse, os valores de potencial harmônico  $k$  das cianinas podem ser comparados com outros tipos de análise teórica ou dados experimentais. Isto porque representam potenciais efetivos de interação de empilhamento, diferente dos potenciais de Morse que neste caso não estão associados a ligações de hidrogênio já que estes, na verdade, não existem nas cianinas. Por exemplo, simulações de dinâmica molecular de Spiriti et al. [56] mostraram uma afinidade mais forte para Cy3-A do que comparado a Cy3-T, o que é consistente com os nossos resultados, veja a figura 29. A comparação entre energias de isomerização *cis-trans*  $E_{\text{iso}}$  [58, 66] e energias calculadas de desempilhamento (*unstacking*)  $\Delta G_{\text{un}}$  [56] estão

	$k$ (eV/nm <sup>2</sup> )	$E_{\text{iso}}$ (kcal/mol)	$\Delta G_{\text{un}}$ (kcal/mol)
Cy3-A	42.6	10.7	6.5
Cy3-C	28.9	8.4	8.4
Cy3-G	19.7	9.9	8.7
Cy3-T	17.3	4.9	4.2

**Tabela 2**

Comparação entre potenciais de empilhamento calculados, energias de ativação da isomerização *cis-trans*  $E_{\text{iso}}$  [58, 66], e energias livres de desempilhamento (*unstacking*)  $\Delta G_{\text{un}}$  calculadas na referência 56.



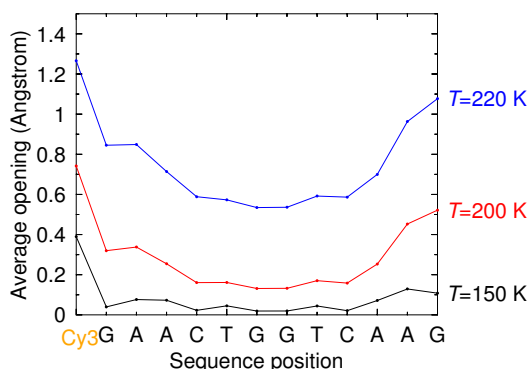
**Figura 30**

Aberturas médias, em três temperaturas de cálculo, para a sequência TCCATGGTACCT, marcada com Cy3 no terminal 5'.

mostradas na Tabela 2. Há uma boa correlação entre os nossos resultados de interação de empilhamento e as energias experimentais de isomerização *cis-trans*, de fato melhores que as energias de desempilhamento com as quais Spiriti et al. [56] as comparam. A tendência de energias livres  $\Delta G_{\text{un}}$  de Kroutil et al. [36] segue  $G \approx C > A > T$ , enquanto que a dos nossos parâmetros de empilhamento seguem  $A > C > G \approx T$ . Kroutil et al. [36], também usando simulação de dinâmica molecular, obtiveram energias livres de desempilhamento de 6.2 kcal/mol, menor do que os resultados de Spiriti et al. [56], sendo que atribuem este valor menor à diferenças nos campos de força utilizados.

Análises de microarranjo mostram campos de fluorescência [35, 44] mais baixos para Cy-C, o que se atribui a uma interação de empilhamento mais fraca, enquanto Cy-G apresenta uma eficiência mais alta, provavelmente, devido a uma interação de empilhamento mais forte. Os nossos resultados indicam uma situação inversa em que Cy5-C mostra uma interação de empilhamento mais forte e Cy5-G a mais fraca. Não é clara a razão dessa diferença, mas é possível que seja devido às interações em microarranjo serem diferentes das interações de DNA em solução ou haja fatores além das interações de empilhamento os quais influenciam a eficiência de fluorescência. No próximo capítulo, discutiremos esta questão em mais detalhe.

O comportamento dos marcadores Cy3 e Cy5 também podem ser estudado pelos perfis de abertura média realizando o cálculo da equação (24), página 29. Esses perfis mostram o deslocamento entre as fitas da molécula de DNA. Na figura 30, mostramos um exemplo para uma das sequências usadas em nosso trabalho. Para cada nucleotídeo  $m$  é calculado um valor médio de deslocamento das ligações de hidrogênio  $\langle y_m \rangle$  que depende da temperatura de trabalho  $T$ . Sobre esta temperatura  $T$  há duas observações importantes a fazer, a primeira é que ela não possui relação com a temperatura de desnaturação que é calculada pela equação (29), página 31.



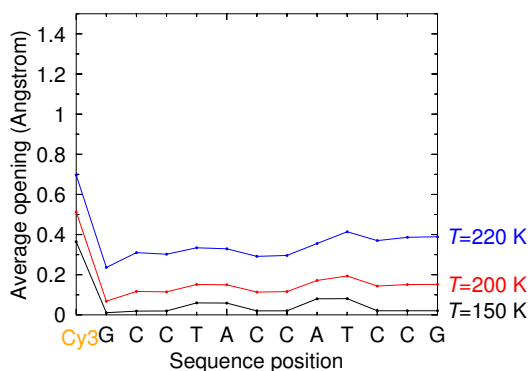
**Figura 31**

Aberturas médias, em três temperaturas de cálculo, para a sequência GAAGTGGTCAAG, marcada com Cy3 no terminal 5'.

A segunda é que para obter perfis de deslocamento menores que 1 Å para duplexos menores que 30 pares de bases, é necessário reduzir esta temperatura a valores pouco realistas. A razão básica disto é que o modelo **PB** é um modelo bidimensional, veja a Figura 20, página 27, ou seja, falta um grau de liberdade importante para que a descrição termodinâmica seja realista. Apesar destas ressalvas, estes perfis são úteis na análise qualitativa de duplexos de **DNA**. Por exemplo, observe o aumento acentuado do deslocamento nos terminais da sequência mostrada na Figura 30, isto corresponde ao *end fraying* conhecido em **DNA** e **RNA** [94]. Tipicamente, na região dos pares AT há um aumento do deslocamento devido a menor estabilidade térmica. Na medida em que a temperatura  $T$  é aumentada, esse deslocamento aumenta e as diferenças entre regiões diferentes do duplexo ficam menos pronunciadas. A abertura média no marcador Cy3 é muito similar, neste caso, a de um terminal AT (lado direito da Figura 30), o que já era esperado dado que os potenciais de Morse e de empilhamento serem muito similares a de bases canônicas.

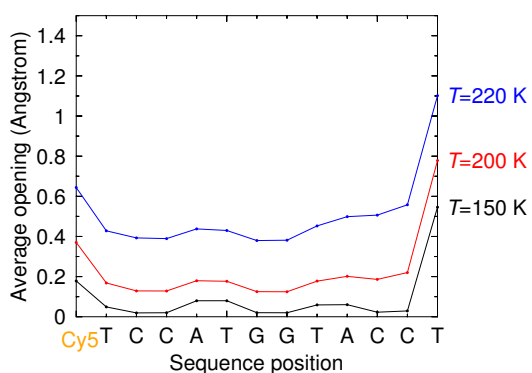
Nas Figuras 31 e 32, observamos duas sequências com a presença de Cy3 e como o ambiente sequencial delas influencia na estabilidade do marcador com a molécula de **DNA**. Na primeira figura, vemos Cy3 ligado à guanina seguido de duas adeninas e aberturas médias altas em 200 K e 220 K. Em comparação à segunda figura, é evidente como guanina seguida de duas citosinas mantém o duplexo mais estável. Nas Figuras 33 e 34, vemos Cy5 ligado a uma timina seguida, na primeira figura, por uma citosina e, na segunda figura, por outra timina. Comparando as duas figuras, notamos que Cy5 ligado a uma timina seguida de uma citosina mantém o duplexo mais estável do que quando a timina é seguida por outra timina. Isso é mais evidente para as temperaturas mais elevadas. Essas quatro figuras mostram, o que já foi dito na literatura (mencionado em 6.2), de que os marcadores Cy3 e Cy5 são dependentes da sequência das bases nitrogenadas. As Figuras 35 e 36 trazem a sutileza dessa influência mostrando que Cy5 ligado à uma citosina seguida de uma timina e guanina, nessa ordem, apresenta aberturas médias mais elevadas em comparação com Cy5 ligado à citosina seguida de guanina e timina.

Podemos observar também como Cy5 se mostra um marcador mais estável do que Cy3. Nas Figuras 30 e 33, temos exatamente a mesma sequência, em uma figura com a presença do marcador Cy3 e na outra com Cy5. Vemos que Cy5 estabiliza muito mais essa sequência em comparação a Cy3. A mudança brusca nos valores de abertura média dessa situação mostram



**Figura 32**

Aberturas médias, em três temperaturas de cálculo, para a sequência GCCTACCATCCG, marcada com Cy3 no terminal 5'.

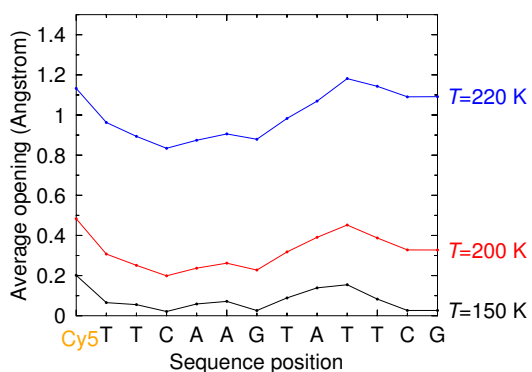


**Figura 33**

Aberturas médias, em três temperaturas de cálculo, para a sequência TCCATGGTACCT, marcada com Cy5 no terminal 5'.

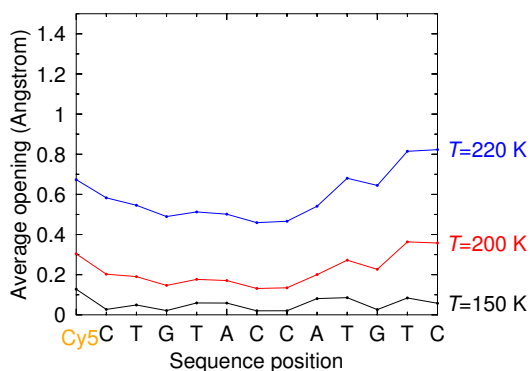
uma interação mais rígida para Cy5 do que para Cy3. O mesmo pode ser observado nas Figuras 37 e 35. Temos a mesma sequência e na figura em que o marcador usado é o Cy3 o duplexo se desestabiliza mais do que aquela em que tem-se o marcador Cy5. Há outra comparação para uma mesma sequência marcada com Cy3 e Cy5 em A.2.

Pelos gráficos de abertura média, observa-se que o ambiente sequencial em que os marcadores estão inseridos influencia o comportamento das cianinas, como é relatado na literatura [45]. Nossos resultados mostram e corroboram o comportamento de Cy3 e Cy5 como um par de bases adicional quando atados ao terminal da molécula de DNA. Os valores que obtivemos para o potencial de Morse evidenciam um comportamento muito similar ao par canônico AT mantendo a estabilidade do duplexo e interagindo com os vizinhos pelo potencial harmônico.



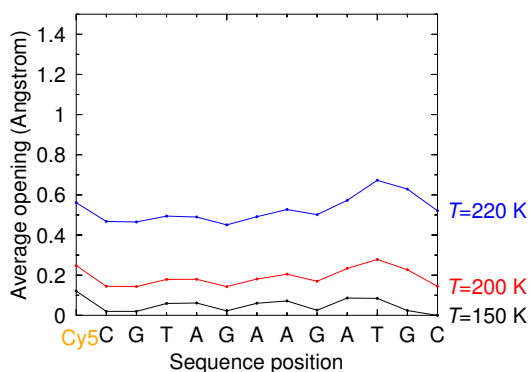
**Figura 34**

Aberturas médias, em três temperaturas de cálculo, para a sequência TTCAAGTATTCG, marcada com Cy5 no terminal 5'.



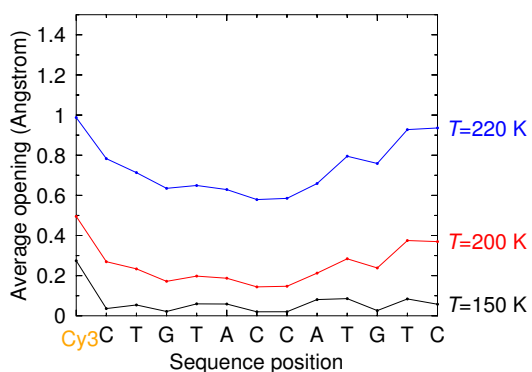
**Figura 35**

Aberturas médias, em três temperaturas de cálculo, para a sequência CTGTACCATGTC, marcada com Cy5 no terminal 5'.



**Figura 36**

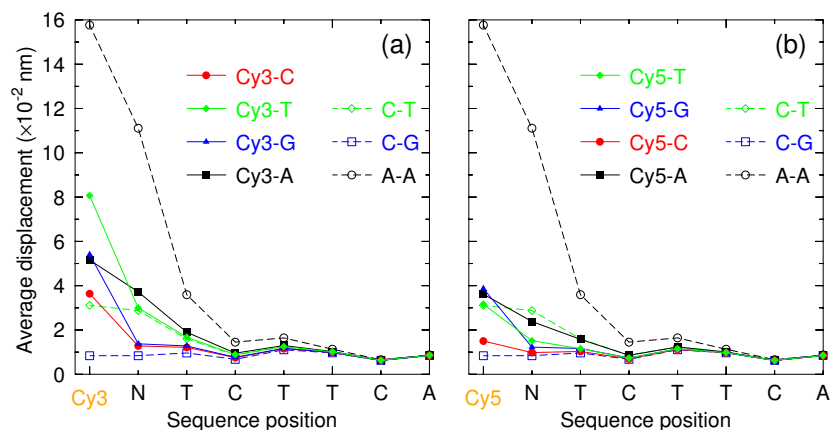
Aberturas médias, em três temperaturas de cálculo, para a sequência CGTAGAAGATGC, marcada com Cy5 no terminal 5'.



**Figura 37**

Aberturas médias, em três temperaturas de cálculo, para a sequência CTGTACCATGTC, marcada com Cy3 no terminal 5'.





**Figura 38**

Posição da sequência pela abertura média. Na primeira posição está o marcador, Cy3 (à esquerda) e Cy5 (à direita), atado a uma base N, a qual pode ser A (quadrados pretos), C (bolinhas vermelhas), G (triângulos azuis) e T (losangos verdes). Para referência, as linhas tracejadas com símbolos vazados são para o par de bases CT (losangos verdes), CG (quadrados azuis) e AA (círculos pretos). Somente os sete primeiros pares de bases de uma sequência de 15 pares, da referência [56], são mostrados.

A Figura 38, mostra as aberturas calculadas para as sequências de Spiriti et al. [56] e resume a discussão sobre os perfis de abertura, onde também são incluídas as sequências sem marcadores. Note que na ausência de Cy3 os dois pares AT terminais consecutivos (linha preta tracejada da figura 38a) criam uma instabilidade muito acentuada que diminui quando o primeiro par AT é substituído por uma cianina. Por outro lado, dois pares CG terminais consecutivos (linha azul tracejada) ainda é a situação de maior estabilidade, neste caso a substituição de CG por uma cianina é um fator de instabilidade.

## 8.5 Conclusão

Mostramos que é possível representar marcadores fluorescentes no modelo mesoscópico PB como se fossem um outro par de bases. Os resultados de interação de empilhamento possuem uma correspondência com as energias de isomerização *cis-trans* de Cy3/Cy5, e os perfis de abertura média indicam claramente a maneira como as cianinas estabilizam o duplexo. No próximo capítulo, vamos explorar mais os perfis de abertura para procurar entender se estes poderiam ser usados como preditores de fluorescência dos marcadores.

## 9 Correlação da Intensidade de Fluorescência

No capítulo anterior, mostramos que os nossos resultados indicam os marcadores Cy3 e Cy5 estabilizarem o duplexo de **DNA** e que eles podem ser representados no modelo **PB** considerando-os como um par de bases. Neste capítulo, abordamos o estudo sobre as intensidades de fluorescência de dados de microarranjo correlacionadas às aberturas médias. Apresentamos os dados experimentais utilizados; as etapas dos procedimentos conduzidos; os resultados que encontramos para cada etapa executada; e, por fim, discutimos os resultados obtidos.

### 9.1 Dados Experimentais

Nessa segunda parte do trabalho, usamos dados de intensidade de fluorescência publicados por Agbavwe and Somoza [35]. Os autores fizeram experimentos de microarranjo de oligômeros de **DNA** fita simples para obterem dados de intensidade de fluorescência para 1024 sequências. Essas sequências foram ligadas à Cy3 e Cy5 pelo terminal 5' e sintetizadas a partir do esquema a seguir:

5' - Cy\* - N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>N<sub>4</sub>N<sub>5</sub> - T<sub>15</sub> - (ACGT-N<sub>1</sub>) - (ACGT-N<sub>2</sub>) - (ACGT-N<sub>3</sub>) - (ACGT-N<sub>4</sub>) - (ACGT-N<sub>5</sub>) - T<sub>5</sub> - *vidro*

onde N representa uma base nitrogenada e o asterisco (Cy\*) o marcador fluorescente (Cy3 ou Cy5). O pentâmero atado à cianina está ligado ao restante da sequência por um *linker* de 15 timinas e determina a última base dos pentâmeros seguintes. Essas cinco bases também são o ponto de partida para variar todas as combinações de bases nitrogenadas. Por exemplo, o primeiro pentâmero gerado foi AAAAA (primeira sequência do conjunto), o seguinte foi AAAAC (segunda sequência do conjunto), mais à frente, tem-se AAACC e seguindo dessa forma [35]. A fita simples de **DNA** termina com a sequência de 5 timinas, as quais foram presas à superfície de vidro para realizar o microarranjo.

Com esse esquema, os autores conseguiram 1024 sequências distintas e uniformes com o intuito de alcançar todas as diferentes interações das bases canônicas entre si e os marcadores e garantir a mesma densidade de cada sequência no microarranjo. Como eles usaram dois marcadores fluorescentes, dividimos os dados em dois conjuntos, cada qual com as mesmas sequências, porém um conjunto com sequências atadas à Cy3 e o outro com as sequências atadas à Cy5.

### 9.2 Procedimentos

Para o nosso trabalho, resolvemos estudar os dados de intensidade de fluorescência tomando as sequências de Agbavwe and Somoza [35] como fitas duplas, ao invés das fitas simples customizadas e estudadas pelos autores. Tendo gerado as fitas complementares, criamos dois arquivos, um para as sequências ligadas à Cy3 e outro com as de Cy5. Em cada arquivo, inserimos as 1024 sequências de **DNA** e seus complementares, com X na posição de Cy3 e Y na posição

de Cy5, e as intensidades de fluorescência respectivas à cada sequência, similar ao que usamos no capítulo 8. Cada conjunto de dados foi analisado seguindo os mesmos passos que serão apresentados a seguir.

### 9.2.1 Passo 1

Obtivemos os dados de intensidade de fluorescência do material suplementar de Agbavwe and Somoza [35] a partir do endereço eletrônico indicado no artigo. Geramos as sequências completas a partir do esquema apresentado em 9.1, pois em seu arquivo eles indicam apenas o primeiro pentâmero, e seus complementares. Com todas as sequências geradas, montamos os arquivos para ambos os marcadores constituídos das sequências, seus complementares e as intensidades de fluorescência.

Criamos também um arquivo com temperaturas de cálculo, como mencionado em 7.2 para o modelo PB, a serem usadas para os cálculos das aberturas médias. O intervalo de temperaturas que utilizamos para o cálculo foi de 150 K a 350 K (com variação de 10 em 10). No caso deste passo, queríamos obter as aberturas médias ( $\langle y \rangle$ ) de todas as sequências. Como parâmetros iniciais, usamos para as bases canônicas valores obtidos em trabalhos anteriores equivalentes para a concentração salina de DNA a 1020 mM e para os marcadores fluorescentes valores obtidos dos procedimentos de 8. Os scripts usados para gerar as sequências e seus complementares e para obtenção das aberturas médias encontram-se em B.1, B.2 e B.3, página 66.

### 9.2.2 Passo 2

Calculadas as aberturas médias de todas as sequências, extraímos para um novo arquivo apenas os valores de abertura média dos pares formados pelos marcadores, XV e YV. A partir desse ponto, tínhamos dois conjuntos de valores: os valores de intensidade de fluorescência e os valores de abertura média dos marcadores. Com esses dois conjuntos distintos, calculamos o coeficiente de correlação de Pearson (7.4) — intensidades de fluorescência por aberturas médias — para cada uma das temperaturas de cálculo que mencionamos anteriormente em 9.2.1. Os scripts para extrair os valores de abertura média dos marcadores e para o cálculo do coeficiente de correlação encontram-se em B.4 e B.5, respectivamente, página 69.

### 9.2.3 Passo 3

Do procedimento de calcular as aberturas médias em diante, fizemos novamente os passos 9.2.1 e 9.2.2, variando a concentração salina de DNA. Em 9.2.1, usamos parâmetros para concentração salina de 1020 mM. Agora, fizemos uma nova rodada para cada conjunto de parâmetros, obtidos de trabalhos anteriores, de quatro diferentes concentrações de sódio: 69 mM, 119 mM, 220 mM e 621 mM. Calculamos as aberturas médias e o coeficiente de correlação para elas e as intensidades de fluorescência.

#### 9.2.4 Passo 4

Refizemos as rodadas nesse passo variando os valores de potencial de Morse  $D$ . Mantivemos como parâmetros iniciais os valores referentes à concentração salina de DNA a 1020 mM. Usamos quatro valores diferentes de potencial de Morse e diferentes também entre os marcadores fluorescentes. Calculamos as aberturas médias para todas as sequências em cada variação do potencial de Morse e, em seguida, calculamos o coeficiente de correlação de Pearson.

#### 9.2.5 Passo 5

Ainda mantendo os parâmetros iniciais para a concentração salina a 1020 mM, sem alterar os valores de potencial de Morse — retomando os valores parametrizados para Cy3 e Cy5 em 8 —, fizemos novas rodadas variando o valor referente ao empilhamento das bases que chamamos de valor de  $k$  ou potencial harmônico. Variamos o valor de  $k$  para os quatro pares de bases canônicos: AT, CG, GC e TA. A variação de cada um foi distinta, equivalendo a uma rodada para cada par. Dessa forma, tivemos ao final quatro novas rodadas para cada cianina. Em cada uma, calculamos as aberturas médias e o coeficiente de correlação.

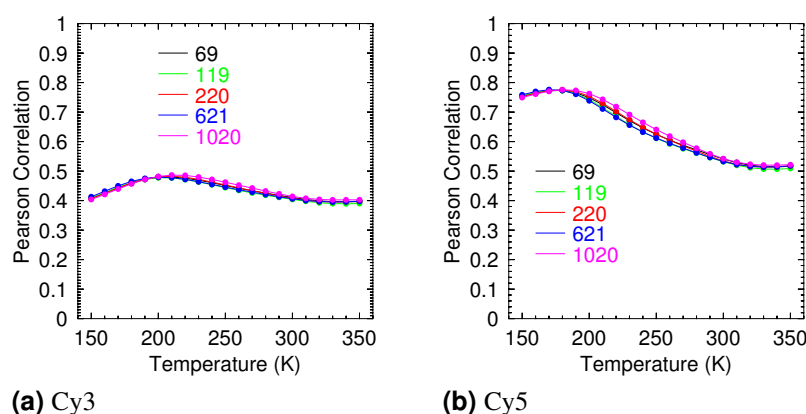
#### 9.2.6 Passo 6

Nesse passo, usamos os parâmetros iniciais para a concentração salina a 1020 mM e mantivemos fixo o valor de potencial de Morse. Encontramos o melhor resultado para cada par canônico das rodadas de 9.2.5 e variamos os parâmetros para o segundo vizinho de cada um. Por exemplo, encontramos o melhor resultado para o par AT, mantivemos o valor de potencial harmônico de AT fixo e variamos o potencial harmônico do vizinho seguinte, o qual poderia ser AT, CG, GC e TA. Fazendo uma rodada diferente para cada diferente vizinho do melhor valor de AT. Para o par CG temos os vizinhos AT, CG e GC. Para o par GC, os vizinhos são AT e CG. E para o par TA, um único vizinho, o par AT.

#### 9.2.7 Passo 7

Otimizamos os coeficientes de correlação de Pearson, usando o script de B.7, página 75. Fizemos quatro rodadas de otimização em que variamos certos parâmetros, enquanto mantivemos outros fixos. Os parâmetros passíveis de variação foram: potencial de Morse, potencial harmônico e temperatura. Quando mencionamos qual parâmetro foi variado subentende-se que os outros parâmetros foram mantidos fixos. As otimizações realizadas foram as seguintes:

1. Variação do potencial de Morse para os pares de bases AT e CG.
2. Variação do potencial de Morse para os pares de bases AT e CG e da temperatura.
3. Variação dos potenciais harmônicos para todos os pares de bases vizinhos.



**Figura 39**

Coefficientes de correlação de Pearson para cinco diferentes concentrações salinas de DNA para as sequências marcadas com (a) Cy3 e (b) Cy5, obtidos segundo o passo 9.2. No eixo horizontal estão as temperaturas de cálculo usadas no modelo estatístico.

4. Variação dos potenciais harmônicos e os potenciais de Morse para os pares de bases AT e CG.

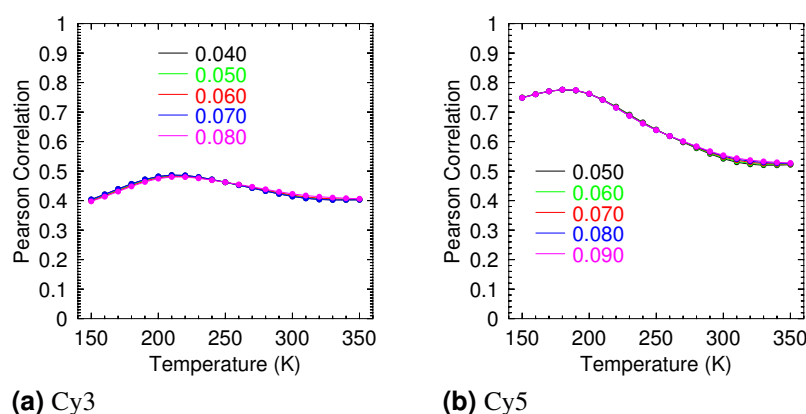
## 9.3 Resultados

Para encontrarmos um melhor preditor para Cy3 e Cy5, analisamos como as intensidades de fluorescência do trabalho de Agbavwe and Somoza [35] correlacionavam com as aberturas médias das sequências sintetizadas pelos autores. Aplicamos o modelo Peyrard-Bishop usando os parâmetros de potencial de Morse e potencial harmônico para as bases canônicas de trabalhos anteriores e os parâmetros para Cy3 e Cy5 que encontramos pelos procedimentos de 8.

### 9.3.1 Variação da Concentração Salina

Para calcular as aberturas médias, os parâmetros a serem usados devem ser de acordo com condições estabelecidas. Uma dessas condições é a concentração salina do DNA. Calculamos as aberturas médias em cinco diferentes concentrações salinas: 69 mM, 119 mM, 220 mM, 621 mM e 1020 mM. Seguindo os procedimentos de 9.2, calculamos o coeficiente de correlação para as intensidades de fluorescência e as aberturas médias. Na Figura 39, estão as temperaturas de cálculo pelos coeficientes de correlação de Pearson das cinco concentrações salinas usadas para Cy3 e Cy5.

Observamos que tanto para Cy3 quanto para Cy5, a relação temperatura pelo coeficiente de correlação é muito similar entre as diferentes concentrações. Para Cy5, Figura 39b, há uma diferença mais notável no intervalo de 200 K a 280 K. Enquanto que Cy3 apresenta uma diferença não tão distinta, mas observável no intervalo de 210 K a 280 K. Vemos também que para ambos os marcadores, a similaridade entre os resultados das concentrações é basicamente universal, alcançando o máximo de similaridade na temperatura de 210 K para Cy3 e 180 K



**Figura 40**

Coefficientes de correlação de Pearson para valores diferentes de potencial de Morse (em eV) para as sequências marcadas com (a) Cy3 e (b) Cy5.

para Cy5. Essas temperaturas equivalem ao valor máximo encontrado para o coeficiente de correlação para cada cianina.

Na Figura 39a, encontramos a maior correlação para o marcador Cy3 ao redor de 0.5. Isso indica uma correlação positiva, em todas as concentrações que usamos, e uma correlação mais alta no intervalo de temperatura de 200 K a 220 K. Para Cy5 (Figura 39b), vemos uma correlação positiva alta ao redor de 0.8, no intervalo de temperatura de 170 K a 190 K.

### 9.3.2 Variação do Potencial de Morse

Para a variação do potencial de Morse para Cy3 e Cy5, partimos do valor do potencial de Morse de cada marcador para variar os valores dessa etapa. Como encontrado em 8, temos o valor de 40.2 meV para Cy3. Então variamos os valores de potencial de Morse em 50 meV, 60 meV, 70 meV e 80 meV. Para Cy5, encontramos o valor de 51.2 meV e variamos os valores de potencial de Morse em 60 meV, 70 meV, 80 meV e 90 meV. Todas essas variações foram realizadas com parâmetros iniciais para a concentração salina de DNA a 1020 mM.

Os resultados dessa variação podem ser vistos na Figura 40. Observamos que mesmo variando os valores de potencial de Morse, a relação entre temperaturas e coeficientes de correlação não se alteram muito em comparação aos resultados da variação da concentração salina apresentada em 9.3.1. O valor máximo de coeficiente de correlação alcançado por Cy3 e Cy5 foi o mesmo encontrado para a variação da concentração salina — ao redor de 0.5 e 0.8, respectivamente. Assim como a temperatura referente ao maior coeficiente encontrado, 210 K para Cy3 e 180 K para Cy5.

O que notamos de diferente é que para Cy5, que apresentava uma leve distinção entre os coeficientes à temperaturas mais baixas, há uma similaridade absoluta para as variações até alcançar a temperatura aproximada de 280 K, quando observa-se uma leve distinção entre os diferentes valores para o potencial de Morse. Quanto a Cy3, a mudança está na similaridade maior no intervalo de 240 K a 270 K em comparação à Figura 39 em que vemos uma discrepância

entre os coeficientes de correlação nesse intervalo.

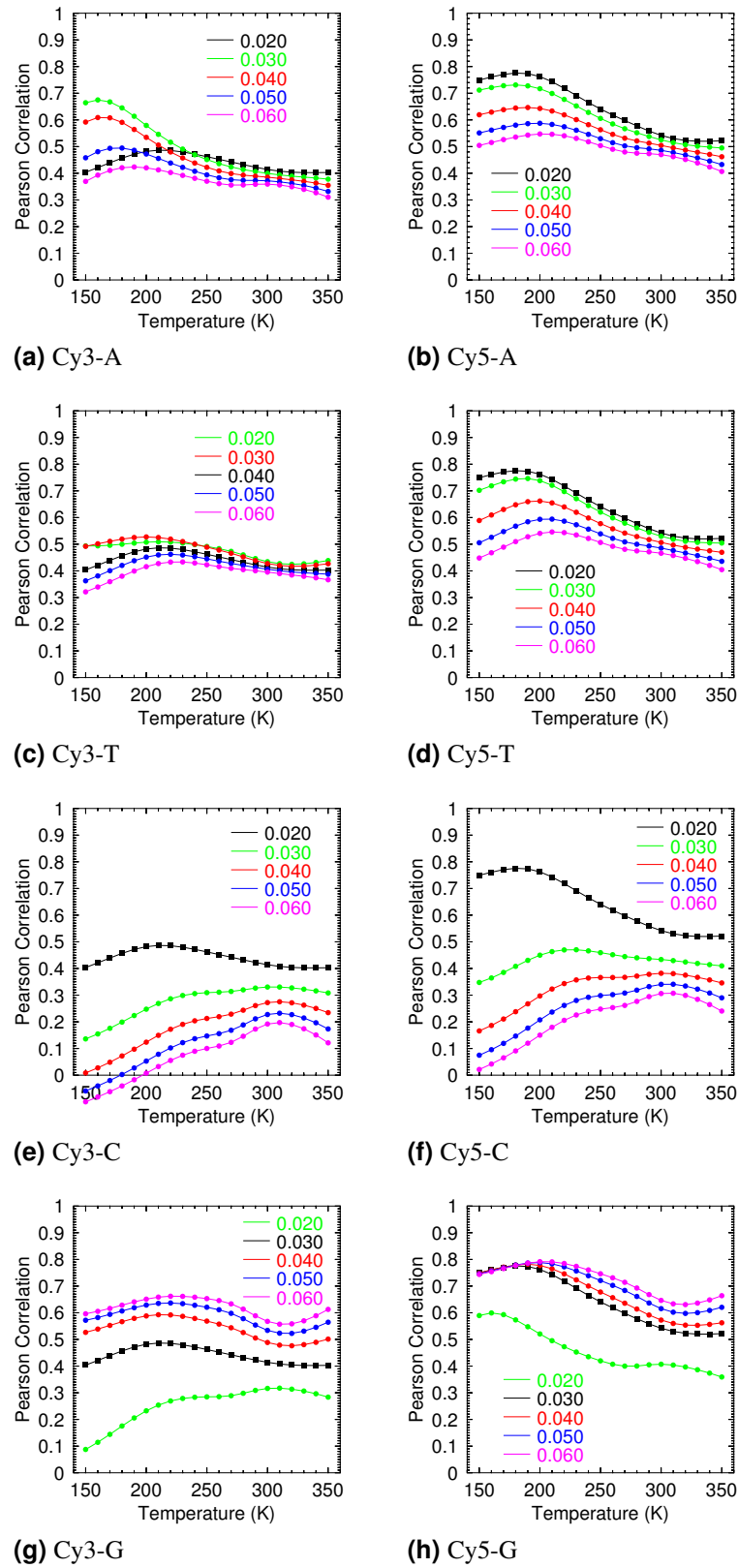
### 9.3.3 Variação do Potencial Harmônico

Nessa etapa, variamos o potencial harmônico referente ao marcador e o primeiro par de bases. Por exemplo, Cy3 (XV) ligado ao par de bases AT, ao par CG, ao par GC e ligado ao par TA. A variação dos valores de potencial harmônico tiveram como referência os parâmetros encontrados para esse potencial nos procedimentos descritos em 8. Para essa etapa, todas as variações foram realizadas com os parâmetros iniciais para o DNA à concentração salina de 1020 mM.

Nas figuras, a seguir, para a relação de temperatura com coeficiente de correlação é evidente a discrepância entre os diferentes valores usados para os potenciais harmônicos para os dois marcadores. Na Figura 41a, em que Cy3 está ligado ao par AT, vemos o valor de referência — concentração salina a 1020 mM e parâmetro estabelecido anteriormente para  $k$  — representados pela linha com quadrados (linha preta). Observamos não haver um intervalo de similaridade para os coeficientes como vimos para as variações para concentração salina (9.3.1) e potencial de Morse (9.3.2). À temperaturas baixas, a discrepância é muito alta dentro do intervalo de coeficiente de correlação de aproximadamente 0.3 a 0.7. Sendo também o intervalo de maior alcance da correlação encontrada e aumentando em comparação aos resultados de 9.3.1 e 9.3.2 — de uma correlação máxima de 0.5 para 0.7. À temperaturas mais altas, o intervalo de correlação estreita e o valor máximo de correlação cai — entre 0.3 e 0.4. Algo similar é visto na Figura 41c, em que Cy3 está ligado ao par TA (linha preta com quadrados). O valor máximo do coeficiente de correlação se mantém em 0.5. Entretanto, o intervalo de correlações aumenta em comparação às variações de concentração salina e de potencial de Morse, passando a ser entre 0.3 e 0.5. Assim como vimos para a ligação com o par AT, à baixas temperaturas, a discrepância é maior do que à temperaturas mais altas.

As Figuras 41b, 41d e 41f mostram as temperaturas pelos coeficientes de correlação para Cy5 atado aos pares AT, TA e CG, respectivamente. Observa-se que a curva (preta com quadrados) para concentração salina a 1020 mM e parâmetro de potencial harmônico encontrado em nosso trabalho anterior (8) é o alcance máximo dos coeficientes de correlação para Cy5 ligado a esses pares de bases. Nas variações dessa etapa há uma elevada discrepância entre as correlações. O intervalo dos coeficientes varia extremamente em comparação aos intervalos encontrados em 9.3.1 e 9.3.2. Quando ligado aos pares AT e TA, o intervalo de correlação para Cy5, à temperaturas baixas, se encontra entre 0.4 e 0.8, se estreitando à temperaturas mais altas, porém diminuindo o coeficiente máximo encontrado. A maior discrepância foi encontrada para o par CG (Figura 41f), em que o intervalo das correlações para temperaturas baixas foi de 0.0 a 0.8. Em compensação, à temperaturas mais altas, o intervalo se estreita para 0.2 a 0.6, diminuindo em comparação ao valor máximo encontrado de 0.8.

A discrepância mencionada para Cy5 atado ao par CG também é observada para Cy3 atado ao mesmo par, veja Figura 41e. Porém a correlação cai drasticamente à temperaturas baixas



**Figura 41**

Coeficientes de correlação de Pearson para valores diferentes de potencial harmônico ligado ao par AT e TA (em  $\text{eV}/\text{\AA}^2$ ) para as sequências marcadas com Cy3 e Cy5.



encontrando-se entre negativa e positiva — intervalo entre -0.2 e 0.5. Para temperaturas mais altas, o intervalo retoma, ainda que com valores baixos, para positivamente correlacionado entre 0.1 e 0.4.

O que vemos de mais interessante é a mudança nas curvas para Cy3 e Cy5 atados ao par canônico GC, Figuras 41g e 41h. A ordem das curvas, em comparação aos resultados dos outros pares de bases, inverte para os dois marcadores. A maior variação do valor de potencial harmônico — para ambos os marcadores, 60 meV — alcança o valor máximo do coeficiente de correlação, sendo aproximadamente 0.7 para Cy3 e 0.8 para Cy5. Para Cy3, à temperaturas baixas, o intervalo de correlação cai levemente abaixo de 0.1 alcançando 0.6. O valor máximo, aproximadamente 0.7, é encontrado para o intervalo de temperatura entre 210 K e 240 K. Para Cy5 (Figura 41h), até a temperatura de 180 K, as curvas são bastante similares, à exceção do potencial harmônico mais baixo usado. E é nesse ponto em que temos o maior coeficiente de correlação dessa etapa. A partir de 180 K, as correlações caem e se diferenciam até se elevarem sutilmente em 310 K em diante.

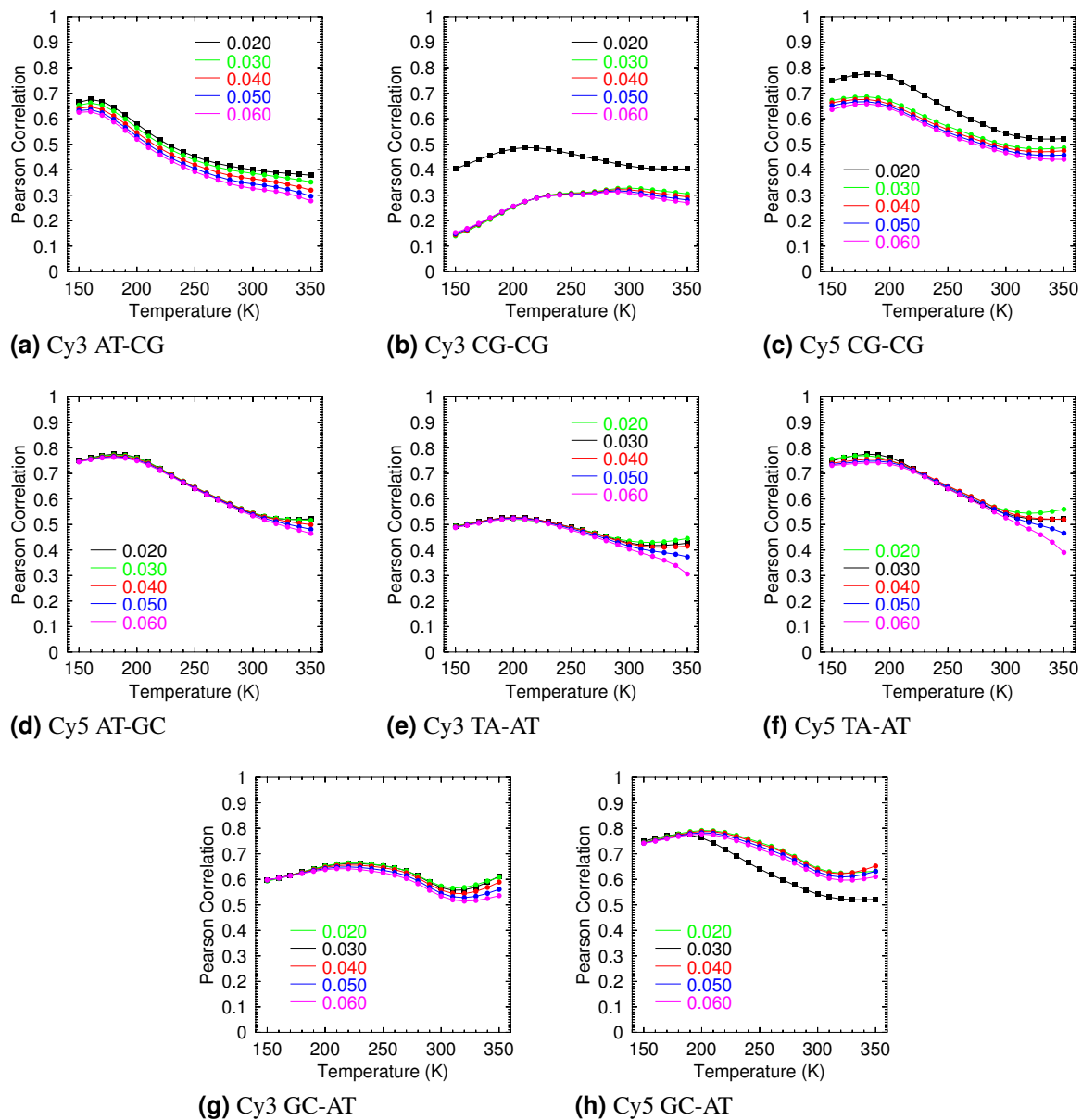
### 9.3.4 Variação dos Segundos Vizinhos

Até este ponto estudamos as variações das concentrações salinas de DNA, do potencial de Morse e do potencial harmônico das cianinas com a base a qual se ligam. Da etapa descrita em 9.2.6, mantivemos fixo o melhor valor obtido em 9.2.5 para cada par de bases e variamos os segundos vizinhos da base ligada ao marcador. Usamos os parâmetros iniciais de DNA à concentração salina de 1020 mM e, para os segundos vizinhos, variamos ao redor do valor do melhor resultado encontrado em 9.3.3. Por exemplo, encontramos o melhor valor de potencial harmônico para o par AT, o mantivemos fixo e variamos o valor de  $k$  para cada vizinho de AT — AT, CG, GC e TA.

Os resultados dessa etapa mostram uma discrepância menor em comparação ao discutido em 9.3.3. Ainda assim, os coeficientes de correlação sofrem variações em relação aos diferentes valores de potencial harmônico usados para os segundos vizinhos. Na Figura 42a, o melhor resultado de Cy3-AT ligado a CG, temos a curva de referência em preto e com quadrados, como nas descrições anteriores. Vemos que o intervalo de coeficientes é estrito ao longo de todo eixo horizontal e alcança o coeficiente máximo ao redor de 0.7 — a curva mais elevada é a de referência. Também nesse caso as maiores correlações estão no intervalo das temperaturas mais baixas.

Para o melhor resultado de Cy3-CG ligado ao par CG, vemos na Figura 42b que a curva de referência se destaca das outras curvas. Os outros valores usados nessa variação (0.030, 0.040, 0.050 e 0.060) caem no intervalo de correlação. O mesmo pode ser visto para a mesma interação de pares de bases para Cy5 (Figura 42c). A curva de referência se mantém acima das outras curvas que caem um pouco em relação à de referência, porém a diferença é menos drástica do que em comparação aos resultados de Cy3 (Figura 42b).

Quando observamos os resultados de Cy5-AT ligado ao par GC, na Figura 42d, vemos que



**Figura 42**

Coeficientes de correlação de Pearson para valores diferentes de potencial harmônico de segundo vizinho (em eV/Å<sup>2</sup>) para as sequências marcadas com Cy3 e Cy5.

a diferença entre as curvas é extremamente sutil se tornando mais evidente à temperaturas mais elevadas. Mas ainda assim, com um intervalo de coeficientes bem restrito e valor máximo ao longo de toda a curva em 0.8. A Figura 42f mostra os resultados de Cy5-TA ligado ao par AT. Há uma similaridade no intervalo de 230 K a 260 K, entretanto a partir dessa temperatura os coeficientes começam a se diferenciar, tornando-se mais evidente essa diferença em 300 K em diante.

Para Cy3-TA ligado ao par AT, na Figura 42e, os coeficientes de correlação são extremamente similares até à temperatura aproximada de 250 K quando começam a se diferenciar e, a partir de 300 K, essa diferença se torna mais evidente. À temperaturas mais altas, o coeficiente de correlação da curva de referência compete com as variações de pontecial harmônico para AT de 0.020 e 0.040.

Na Figura 42g, temos Cy3-GC ligado ao par AT, e podemos ver que, à baixas temperaturas, os coeficientes de correlação são similares. Porém, à medida que as curvas evoluem, elas se diferenciam entre si num intervalo de diferença restrito. Nesse caso, as correlações se encontram num valor alto de correlação positiva. Entretanto, para Cy5-GC ligado ao par AT, a curva de referência cai drasticamente em comparação às outras curvas, Figura 42h. Os coeficientes dos outros valores de potencial harmônico, que não sejam da referência, se mantêm em um intervalo similar se diferenciando levemente. Os coeficientes da referência, a partir da temperatura de 200 K, caem até aproximadamente 0.5.

### 9.3.5 Otimização dos parâmetros

Nas seções anteriores nós vimos que há uma considerável sensibilidade dos coeficientes de correlação com os parâmetros do modelo, mesmo em nível de segundos vizinhos. Aqui, nós vamos automatizar o procedimento como descrito na seção 9.2.7. A otimização teve o objetivo de verificar qual o valor máximo possível para o coeficiente de correlação de Pearson.

Esta etapa ainda está em andamento e em fase de conclusão. Temos alguns resultados preliminares (Tabela 3) que mostram similaridade com os resultados das etapas anteriores (9.3). Quando variamos também o potencial de Morse, em 9.3.2, encontramos para Cy3 um coeficiente de correlação ao redor de 0.5 e para Cy5, ao redor de 0.8. Esses resultados preliminares são similares aos resultados obtidos no capítulo 8, o que é bastante encorajador. No entanto esta parte do nosso trabalho ainda não está finalizada, temos de ampliar o número de parâmetros a serem otimizados o que é computacionalmente bastante demorado.

## 9.4 Discussão

Vimos que variar as concentrações salinas de DNA e os valores de potencial de Morse não altera significativamente as correlações encontradas para intensidades de fluorescência e aberturas médias. O coeficiente máximo de correlação que encontramos para Cy3 variando tanto a concentração salina quanto o potencial de Morse está ao redor de 0.5. Seguindo as

Variação do potencial de Morse $D$ (eV)			
		Cy3	Cy5
	Valor Inicial	Otimização	
AT	30.0	23.4	25.7
CG	70.0	43.0	50.0
$r$		0.5	0.8

**Tabela 3**

Resultados preliminares da otimização dos pares canônicos variando o potencial de Morse e mantendo fixos os potenciais harmônicos e a temperatura. Na última linha,  $r$ , o coeficiente de correlação de Pearson para a otimização.

mesmas variações, encontramos para Cy5 o coeficiente máximo ao redor de 0.8, como pode ser recordado em 9.3.1 e 9.3.2.

Para a variação dos valores de potencial harmônico, vimos para os dois marcadores discrepâncias evidentes entre os resultados. Em todas as ligações, com exceção de Cy5 atado ao par GC, a distinção maior entre os resultados é observada à temperaturas mais baixas, ampliando assim o intervalo de coeficientes de correlação encontrados. Para Cy5 ligado ao par GC (Figura 41h), à temperaturas baixas, apenas uma curva se distingue das outras as quais encontram-se extremamente similares. Porém, a partir da temperatura de 190 K, as curvas similares começam a se diferenciar até chegar à altas temperaturas, enquanto aquela distinta desde temperaturas baixas se mantém a coeficientes mais baixos em relação aos outros. O que notamos é que para Cy3 há uma variação do coeficiente de correlação máximo encontrado comparando-se o comportamento da ligação do marcador aos quatro pares de bases, Figuras 41a, 41e, 41g e 41c. Quando variamos a concentração salina do DNA e o potencial de Morse, encontramos para Cy3 um coeficiente máximo ao redor de 0.5, enquanto para os resultados de potencial harmônico, os coeficientes variam entre 0.5 e 0.7. Por sua vez, Cy5 obteve um coeficiente máximo de correlação ao redor de 0.8 nas variações de concentração salina, potencial de Morse e potencial harmônico.

Após essas etapas, encontramos o melhor resultado para cada interação de empilhamento do marcador com a primeira base nitrogenada e o mantivemos fixo para variar a interação com o segundo vizinho. As diferenças entre os coeficientes encontrados são evidentes, porém menos do que as diferenças presentes nos resultados da variação somente do potencial harmônico da base nitrogenada ligada ao marcador. Para Cy3-GC ligado aos pares AT e CG, o coeficiente máximo encontrado é ao redor de 0.7, enquanto Cy3-CG ligado aos pares AT, CG e GC mostram um coeficiente máximo ao redor de 0.5. Para Cy3-TA ligado a AT, o coeficiente máximo é levemente acima de 0.5. Em todos os casos de Cy5 ligado a qualquer par de bases e seguido por qualquer outro par, encontramos um coeficiente máximo ao redor de 0.8.

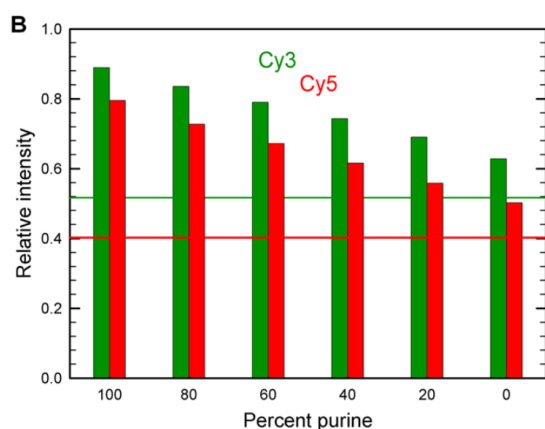
Embora no caso de Cy3 atado ao par CG tenhamos obtido correlações negativas, para todas as outras interações encontramos correlações positivas entre as intensidades de fluorescência e as aberturas médias das sequências. Cy5 apresentou um coeficiente de correlação alto ao redor de 0.8, em todos os casos, o que se aproxima de uma correlação positivamente perfeita, como descrito em 7.4. Para Cy3, encontramos um coeficiente um pouco menor, oscilando entre 0.5

e 0.7, mas ainda assim é uma correlação positiva que torna nossos resultados coerentes com nossa proposta. Além disso, observamos que é possível encontrar coeficientes de correlação, para intensidade de fluorescência e abertura média, melhores em um intervalo de temperaturas específico para cada marcador. Lembrando que as temperaturas desse intervalo são temperaturas de cálculo para o modelo mesoscópico que é um modelo estatístico.

Para verificar se os resultados não foram obtidos por acaso, calculamos o  $p$ -valor para Cy3 e Cy5. Calculamos os coeficientes de correlação de Pearson novamente a partir de valores aleatórios de potencial de Morse  $D$  para os pares canônicos, dentro do intervalo de 0.01 a 0.09 eV. Realizamos esses cálculos 1000 vezes e, ao afinal, dividimos o total de resultados acima da correlação encontrada em 9.3 por 1000. Os scripts para o cálculo do  $p$ -valor podem ser vistos em B.8. Os valores que encontramos para Cy3 e Cy5, respectivamente, foram  $2.6 \times 10^{-2}$  e  $1.6 \times 10^{-2}$ , o que indica que é pouco provável que os valores de parâmetros e Morse que correlacionam bem com as eficiências e luminescência tenham sido obtidos por acaso.

Da etapa de otimização, temos resultados preliminares que indicam uma similaridade com os resultados obtidos em 9.3. Para a variação do potencial de Morse, mantendo fixos os potenciais de empilhamento e a temperatura, encontramos correlações semelhantes às anteriores — ao redor de 0.5 para Cy3 e de 0.8 para Cy5.

De acordo com Kretschy et al. [44] e Agbavwe and Somoza [35], a quantidade de purinas presente em uma sequência de DNA em detrimento da quantidade de pirimidinas contribui para a fluorescência do oligômero. Eles também sugerem que mais pirimidinas do que purinas em uma sequência leva à queda de fluorescência, principalmente, se houver uma quantidade considerável de citosinas [35]. Agbavwe and Somoza [35], em seu artigo no qual coletamos os dados que usamos [35], correlacionam a intensidade de fluorescência pela quantidade de purinas presentes no pentâmero ligado aos marcadores em cada sequência. Eles sugerem que a quantidade de purinas no pentâmero que esteja ligado a Cy3 ou Cy5 pelo terminal 5' do DNA influencia a intensidade de fluorescência sendo um bom preditor, como mostrado na Figura 43.



**Figura 43**

Gráfico da porcentagem de purinas presentes nos pentâmeros ligados aos marcadores fluorescentes em relação às intensidades de fluorescência. Nas duplas barras, as da esquerda são para Cy3 e as da direita para Cy5. Figura retirada da referência [35].

O que observamos dos nossos resultados em relação à sugestão de Agbavwe and Somoza [35] é que, embora a quantidade de purina seja sugerida na literatura e seja um potencial preditor, os valores de intensidade de fluorescência se correlacionam melhor com as aberturas médias

das sequências. Ainda assim, é importante ressaltar que a quantidade de purinas influencia a fluorescência levando em consideração todas as bases nitrogenadas presentes na sequência, não somente um x-mero próximo ao marcador. No caso apresentado por Agbavwe and Somoza [35], eles consideram a quantidade de purinas estritamente ao pentâmero ligado ao marcador desconsiderando as bases ligadas ao longo de toda a sequência após o pentâmero.

## 9.5 Conclusão

Os nossos resultados mostram uma correlação importante entre as intensidades de fluorescência e as aberturas médias calculadas no modelo PB, o que abre a possibilidade de usar este modelo como preditor de eficiência de sondas. Há uma importante dependência da correlação com os parâmetros do modelo, estudamos como os parâmetros de Morse e de empilhamento afetam esta correlação. A partir desta conclusão, elaboramos uma maneira automatizada para obter os parâmetros que maximizam a correlação. Esta parte ainda encontra-se em andamento, mas os resultados preliminares apontam que os parâmetros novos obtidos não divergem de maneira significativa dos parâmetros otimizados com temperaturas de desnaturação do capítulo 8.

## 10 Conclusões

Nos últimos anos, marcadores fluorescentes têm sido muito usados para marcação de ácidos nucleicos e outras biomoléculas [31, 34, 59]. Podemos encontrá-los sendo empregados amplamente nas áreas de biomedicina, biologia molecular, tecnologia, biofísica e farmacologia [31–33, 58]. Sondas fluorescentes têm sido estudadas para o desenvolvimento e aprimoramento de técnicas de diagnóstico e terapia, além de ferramentas tais como reação em cadeia da polimerase (PCR) [34, 35, 44], hibridização fluorescente *in situ* (FISH) [44, 65] e microarranjo de DNA [44]. Dois marcadores que têm ganhado extrema atenção são os marcadores cianina Cy3 e Cy5. Usados como par doador-aceptor na técnica de FRET, esses fluoróforos são sensíveis ao ambiente sequencial em que se encontram [34, 35, 41, 44], possuem uma eficiência quântica de fluorescência mais alta [35, 56] e maior fotoestabilidade [10, 36, 42, 42, 61].

Compartilhando desse interesse, nosso trabalho consistiu em parametrizar os marcadores Cy3 e Cy5 para podermos incorporá-los aos modelos mesoscópicos. A contribuição desses novos parâmetros ampliam as possibilidades de análises de não somente sequências puras de DNA e RNA, como também sequências marcadas com esses e outros fluoróforos. Isso poderia ser aplicado, por exemplo, para o desenvolvimento de sondas de DNA marcadas com marcadores fluorescentes.

Em nosso trabalho, usamos dados publicados por Moreira et al. [45] de valores experimentais de temperaturas de *melting* de 35 sequências de DNA. Essas sequências foram marcadas com Cy3 e Cy5, 19 sequências com o primeiro marcador e 16 com o segundo. Aplicamos o modelo PB para minimizarmos os dados variando somente os parâmetros relacionados com os marcadores, no caso, o potencial de Morse para Cy3 e Cy5, os considerando como um par de bases (XV e YV), e os potenciais harmônicos para as interações de empilhamento do marcador com a base nitrogenada ao qual se encontravam atados. Executamos três rodadas de minimizações para cada marcador como descrito em 8.3. Encontramos parâmetros para Cy3 e Cy5, tanto para potencial de Morse como potencial harmônico, que indicam uma estabilidade dos marcadores quando atados à molécula de DNA, além de reforçarem o comportamento como um par de bases adicional mencionado na literatura [34, 44, 53, 54, 56, 62].

Vimos também que Cy5 estabiliza o duplexo de DNA melhor que Cy3 quando atados ao terminal da molécula. Os perfis de abertura apresentados em 8.4 mostram como o deslocamento das fitas das sequências de DNA com Cy5 são menores do que aquelas com Cy3 quando comparamos a mesma sequência atada a cada um deles, veja uma comparação nas Figuras 30 e 33. Observamos pelos perfis que o ambiente sequencial do duplexo influencia o comportamento dos marcadores como alguns autores sugerem [34, 35, 41, 44].

Com os valores parametrizados que obtivemos para as cianinas Cy3 e Cy5, analisamos outro conjunto de dados para esses fluoróforos. Agbavwe e Somoza publicaram dados experimentais de microarranjo de DNA para sequências marcadas com Cy3 e Cy5 [35]. Eles coletaram dados de intensidade de fluorescência para 1024 fitas simples de DNA atadas à Cy3 e 1024 fitas



simples, idênticas às de Cy3, atadas à Cy5.

Nosso intuito era que se houvesse alguma correlação entre as intensidades de fluorescência e as aberturas médias dessas sequências, poderíamos obter um bom preditor para intensidades de fluorescência. Para esse trabalho, montamos as sequências seguindo o esquema de customização que os autores descreveram em seu artigo. Além de montá-las, também as transformamos de fitas simples para fitas duplas. Calculamos as aberturas médias para todas as sequências e extraímos os valores de abertura média apenas dos marcadores. Com os dois conjuntos de dados — intensidades de fluorescência e aberturas médias dos marcadores —, calculamos o coeficiente de correlação de Pearson variando parâmetros específicos.

Variamos as concentrações salinas de **DNA** e encontramos resultados extremamente similares tanto para Cy3 quanto para Cy5. O mesmo comportamento foi observado quando variamos os valores de potencial de Morse dos marcadores. Para os dois casos, encontramos o coeficiente de correlação máximo para Cy3 ao redor de 0.5 e para Cy5 ao redor de 0.8. Variando o potencial harmônico da interação dos marcadores com a base nitrogenada ao qual estavam ligados, observamos uma maior discrepância entre os resultados. Ainda assim, para Cy5, o coeficiente máximo se manteve ao redor de 0.8, enquanto para Cy3 observa-se uma oscilação entre 0.5 e 0.7, como pode ser visto em 9.3.3. Realizadas essas variações, encontramos o melhor resultado para cada cianina-par de bases e o mantivemos fixo. Variamos então o potencial harmônico do segundo vizinho. Nessa etapa, a discrepância entre os resultados foi menor em comparação de quando variamos apenas a cianina e a base ligada a ela. O coeficiente de correlação máximo para Cy5 se manteve ao redor de 0.8 e para Cy3 se manteve o intervalo entre 0.5 e 0.7.

Vemos que, apesar de alguns coeficientes negativos encontrados para Cy3, as correlações que obtivemos são altas, correlacionando positivamente as intensidades de fluorescência e as aberturas médias para os marcadores. Variar as concentrações salinas e os potenciais de Morse mostrou-se não influenciar significativamente nos coeficientes de correlação, o que foi reforçado pelos resultados encontrados pela maximização dos coeficientes de correlação variando os valores de potencial de Morse. Porém, notamos que a variação dos potenciais harmônicos, tanto do par ligado ao marcador quanto o segundo vizinho, produzem uma oscilação nos coeficientes de correlação. Ainda assim, essa oscilação mantém a correlação positiva e consideravelmente alta.

Com a crescente importância dos marcadores fluorescentes Cy3 e Cy5, compreender melhor essas cianinas torna-se fundamental. Em nosso trabalho, parametrizamos esses marcadores e encontramos resultados consistentes e significativos para a incorporação desses fluoróforos aos modelos mesoscópicos. Verificamos que as cianinas tendem a se comportarem como um par de bases adicional quando acopladas ao terminal da molécula de **DNA** e que as bases nitrogenadas presentes na sequência influenciam o comportamento dos marcadores. Com os resultados obtidos para Cy3 e Cy5 na primeira parte do trabalho, pudemos aplicar o modelo Peyrard-Bishop para encontrarmos as aberturas médias das sequências sintetizadas por Agbavwe e Somoza [35] e correlacioná-las com as intensidades de fluorescência. Encontramos coeficientes de correlação



positivamente altos para ambas as cianinas, indicando que o modelo mesoscópico é capaz de prever as intensidades de fluorescência a partir das sequências de DNA. Dessa forma, temos um potencial preditor para a intensidade de fluorescência que se mostrou mais eficaz do que o preditor sugerido por Agbavwe e Somoza [35]. Os autores tomaram como preditor a correlação das intensidades de fluorescência com a quantidade de purinas presentes exclusivamente no pentâmero ligado aos marcadores. Enquanto nós tomamos as correlações da sequência completa, o que sugere a melhor eficácia do modelo mesoscópico.

## 11 Perspectivas

A modelagem realizada para as cianinas Cy3 e Cy5 dos parâmetros de potencial de Morse — representando as ligações de hidrogênio — e de potencial harmônico — representando as interações de empilhamento — para uso em modelos mesoscópicos amplia as aplicações desses modelos para outras situações.

Os resultados obtidos nesse trabalho nos motivaram a novos objetivos em relação à marcadores fluorescentes e sequências de DNA. Além de ter potencial para estender à sequências de RNA. Nas nossas perspectivas, consideramos a modelagem de outros fluoróforos atados à moléculas de DNA e RNA e a inserção de modificações estruturais de ácidos nucleicos em modelos mesoscópicos.

No momento, já iniciamos estudos de sondas de DNA e RNA modificadas com LNA (*locked nucleic acids*) e marcadas com Cy3, Cy5, PER (*perylene*) e PEP (*phenylethynyl pyrene*). A partir do trabalho apresentado nesta dissertação e dos resultados obtidos até o momento do estudo dessas sondas, acreditamos ser possível desenvolver sondas mais específicas e sensíveis para detecção de oncogenes aplicando modelos mesoscópicos.

## A Dados referentes aos procedimentos do capítulo 8

### A.1 Sequências modificadas de DNA e Temperaturas de *melting*

As tabelas a seguir (4 e 5) apresentam as sequências de DNA usadas nos procedimentos para a parametrização de temperaturas de *melting* descritos em 8.3. Esses dados foram retirados da referência [45]. Na tabela 4, X representa o marcador Cy3 e V seu "par de base"(8.3). Na tabela 5, Y é o marcador Cy5 e V também seu "par de base"(8.3). Estão presentes também os valores experimentais de temperaturas de *melting* obtidos por Moreira et al. [45], na terceira coluna, e os valores preditos de temperaturas de *melting* encontrados em nosso trabalho, quarta coluna. As quatro primeiras temperaturas experimentais da Tabela 4 foram recalculadas por serem referentes à sequências auto-complementares.

**Tabela 4**

Duplexos de DNA modificados com Cy3 usados na parametrização e os valores experimentais de temperaturas de *melting*, retirados da referência [45]. Na última coluna estão os valores preditos de temperaturas de *melting* encontrados em nosso trabalho.

Duplexos de DNA Modificados com Cy3 e $T_m$ 's experimentais e preditas (°C)			
Sequência 5' → 3'	Sequência 3' → 5'	$T_m$ experimental	$T_m$ predita
XACGATCGTV	VTGCTAGCAX	48.3	48.1256
XCTGATCAGV	VGACTAGTCX	44.1	43.8476
XGATGCATCV	VCTACGTAGX	43.9	44.0639
XTCACGTGAV	VAGTGCACTX	44	46.3536
XAGCGTCCA	VTCGCAGGT	46.6	48.6828
XAGCGTCCAV	VTCGCAGGTX	51.9	51.6459
AGCGTCCAV	TCGCAGGTX	47.7	47.1884
XGGCATCGG	VCCGTAGCC	47.5	48.227
GGCATCGGV	CCGTAGCCX	47.6	49.1059
XATCGTTGCTA	VTAGCAACGAT	48.9	48.4679
ATCGTTGCTAV	TAGCAACGATX	47.9	47.3473
XCAGCAGGCAC	VGTCGTCCGTG	56.3	56.4411
CAGCAGGCACV	GTCGTCCGTGX	55.5	55.7213
XGCGAGGAGGCTT	VCGCTCCTCCGAA	64.1	62.7685
XGCGAGGAGGCTTV	VCGCTCCTCCGAAX	66.8	66.3509
GCGAGGAGGCTTV	CGCTCCTCCGAAX	62.9	63.6771
XTTCAAGTATTCG	VAAGTTCATAAGC	48.2	48.6425
XTTCAAGTATTCGV	VAAGTTCATAAGCX	53.3	52.1892
TTCAAGTATTCGV	AAGTTCATAAGCX	49.8	49.6992

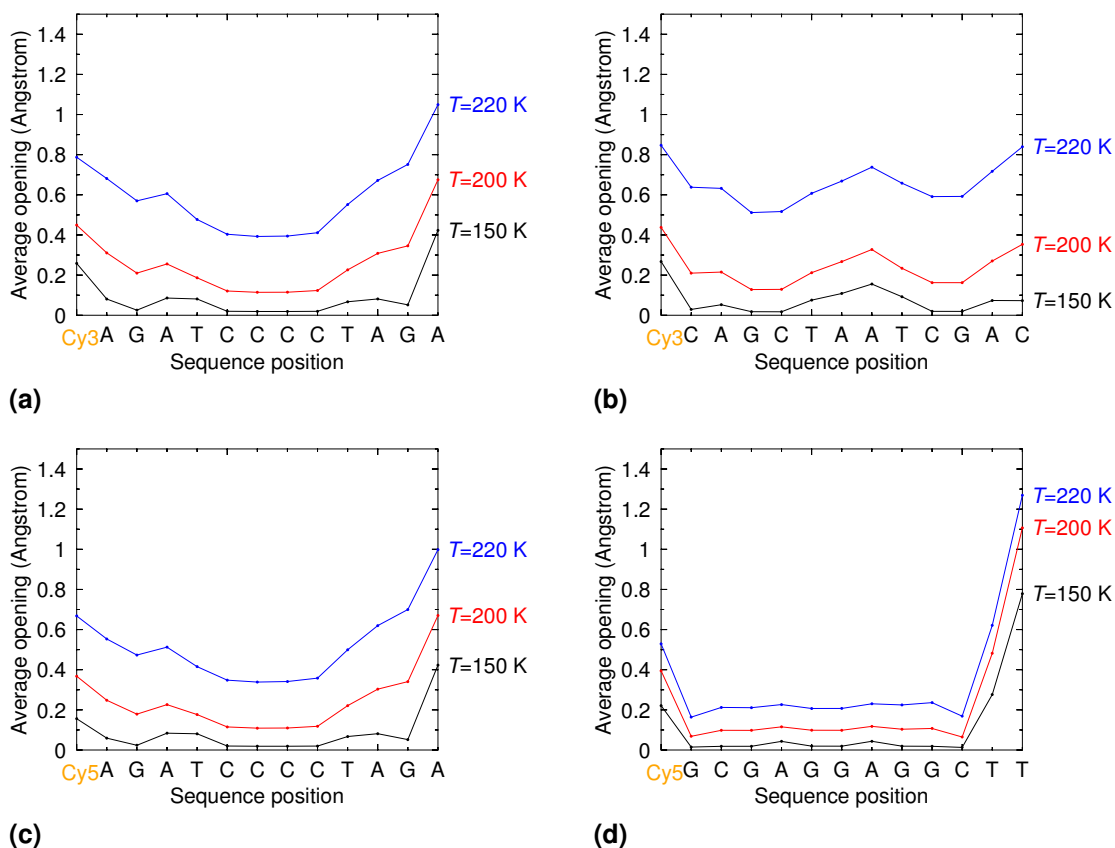
**Tabela 5**

Duplexos de DNA modificados com Cy5 usados na parametrização e os valores experimentais de temperaturas de *melting*, retirados da referência [45]. Na última coluna estão os valores preditos de temperaturas de *melting* encontrados em nosso trabalho.

Duplexos de DNA Modificados com Cy5 e $T_m$ 's experimentais e preditas (°C)			
Sequência 5' → 3'	Sequência 3' → 5'	$T_m$ experimental	$T_m$ predita
YACGATCGTV	VTGCTAGCAY	48	47.9223
YCTGATCAGV	VGACTAGTCY	44.9	45.1937
YGATGCATCV	VCTACGTAGY	42.4	43.2507
YTCACGTGAV	VAGTGCACTY	47.5	48.4327
YAGCGTCCA	VTCGCAGGT	46.2	48.2888
AGCGTCCAV	TCGCAGGTY	47.7	48.1789
YGGCATCGG	VCCGTAGCC	46	47.6018
GGCATCGGV	CCGTAGCCY	47.7	47.4605
YATCGTTGCTA	VTAGCAACGAT	48.4	47.8615
ATCGTTGCTAV	TAGCAACGATY	48.7	48.0102
YCAGCAGGCAC	VGTCGTCCGTG	56.5	56.8375
CAGCAGGCACV	GTCGTCCGTGY	55.7	55.1746
YGCGAGGAGGCTT	VCGCTCCTCCGAA	63.4	62.2628
GCGAGGAGGCTTV	CGCTCCTCCGAAY	62.7	63.4645
YTTCAAGTATTCG	VAAGTTCATAAGC	49	49.2175
TTCAAGTATTCGV	AAGTTCATAAGCY	49.6	49.8611

## A.2 Perfis de abertura média para Cy3 e Cy5

Perfis de abertura média suplementares que não foram discutidos no capítulo 8. Esses perfis foram gerados a partir dos dados referentes à parametrização dos marcadores fluorescentes. Os procedimentos estão descritos em 8.3.



**Figura 44**

Aberturas médias (no eixo vertical), em três temperaturas de cálculo, para a sequência marcada com Cy3 no terminal 5' descrita no eixo horizontal.

## B Dados referentes aos procedimentos do capítulo 9

### B.1 Script para montagem das sequências

Script criado para extrair a sequência dos 5-mers do artigo de Agbavwe and Somoza [35] e montar as sequências seguindo o esquema apresentado pelos autores e presente neste trabalho em 9.1. As intensidades de fluorescência também foram extraídas do mesmo arquivo usando o mesmo script.

1. O script chama o arquivo `dataset.csv`, o qual contém os 5-mers e as intensidades de fluorescência.
2. Lê a primeira linha, que contém o cabeçalho, e a ignora, passando a leitura a partir da segunda linha do arquivo.
3. Inicia um *loop* finito que lê o arquivo a partir da segunda linha. Quebra a linha quando lê uma vírgula e cria um *array* — `@dat`.
4. Com os elementos do *array*, cria um *hash* para "cy3" e "cy5". Para "cy3", o primeiro elemento (o 5-mer) de `@dat` é a chave e o segundo (a intensidade de fluorescência) é o valor. Para "cy5", o quinto elemento (o 5-mer) é a chave e o sexto (a intensidade de fluorescência) é o valor.
5. Cria o arquivo `cy3_int.txt`. Passa por cada chave, a quebra e cria um *array*.
6. Monta a sequência completa e salva no arquivo criado anteriormente.
7. Finalizadas as sequências para Cy3, o script fecha o arquivo.
8. Cria um arquivo para Cy5 (`cy5_int.txt`) e segue o mesmo processo do item anterior para Cy3.
9. Percorre cada chave, a quebra colocando cada caracter em um elemento do *array* e monta a sequência completa.
10. Finalizadas as sequências para Cy5, o script fecha o arquivo e encerra o programa.

```
1  #!/usr/bin/perl
2
3  #criar arquivo com as sequencias completas
4  #do artigo agbavwell
5
6  use strict;
7  use warnings;
8
9  my ($tab, $line);
10 my (%tab, @dat);
11
12 #abre o arquivo, le as linhas e cria um hash dos dados
13 open (DATASET, "dataset.csv");
14     $line = <DATASET>;      #le a primeira linha
15     $line = <DATASET>;      #le a segunda linha
16
17 while ($line = <DATASET>){
```

```

18     @dat = split ('', $line);
19     $tab {"cy3"}{$dat[0]} = $dat[1];
20     $tab {"cy5"}{$dat[4]} = $dat[5];
21 }
22
23 my $s;
24
25 #cria um arquivo com as sequencias e os respectivos valores para cy3
26 open (CY3, '>', "cy3_int.txt") || die "O arquivo nao deu certo.";
27
28 foreach $s (sort keys %{$tab{"cy3"}}){
29     foreach ($s){
30         my @insert = split ('', $s);    #joga cada caracter num elemento do array
31         my $base = "X" . $s . "TTTTTTTTTTTTTTT" . "ACGT" . $insert[0] .
32         "ACGT" . $insert[1] . "ACGT" . $insert[2] . "ACGT" . $insert[3] .
33         "ACGT" . $insert[4] . "TTTTT";
34         print CY3 "$base $tab{"cy3"}{$s}\n";
35     }
36 }
37
38 close CY3;
39
40 #cria um arquivo com as sequencias e os respectivos valores para cy5
41 open (CY5, '>', "cy5_int.txt") || die "O arquivo nao deu certo.";
42
43 foreach $s (sort keys %{$tab{"cy5"}}){
44     foreach ($s){
45         my @insert = split ('', $s);    #joga cada caracter num elemento do array
46         my $base = "Y" . $s . "TTTTTTTTTTTTTTT" . "ACGT" . $insert[0] .
47         "ACGT" . $insert[1] . "ACGT" . $insert[2] . "ACGT" . $insert[3] .
48         "ACGT" . $insert[4] . "TTTTT";
49         print CY5 "$base $tab{"cy5"}{$s}\n";
50     }
51 }
52
53 close CY5;

```

## B.2 Scripts para gerar sequências complementares

Scripts para gerar as sequências complementares para Cy3 e Cy5. Dessa forma, transformamos as 1024 fitas simples de Agbavwe and Somoza [35] em fitas duplas. O terceiro script desta seção reúne em um único arquivo para cada marcador as sequências e seus complementares e as intensidades de fluorescência. Os scripts apresentados estão para Cy3, mas basta trocar o 3 por 5, e X por Y, para termos os mesmos scripts para Cy5. Os procedimentos foram realizados em diretórios específicos e os scripts similares para cada marcador.

### Gerando sequências complementares

1. O script cria o arquivo `comp_cy3.txt` e abre um arquivo que contém somente as fitas simples.
2. Lê cada linha e quebra em um *array*.
3. Faz a tradução — as fitas complementares — e salva no arquivo criado.

4. Após percorrer todo o arquivo com as sequências, o script fecha os dois arquivos e encerra o programa.

```
1  #!/usr/bin/perl
2
3  #gerando as sequencias complementares com cy3
4  #sequencias artigo agbavwell
5
6  use strict;
7  use warnings;
8
9  open (COMP, ">", "comp_cy3.txt") || die "Arquivo nao criado.";
10
11 open (CY3, "seq_cy3.txt");
12
13 while (my $line = <CY3>){
14     chomp $line;
15     my @seq = split ('', $line);
16     $line =~ tr/ACGTX/TGCAV/;
17     print COMP $line;
18     print COMP "\n";
19 }
20
21 close COMP;
22 close CY3;
```

### Montagem do arquivo completo com sequências, complementares e intensidades de fluorescência

1. O script cria o arquivo `cy3\_all.dat` e abre outros três arquivos, resultantes dos scripts anteriores.
2. Lê linha a linha dos três arquivos abertos e quebra a linha do terceiro arquivo em um *array* para guardar as intensidades de fluorescência.
3. O script concatena as informações dos três arquivos — sequências, complementares e intensidades de fluorescência — e salva no arquivo criado, mencionado acima.
4. Fecha os quatro arquivos e encerra o programa.

```
1  #!/usr/bin/perl
2
3  #juntando as sequencias com seus complementares
4  #acrescentando as intensidades de fluorescencia
5
6  use strict;
7  use warnings;
8
9  open (ALL, ">", "cy3_all.dat") || die "Arquivo nao criado.";
10
11 open (SEQS, "seq_cy3.txt");
12
13 open (COMP, "comp_cy3.txt");
14
15 open (INT, "cy3_int.txt");
```



```

16
17 while ((my $line1 = <SEQS>) and (my $line2 = <COMP>) and (my $line3 = <INT>))
18 {
19     chomp $line1;
20     chomp $line2;
21     my @int = split (/ /, $line3);
22     print ALL $line1 . " " . $line2 . " " . $int[1];
23 }
24
25 close SEQS;
26
27 close COMP;
28
29 close INT;
30
31 close ALL;

```

### B.3 Script para calcular aberturas médias

Script criado para calcular as aberturas médias de todas as sequências com Cy3 e Cy5 aplicando o modelo mesoscópico Peyrard-Bishop (PB). Como os procedimentos foram realizados em diretórios diferentes, trocando-se o 3 por 5 tem-se o script para Cy5.

1. Declaração das variáveis.
2. O script chama o programa.
3. As opções para execução do programa estão declaradas.

```

1 #!/bin/bash
2
3 T=$1
4 SEQ=$2
5 CSEQ=$3
6
7 tfreg -m=pb -pbc=0 -t=$T -cutoff=10 -int=-1:200/400 -res=averagey \
8 -par=/usr/share/TfReg/data/dna_pb_1020.par,/usr/share/TfReg/data/dna_pb_cy3.par
9 -o=cy3-$SEQ-$T -pm=2 -dict=X:V \
10 -seq=$SEQ -cseq=$CSEQ

```

### B.4 Script para extrair aberturas médias dos marcadores

Script criado para extrair dos arquivos resultantes de B.3 as aberturas médias referentes apenas aos marcadores Cy3 e Cy5. Para Cy5, basta trocar o 3 por 5.

1. Declaração de variável que chama todos os arquivos com os resultados de abertura média.
2. Cria o arquivo `meds_$T.txt`, em que `$T` é a variável para cada temperatura e guarda os valores de aberturas médias para os marcadores.
3. Passa por todos os arquivos de abertura média, quebra cada linha e salva o valor de abertura média do marcador.
4. Fecha os arquivos e encerra o programa.

```

1  #!/usr/bin/perl
2
3  #valores de <y> do cy3
4
5  use strict;
6  use warnings;
7
8  my $T = 150;
9
10 my @temps = glob ("../temperaturas/cy3-*$T.dat");
11
12 open (MEDS, ">", "meds_$T.txt") || die "Arquivo nao criado.";
13
14 for (my $i = 0; $i <= $#temps; $i++){
15     open (TEMPS, "<", $temps[$i]) || die "Fez errado!";
16     my $line = <TEMPS>;
17     chomp $line;
18     my @split = split (/ /, $line);
19     my $file = $temps[$i];
20     my @name = split (/-/ , $file);
21     print MEDS "$name[1]" . " " . "$split[1]\n";
22 }
23
24 close TEMPS;
25
26 close MEDS;

```

## B.5 Script para calcular coeficiente de correlação de Pearson

Script criado para calcular o coeficiente de correlação de Pearson para as intensidades de fluorescência e as aberturas médias para os marcadores. Mais uma vez, troca-se 3 por 5 para proceder com Cy5.

1. Abre o arquivo com as aberturas médias dos marcadores.
2. Lê cada linha do arquivo, quebra a linha e guarda os valores em um *array* — @temp.
3. Abre o arquivo com as intensidades de fluorescência.
4. Lê cada linha do arquivo, quebra a linha e guarda os valores em um *array* — @int.
5. Inicia os procedimentos para a correlação de Pearson.
6. Cria arquivo para guardar os coeficientes de correlação calculados.
7. Cria arquivo para guardar as aberturas médias e as intensidades de fluorescência juntas.
8. Soma todos os valores para intensidade de fluorescência (\$soma\_i) e abertura média (\$soma\_t) e salva no arquivo.
9. Declaração de variáveis.
10. Calcula a diferença das intensidades de fluorescência pela média (\$dif\_i) e das aberturas médias pela média (\$dif\_t) e calcula o produto dessas diferenças (@prod).
11. Soma os produtos (\$soma\_tot).

12. Calcula o quadrado da diferença pela média para as intensidades de fluorescência (@quad\_i) e as aberturas médias (@quad\_t).
13. Soma os quadrados das intensidades de fluorescência (\$soma\_qi) e das aberturas médias (\$soma\_qt).
14. Calcula a raiz quadrada da soma dos quadrados das intensidades de fluorescência (\$raiz\_i) e das aberturas médias (\$raiz\_t).
15. Calcula o coeficiente de correlação de Pearson.
16. Salva o resultado no arquivo "cor\_\$T.txt" e o fecha.
17. Imprime na linha de comando o coeficiente de correlação e encerra o programa.

```
1  #!/usr/bin/perl
2
3  #Pearson Correlation
4  #medias e intensidades para cy3
5  #valor medio de cy3
6
7  use strict;
8  use warnings;
9
10 my $T = 350;
11
12 open (TEMP, "<", "meds_$T.txt") || die "Arquivo nao abre.";
13
14 my @temp;
15
16 while (my $line = <TEMP>){
17     chomp $line;
18     my @split = split (/ /, $line);
19     push (@temp, $split[1]);
20 }
21
22 close TEMP;
23
24 open (INT, "<", "../..cy3_int.txt") || die "Arquivo nao abre.";
25
26 my @int;
27
28 while (my $linha = <INT>){
29     chomp $linha;
30     my @insert = split (/ /, $linha);
31     push (@int, $insert[1]);
32 }
33
34 close INT;
35
36 #CORRELACAO DE PEARSON
37
38 open (PEARSON, ">", "cor_$T.txt") || die "Arquivo nao pode ser criado.";
39
40 my $size = scalar @temp;
41
42 my $soma_t = 0;
```

```

43 my $soma_i = 0;
44
45 open (POT, ">scatter_$T.dat");
46
47 #soma todos os elementos e faz a media
48 for (my $i = 0; $i < $size; $i++){
49     $soma_t += $temp[$i];
50     $soma_i += $int[$i];
51     print POT "$temp[$i] $int[$i]\n";
52 }
53
54 close POT;
55
56 my $med_t = $soma_t / $size;
57 my $med_i = $soma_i / $size;
58
59 my @prod;
60 my $prod = 0;
61 my $dif_t = 0;
62 my $dif_i = 0;
63
64 #diferenca dos elementos pela media e produto dessas diferencas
65 for (my $i = 0; $i < $size; $i++){
66     $dif_t = $temp[$i] - $med_t;
67     $dif_i = $int[$i] - $med_i;
68     $prod = $dif_t * $dif_i;
69     push @prod, "$prod";
70 }
71
72 my $soma_tot = 0;
73
74 #soma os produtos de x e de y
75 for (my $i = 0; $i < $size; $i++){
76     $soma_tot += $prod[$i];
77 }
78
79 my ($quad_t, $quad_i);
80 my (@quad_t, @quad_i);
81
82 #quadrado da diferenca pela media
83 for (my $i = 0; $i < $size; $i++){
84     $dif_t = $temp[$i] - $med_t;
85     $quad_t = $dif_t ** 2;
86     push @quad_t, $quad_t;
87 }
88
89 for (my $i = 0; $i < $size; $i++){
90     $dif_i = $int[$i] - $med_i;
91     $quad_i = $dif_i ** 2;
92     push @quad_i, $quad_i;
93 }
94
95 my $soma_qt = 0;
96 my $soma_qi = 0;
97
98 #soma dos quadrados de temp e de int
99 for (my $i = 0; $i < $size; $i++){
100     $soma_qt += $quad_t[$i];

```

```

101 }
102
103 for (my $i = 0; $i < $size; $i++){
104     $soma_qi += $quad_i[$i];
105 }
106
107 my ($raiz_t, $raiz_i);
108
109 #raiz quadrada da soma dos quadrados de temp e de int
110 $raiz_t = sqrt $soma_qt;
111 $raiz_i = sqrt $soma_qi;
112
113 my $raizes;
114
115 #produto das raizes de temp e int
116 $raizes = $raiz_t * $raiz_i;
117
118 my $pearson;
119
120 #resultado da correlacao de Pearson
121 $pearson = $soma_tot / $raizes;
122
123 print PEARSON "$pearson\n";
124
125 close PEARSON;
126
127 print "A correlacao de Pearson e $pearson.\n";

```

## B.6 Coeficientes de correlação $\times$ Temperaturas de cálculo

**Tabela 6**

Coeficientes de correlação de Pearson das 1024 sequências de DNA marcadas com Cy3 e com Cy5 referentes à cada temperatura de cálculo usada em nosso trabalho.

Temperaturas x Coeficientes de Correlação		
Temperaturas (K)	Sequências com Cy3	Sequências com Cy5
150	0.403736235836775	0.749340765497487
160	0.421167175542713	0.76095468053743
170	0.439306322732145	0.770605583506817
180	0.456738071468838	0.775654037719196
190	0.471570795761116	0.773445020803393
200	0.481875442715967	0.762430263738496
210	0.486447862113469	0.743204418657716
220	0.485377458308387	0.718429044047636
230	0.479898618616161	0.691513682226625
240	0.471681941194221	0.665164086654292
250	0.462165392559777	0.640749462747978
260	0.452263512497788	0.618433470895638
270	0.442365383528589	0.597643976776455
280	0.432447185186504	0.577580227421955
290	0.422708870719465	0.558351616057148
300	0.414179010394348	0.541937074643673
310	0.407805553346256	0.530092583449722
320	0.403957642221066	0.523035580999466
330	0.402426814294252	0.520067390455618
340	0.402390025023674	0.520026915967525
350	0.402737678369181	0.521710366848259

## B.7 Script para otimização das correlações de Pearson

Após calcular os coeficientes de correlação de Pearson para as intensidades de fluorescência e as aberturas médias dos marcadores Cy3 e Cy5, otimizamos esses coeficientes.

### Script bash para executar subrotina

1. Executa o programa em Perl.
2. Recebe os argumentos de 1 a 7.

```
1 #!/bin/bash
2 #SBATCH --mem=10000 --reservation=bioinf_17
3
4 srun perl minimization.pl $1 $2 $3 $4 $5 $6 $7;
```

### Script para executar o módulo Perl

1. Cria arquivo para salvar resultado.
2. Chama a subrotina e recebe três argumentos pela linha de comando.
3. Guarda os argumentos em um *array*.
4. Declaração da escala.
5. Executa módulo de maximização.
6. Guarda resultado.
7. Salva resultado e encerra o programa.

```
1 #!/usr/bin/perl
2
3 #algoritmo de minimizacao
4 #usando o metodo nelder-mead
5 #metodo downhill simplex
6
7 use lib '.';
8 use lib "/users/miranda/lib/lib/perl5/site_perl/5.26.1";
9 use correlation;
10 use Math::Amoeba qw(MinimiseND);
11
12 open (RES, ">output/res.dat");
13
14 cor ($ARGV[0], $ARGV[1], $ARGV[2]); #valores de Morse de AT e CG e temperatura
15
16 my @guess = ($ARGV[0], $ARGV[1], $ARGV[2]);
17 my @scale = (0.01, 0.01, 10);
18 my ($x, $y) = MinimiseND(\@guess, \@scale, \&cor, 1e-7, 300);
19
20 system ("echo $y >> output/resultado.txt");
21
22 print RES "resultado: ";
23 print RES "(", join(',', @{$x}), ") = $y\n";
24
25 close RES;
```

## Subrotina para gerar as aberturas médias e calcular os coeficientes de correlação

1. Inicia a subrotina
2. Declaração de variáveis.
3. Variáveis recebem argumentos da linha de comando.
4. Cria pasta.
5. Declaração de variáveis.
6. Criação de arquivos para guardar progresso da execução do programa.
7. Cria arquivo para parâmetros do potencial de Morse.
8. Executa programa para calcular as aberturas médias.
9. Abre arquivo.
10. Lê cada linha, quebra a linha e guarda a intensidade de fluorescência em um *array* — @int.
11. Fecha arquivo.
12. Abre arquivo.
13. Lê cada linha, quebra a linha e guarda as aberturas médias do marcador em um *array* — @ave.
14. Fecha o arquivo.
15. Inicia os procedimentos para a correlação de Pearson.
16. Declaração de variáveis.
17. Soma todos os valores para intensidade de fluorescência (\$soma\_i) e abertura média (\$soma\_a).
18. Calcula a média das intensidades de fluorescência (\$med\_a) e das aberturas médias (\$med\_a).
19. Declaração de variáveis.
20. Calcula a diferença das intensidades de fluorescência pela média (\$dif\_i) e das aberturas médias pela média (\$dif\_a).
21. Calcula o quadrado da diferença pela média para as intensidades de fluorescência (\$quad\_i) e as aberturas médias (\$quad\_a).
22. Calcula o produto dessas diferenças (@prod).
23. Declaração de variáveis.
24. Soma os produtos (\$soma\_tot) das intensidades de fluorescência e das aberturas médias.
25. Soma os quadrados das intensidades de fluorescência (\$soma\_qi) e das aberturas médias (\$soma\_qa).



26. Declaração de variáveis.
27. Calcula a raiz quadrada da soma dos quadrados das intensidades de fluorescência (\$raiz\_i) e das aberturas médias (\$raiz\_a).
28. Calcula o produto das raízes das intensidades de fluorescência e das aberturas médias.
29. Declaração de variáveis.
30. Calcula o coeficiente de correlação de Pearson e muda o sinal do resultado.
31. Salva o coeficiente de correlação de Pearson.
32. Encerra a subrotina.

```

1  #!/usr/bin/perl
2
3  #recebe parametros e executa tfreq
4  #calcula valores de abertura media
5  #calcula correlacao de pearson
6  #para intensidades de fluorescencia e abertura media do marcador
7
8  package correlation;
9  require Exporter;
10
11 our @ISA = qw (Exporter);
12 our @EXPORT = qw (cor);
13
14 sub cor {
15
16 my $file = $ARGV[5];      #sequencias
17 my $dna = $ARGV[3];      #parametros para concentracao de dna
18 my $cy = $ARGV[4];      #parametros para cianina
19 my $T;                  #temperatura
20 my %morse;              #valores de morse
21
22 $morse{"AT"} = shift;
23 $morse{"CG"} = shift;
24 $T = shift;
25
26 system ("mkdir -p output");
27
28 my $D_AT=$morse{"AT"};
29 my $D_CG=$morse{"CG"};
30
31 system ("echo $D_AT $D_CG >> output/morse-history.txt");
32 system ("echo $T >> output/temp-history.txt");
33
34 open (PAR, ">output/morse.par");
35
36 print PAR "valores\n";
37
38 foreach my $bp (keys %morse)
39 {
40     print PAR "$bp:morse.D $morse{$bp}\n";
41 }
42

```

```

43 my $pot = "output/morse.par";
44
45 #rodando tfreg
46 system ("/users/bioinf/programs/tfreg-1072 -m=pb -pbc=0 -t=$T -cutoff=10
47 -int=-1:200/400 -data=$file -res=averagey -par=$dna,$cy,$pot -o=output/cy3
48 -pm=2 -dict=X:V &> output/cy3.echo");
49
50 #arquivando apenas as intensidades de fluorescencia
51 open (INT, $ARGV[6]);      #abre o arquivo com sequencias, complementares e valores
52
53 my @int;
54
55 while (my $line = <INT>)
56 {
57     chomp $line;
58     my @dat = split (/ /, $line);
59     push (@int, $dat[2]);
60 }
61
62 close INT;
63
64 #arquivando apenas as aberturas medias
65 open (AVE, "output/cy3.dat");
66
67 my @ave;
68
69 while (my $liney = <AVE>)
70 {
71     chomp $liney;
72     my @array = split (/ /, $liney);
73     push (@ave, $array[3]);
74 }
75
76 close AVE;
77
78 #correlacao de pearson
79 my $size = scalar @ave;
80
81 #soma todos os elementos e faz a media
82 my $soma_a = 0;
83 my $soma_i = 0;
84
85 for (my $i = 0; $i < $size; $i++)
86 {
87     $soma_a += $ave[$i];      #a para abertura media
88     $soma_i += $int[$i];      #i para intensidade de fluorescencia
89 }
90
91 my $med_a = $soma_a / $size;  #media das aberturas medias
92 my $med_i = $soma_i / $size;  #media das intensidades de fluorescencia
93
94 my @prod;
95 my $prod = 0;
96 my $dif_a = 0;
97 my $dif_i = 0;
98 my ($quad_a, $quad_i);
99 my (@quad_a, @quad_i);
100

```

```

101 #diferença dos elementos pela média e produto dessas diferenças
102 #quadrado da diferença pela média
103 for (my $i = 0; $i < $size; $i++){
104     $dif_a = $ave[$i] - $med_a;
105     $quad_a = $dif_a ** 2;
106     push @quad_a, $quad_a;
107     $dif_i = $int[$i] - $med_i;
108     $quad_i = $dif_i ** 2;
109     push @quad_i, $quad_i;
110     $prod = $dif_a * $dif_i;
111     push @prod, "$prod";
112 }
113
114 #soma os produtos de ave e de int
115 #soma dos quadrados de ave e de int
116 my $soma_tot = 0;
117 my $soma_qa = 0;
118 my $soma_qi = 0;
119
120 for (my $i = 0; $i < $size; $i++){
121     $soma_tot += $prod[$i];
122     $soma_qa += $quad_a[$i];
123     $soma_qi += $quad_i[$i];
124 }
125
126 #raiz quadrada da soma dos quadrados de ave e de int
127 my ($raiz_a, $raiz_i);
128
129 $raiz_a = sqrt $soma_qa;
130 $raiz_i = sqrt $soma_qi;
131
132 #produto das raízes de ave e int
133 my $raizes;
134
135 $raizes = $raiz_a * $raiz_i;
136
137 #resultado da correlação de Pearson
138 my $pearson;
139 my $cor;
140
141 $cor = $soma_tot / $raizes;
142 $pearson = - $cor;
143
144 system("echo $pearson >> output/morse-history.txt");
145
146 return $pearson;
147 }
148
149 1;

```

## B.8 Script para cálculo do $p$ -valor

Script para calcular as aberturas médias e calcular os coeficientes de correlação de Pearson a partir de valores randômicos para os potenciais de Morse para os pares de bases AT e CG. Para o procedimento com o marcador Cy5, basta trocar o 3 por 5 e X por Y.

## Script bash para executar programa em Perl

1. Repassa o comando de 1000 execuções.
2. Executa o programa em Perl.
3. Um único argumento na linha de comando.

```
1 #!/bin/bash
2 #SBATCH --reservation=biointf_17
3
4 for M in {1..1000};
5 do
6   srun perl p-valor.pl $1
7 done
```

## Script para obtenção dos coeficientes de correlação para cálculo do $p$ -valor

1. Declaração de variáveis.
2. Randomização dos valores para os potenciais de Morse dos pares de bases AT e CG.
3. Cria pasta.
4. Cria arquivo para parâmetros do potencial de Morse.
5. Executa programa para calcular as aberturas médias.
6. Abre arquivo.
7. Lê cada linha, quebra a linha e guarda a intensidade de fluorescência em um *array* — @int.
8. Fecha arquivo.
9. Abre arquivo.
10. Lê cada linha, quebra a linha e guarda as aberturas médias do marcador em um *array* — @ave.
11. Fecha o arquivo.
12. Inicia os procedimentos para a correlação de Pearson.
13. Declaração de variáveis.
14. Soma todos os valores para intensidade de fluorescência (\$soma\_i) e abertura média (\$soma\_a).
15. Calcula a média das intensidades de fluorescência (\$med\_i) e das aberturas médias (\$med\_a).
16. Declaração de variáveis.
17. Calcula a diferença das intensidades de fluorescência pela média (\$dif\_i) e das aberturas médias pela média (\$dif\_a).

18. Calcula o quadrado da diferença pela média para as intensidades de fluorescência (\$squad\_i) e as aberturas médias (\$squad\_a).
19. Calcula o produto dessas diferenças (@prod).
20. Declaração de variáveis.
21. Soma os produtos (\$soma\_tot) das intensidades de fluorescência e das aberturas médias.
22. Soma os quadrados das intensidades de fluorescência (\$soma\_qi) e das aberturas médias (\$soma\_qa).
23. Declaração de variáveis.
24. Calcula a raiz quadrada da soma dos quadrados das intensidades de fluorescência (\$raiz\_i) e das aberturas médias (\$raiz\_a).
25. Calcula o produto das raízes das intensidades de fluorescência e das aberturas médias.
26. Declaração de variáveis.
27. Calcula o coeficiente de correlação de Pearson.
28. Salva o coeficiente de correlação de Pearson.
29. Encerra o programa.

```

1  #!/usr/bin/perl
2
3  #calcular p-valor
4  #randomizar valores de pontecial de Morse para AT e CG
5  #calcular coeficiente de correlacao de Pearson
6  #abertura media de Cy3 x intensidades de fluorescencia
7
8  my $file = "cy3_seq.dat";           #sequencias
9  my $dna = "dna_pb_1020.par";       #parametros para concentracao de dna
10 my $cy = "dna_pb_cy3.par";         #parametros para cianina
11 my $T = $ARGV[0];                 #temperatura
12 my %morse;                         #valores de morse
13
14 $morse{"AT"} = rand(0.09) + 0.01;
15 $morse{"CG"} = rand(0.09) + 0.01;
16
17 system ("mkdir -p output");
18
19 open (PAR, ">>output/morse-$T.par");
20
21 print PAR "valores\n";
22
23 foreach my $bp (keys %morse)
24 {
25     print PAR "$bp:morse.D $morse{$bp}\n";
26 }
27
28 my $pot = "output/morse-$T.par";
29
30 #rodando tfreq

```

```

31 system ("/users/bioinf/programs/tfreg-1072 -m=pb -pbc=0 -t=$T -cutoff=10
32 -int=-1:200/400 -data=$file -res=averagey -par=$dna,$cy,$spot -o=output/cy3-$T
33 -pm=2 -dict=X:V &> output/cy3-$T.echo");
34
35 #arquivando apenas as intensidades de fluorescencia
36 open (INT, "cy3_int.dat"); #abre o arquivo com sequencias, complementares e valores
37
38 my @int;
39
40 while (my $line = <INT>)
41 {
42     chomp $line;
43     my @dat = split (/ /, $line);
44     push (@int, $dat[2]);
45 }
46
47 close INT;
48
49 #arquivando apenas as aberturas medias
50 open (AVE, "output/cy3-$T.dat");
51
52 my @ave;
53
54 while (my $liney = <AVE>)
55 {
56     chomp $liney;
57     my @array = split (/ /, $liney);
58     push (@ave, $array[3]);
59 }
60
61 close AVE;
62
63 #correlacao de pearson
64 open (PEARSON, ">>output/pearson-$T.dat") || die "Arquivo nao pode ser criado.";
65
66 my $size = scalar @ave;
67
68 #soma todos os elementos e faz a media
69 my $soma_a = 0;
70 my $soma_i = 0;
71
72 for (my $i = 0; $i < $size; $i++)
73 {
74     $soma_a += $ave[$i];           #a para abertura media
75     $soma_i += $int[$i];          #i para intensidade de fluorescencia
76 }
77
78 my $med_a = $soma_a / $size;      #media das aberturas medias
79 my $med_i = $soma_i / $size;      #media das intensidades de fluorescencia
80
81 my @prod;
82 my $prod = 0;
83 my $dif_a = 0;
84 my $dif_i = 0;
85 my ($quad_a, $quad_i);
86 my (@quad_a, @quad_i);
87
88 #diferenca dos elementos pela media e produto dessas diferencas

```

```

89 #quadrado da diferenca pela media
90 for (my $i = 0; $i < $size; $i++){
91     $dif_a = $ave[$i] - $med_a;
92     $quad_a = $dif_a ** 2;
93     push @quad_a, $quad_a;
94     $dif_i = $int[$i] - $med_i;
95     $quad_i = $dif_i ** 2;
96     push @quad_i, $quad_i;
97     $prod = $dif_a * $dif_i;
98     push @prod, "$prod";
99 }
100
101 #soma os produtos de ave e de int
102 #soma dos quadrados de ave e de int
103 my $soma_tot = 0;
104 my $soma_qa = 0;
105 my $soma_qi = 0;
106
107 for (my $i = 0; $i < $size; $i++){
108     $soma_tot += $prod[$i];
109     $soma_qa += $quad_a[$i];
110     $soma_qi += $quad_i[$i];
111 }
112
113 #raiz quadrada da soma dos quadrados de ave e de int
114 my ($raiz_a, $raiz_i);
115
116 $raiz_a = sqrt $soma_qa;
117 $raiz_i = sqrt $soma_qi;
118
119 #produto das raizes de ave e int
120 my $raizes;
121
122 $raizes = $raiz_a * $raiz_i;
123
124 #resultado da correlacao de Pearson
125 my $pearson;
126
127 $pearson = $soma_tot / $raizes;
128
129 print PEARSON "$pearson\n";
130
131 close PEARSON;

```

## Referências

- [1] Bernard Valeur. *Molecular fluorescence: principles and applications*. Wiley-VCH Verlag GmbH, 2001.
- [2] Joseph R Lakowicz. *Principles of fluorescence spectroscopy*. Springer Science & Business Media, 2006.
- [3] Sujay Ray, Julia R Widom, and Nils G Walter. Life under the microscope: Single-molecule fluorescence highlights the rna world. *Chemical reviews*, 118(8):4120–4155, 2018.
- [4] John C Croney, David M Jameson, and Robert P Learmonth. Fluorescence spectroscopy in biochemistry: teaching basic principles with visual demonstrations. *Biochemistry and Molecular Biology Education*, 29(2):60–65, 2001.
- [5] Hugh D Young and Roger A Freedman. *Física IV: ótica e física moderna*. Pearson Education do Brasil, 2009.
- [6] Jack A Tuszynski and Michal Kurzynski. *Introduction to molecular biophysics*. CRC press, 2003.
- [7] Peter Atkins and Julio De Paula. *Físico-química*. LTC, 2018.
- [8] D Halliday, R Resnick, and J Walker. Fundamentos de física, 8a. edição, vol. 2, ltc, 2008.
- [9] <https://pt.wikipedia.org/wiki/Ficheiro:Comprimento-de-onda.png>.
- [10] Marcia Levitus and Suman Ranjit. Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Quarterly reviews of biophysics*, 44(01):123–151, 2011.
- [11] Bernardo G Moreira, Yong You, Mark A Behlke, and Richard Owczarzy. Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. *Biochemical and Biophysical Research Communications*, 327(2):473–484, 2005.
- [12] David L Nelson and Michael M Cox. *Princípios de Bioquímica de Lehninger*. Artmed Editora, 2014.
- [13] Tom Strachan and Andrew Read. *Genética Molecular Humana*. Artmed Editora, 2013.
- [14] Benjamin A Pierce. *Genética: um enfoque conceitual*. Guanabara Koogan, 2013.
- [15] Arthur M Lesk. *Introduction to genomics*. Oxford University Press, 2007.
- [16] Bruce R Korf. Basic genetics. *Primary Care: Clinics in Office Practice*, 31(3):461–478, 2004.
- [17] Arnaldo Zaha, Henrique Bunselmeyer Ferreira, and Luciane MP Passaglia. *Biologia Molecular Básica-5*. Artmed Editora, 2014.
- [18] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.



- [19] Shu-ichi Nakano, Daisuke Miyoshi, and Naoki Sugimoto. Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chemical reviews*, 114(5): 2733–2758, 2013.
- [20] Charlotte W Pratt and Kathleen Cornely. *Essential Biochemistry*. Guanabara Koogan, 2004.
- [21] Lauren A Levine, Matthew Junker, Myranda Stark, and Dustin Greenleaf. A DNA melting exercise for a large laboratory class. *Journal of Chemical Education*, 92:1928–1931, 2015. doi:[10.1021/acs.jchemed.5b00049](https://doi.org/10.1021/acs.jchemed.5b00049).
- [22] William S. Klug et al. *Concepts of Genetics*. Pearson Benjamin Cummings, 2012.
- [23] [https://www.brainkart.com/article/Denaturation-of-DNA\\_27539/](https://www.brainkart.com/article/Denaturation-of-DNA_27539/).
- [24] Luis A Marky and Kenneth J Breslauer. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers: Original Research on Biomolecules*, 26(9):1601–1620, 1987.
- [25] Sherrie Schreiber-Gosche and Robert A Edwards. Thermodynamics of oligonucleotide duplex melting. *Journal of Chemical Education*, 86(5):644, 2009.
- [26] Kathleen P Howard. Thermodynamics of DNA duplex formation: A biophysical chemistry laboratory experiment. *J. Chem. Educ*, 77(11):1469, 2000.
- [27] John SantaLucia. Physical principles and visual-omp software for optimal pcr design. In *PCR primer design*, pages 3–33. Springer, 2007.
- [28] Lars Kaderali and Alexander Schliep. Selecting signature oligonucleotides to identify organisms using dna arrays. *Bioinformatics*, 18(10):1340–1349, 2002.
- [29] Richard Owczarzy. Melting temperatures of nucleic acids: discrepancies in analysis. *Biophysical Chemistry*, 117(3):207–215, 2005.
- [30] Jean-Louis Mergny and Laurent Lacroix. Analysis of thermal melting curves. *Oligonucleotides*, 13(6):515–537, 2003.
- [31] HA Shindy. Fundamentals in the chemistry of cyanine dyes: A review. *Dyes and Pigments*, 145:505–513, 2017.
- [32] H Hilal and JA Taylor. Cyanine dyes for the detection of double stranded dna. *Journal of biochemical and biophysical methods*, 70(6):1104–1108, 2008.
- [33] Diego S Pisoni, Letícia Todeschini, Antonio César A Borges, César L Petzhold, Fabiano S Rodembusch, and Leandra F Campo. Symmetrical and asymmetrical cyanine dyes. synthesis, spectral properties, and bsa association study. *The Journal of organic chemistry*, 79(12):5511–5520, 2014.
- [34] Nicole Kretschy and Mark M Somoza. Comparison of the sequence-dependent fluorescence of the cyanine dyes Cy3, Cy5, DyLight DY547 and DyLight DY647 on single-stranded DNA. *PloS one*, 9(1):e85605, 2014.

- [35] Christy Agbavwe and Mark M Somoza. Sequence-dependent fluorescence of cyanine dyes on microarrays. *PLoS One*, 6(7):e22177, 2011.
- [36] Ondřej Kroutil, Ingrid Romancová, Miroslav Šíp, and Zdeněk Chval. Cy3 and Cy5 dyes terminally attached to 5' C end of DNA: Structure, dynamics, and energetics. *J. Phys. Chem. B*, 118(47):13564–13572, 2014.
- [37] Jonathan Ouellet, Stephanie Schorr, Asif Iqbal, Timothy J Wilson, and David MJ Lilley. Orientation of cyanine fluorophores terminally attached to DNA via long, flexible tethers. *Biophys. J.*, 101(5):1148–1154, 2011.
- [38] Alexandre S Boutorine, Darya S Novopashina, Olga A Krasheninina, Karine Nozeret, and Alya G Venyaminova. Fluorescent probes for nucleic acid visualization in fixed and live cells. *Molecules*, 18(12):15357–15397, 2013.
- [39] Bruce A Armitage. Cyanine dye–DNA interactions: intercalation, groove binding, and aggregation. In *DNA binders and related subjects*, pages 55–76. Springer, 2005.
- [40] Fahad Rashid, Vlad-Stefan Raducanu, Manal S Zaher, Muhammad Tehseen, Satoshi Habuchi, and Samir M Hamdan. Initial state of dna-dye complex sets the stage for protein induced fluorescence modulation. *Nature communications*, 10(1):2104, 2019.
- [41] Michael J Morten, Sergio G Lopez, I Emilie Steinmark, Aidan Rafferty, and Steven W Magennis. Stacking-induced fluorescence increase reveals allosteric interactions through dna. *Nucleic acids research*, 46(21):11618–11626, 2018.
- [42] Helen Hwang and Sua Myong. Protein induced fluorescence enhancement (PIFE) for probing protein–nucleic acid interactions. *Chemical Society Reviews*, 43(4):1221–1229, 2014.
- [43] Fabio D Steffen, Roland KO Sigel, and Richard Börner. An atomistic view on carbocyanine photophysics in the realm of rna. *Physical Chemistry Chemical Physics*, 18(42):29045–29055, 2016.
- [44] Nicole Kretschy, Matej Sack, and Mark M Somoza. Sequence-dependent fluorescence of Cy3-and Cy5-labeled double-stranded DNA. *Bioconjugate Chemistry*, 27(3):840–848, 2016.
- [45] Bernardo G Moreira, Yong You, and Richard Owczarzy. Cy3 and Cy5 dyes attached to oligonucleotide terminus stabilize DNA duplexes: Predictive thermodynamic model. *Biophys. Chem.*, 198:36–44, 2015. doi:[10.1016/j.bpc.2015.01.001](https://doi.org/10.1016/j.bpc.2015.01.001).
- [46] Lauren A Ernst, Ravinder K Gupta, Ratnakar B Mujumdar, and Alan S Waggoner. Cyanine dye labeling reagents for sulfhydryl groups. *Cytometry: The Journal of the International Society for Analytical Cytology*, 10(1):3–10, 1989.
- [47] Alan Waggoner. Optical probes of membrane potential. *The Journal of membrane biology*, 27(1):317–334, 1976.
- [48] <http://www.buildingbox.co.uk/slide5/>.
- [49] Melissa Massey, W Russ Algar, and Ulrich J Krull. Fluorescence resonance energy transfer (fret) for dna biosensors: Fret pairs and förster distances for various dye–dna conjugates. *Analytica chimica acta*, 568(1-2):181–189, 2006.

- [50] Frances M Hamer. *The cyanine dyes and related compounds*, volume 60. John Wiley & Sons, 2009.
- [51] Adrian Fegan, Pravin S Shirude, and Shankar Balasubramanian. Rigid cyanine dye nucleic acid labels. *Chem. Commun.*, (17):2004–2006, 2008.
- [52] Asif Iqbal, Sinan Arslan, Burak Okumus, Timothy J Wilson, Gerard Giraud, David G Norman, Taekjip Ha, and David MJ Lilley. Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci. USA*, 105(32):11176–11181, 2008.
- [53] Asif Iqbal, Lihua Wang, Katherine C Thompson, David MJ Lilley, and David G Norman. The structure of cyanine 5 terminally attached to double-stranded DNA: implications for FRET studies. *Biochem.*, 47(30):7857–7862, 2008.
- [54] David G Norman, Richard J Grainger, Dusan Uhrin, and David MJ Lilley. Location of cyanine-3 on double-stranded DNA: importance for fluorescence resonance energy transfer studies. *Biochemistry*, 39(21):6317–6324, 2000.
- [55] Lucy M Hall, Marta Gerowska, and Tom Brown. A highly fluorescent dna toolkit: synthesis and properties of oligonucleotides containing new cy3, cy5 and cy3b monomers. *Nucleic acids research*, 40(14):e108–e108, 2012.
- [56] Justin Spiriti, Jennifer K Binder, Marcia Levitus, and Arjan Van Der Vaart. Cy3-DNA stacking interactions strongly depend on the identity of the terminal basepair. *Biophys. J.*, 100(4):1049–1057, 2011.
- [57] Heiko Ihmels, Mohamed MA Mahmoud, and Brian O Patrick. Optical differentiation between quadruplex DNA and duplex DNA with a [2.2. 2] heptamethinecyanine dye. *Journal of Physical Organic Chemistry*, 30(9):e3736, 2017.
- [58] Billie Jo Harvey, Claudia Perez, and Marcia Levitus. DNA sequence-dependent enhancement of Cy3 fluorescence. *Photochemical & Photobiological Sciences*, 8(8):1105–1110, 2009.
- [59] Billie J Harvey and Marcia Levitus. Nucleobase-specific enhancement of Cy3 fluorescence. *Journal of Fluorescence*, 19(3):443, 2009.
- [60] <https://en.wikipedia.org/wiki/Cyanine>.
- [61] Yijin Liu and David MJ Lilley. Crystal structures of cyanine fluorophores stacked onto the end of double-stranded rna. *Biophysical journal*, 113(11):2336–2343, 2017.
- [62] Linas Urnavicius, Scott A McPhee, David MJ Lilley, and David G Norman. The structure of sulfoindocarbocyanine 3 terminally attached to dsDNA via a long, flexible tether. *Biophys. J.*, 102(3):561–568, 2012.
- [63] Xun Li, Yandong Yin, Xinxing Yang, Zeyong Zhi, and Xin Sheng Zhao. Temperature dependence of interaction between double stranded dna and cy3 or cy5. *Chemical Physics Letters*, 513(4-6):271–275, 2011.
- [64] Ke Jia, Yan Wan, Andong Xia, Shayu Li, Fangbin Gong, and Guoqiang Yang. Characterization of photoinduced isomerization and intersystem crossing of the cyanine dye cy3. *The Journal of Physical Chemistry A*, 111(9):1593–1597, 2007.

- [65] Bernhard M Fuchs, Kazuaki Syutsubo, Wolfgang Ludwig, and Rudolf Amann. In situ accessibility of escherichia coli 23s rna to fluorescently labeled oligonucleotide probes. *Appl. Environ. Microbiol.*, 67(2):961–968, 2001.
- [66] Matthew E Sanborn, Brian K Connolly, Kaushik Gurunathan, and Marcia Levitus. Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA. *J. Phys. Chem. B*, 111(37):11064–11074, 2007.
- [67] V Morozov, A Badasyan, A Grigoryan, M Sahakyan, and Eu Mamasakhlisov. Stacking decreases the cooperativity of melting of homopolymeric dna. *Modern Physics Letters B*, 19(01n02):79–83, 2005.
- [68] John SantaLucia, Jr., H T Allawi, and P A Seneviratne. Improved nearest-neighbour parameters for predicting DNA duplex stability. *Biochem.*, 35:3555–3562, 1996.
- [69] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62(23):2755–2757, 1989.
- [70] T. Dauxois, M. Peyrard, and A. R. Bishop. Entropy-driven DNA denaturation. *Phys. Rev. E*, 47(1):R44–R47, 1993.
- [71] Philip M Morse. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical Review*, 34(1):57, 1929.
- [72] M. Peyrard. Nonlinear dynamics and statistical physics of DNA. *Nonlinearity*, 17:R1, 2004.
- [73] Michel Peyrard, Santiago Cuesta-López, and Guillaume James. Modelling DNA at the mesoscale: a challenge for nonlinear science? *Nonlinearity*, 21(6):T91, 2008.
- [74] Gerald Weber. TfReg: Calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models. *Bioinformatics*, 29:1345–1347, 2013. doi:[10.1093/bioinformatics/btt133](https://doi.org/10.1093/bioinformatics/btt133). URL <http://bioinformatics.oxfordjournals.org/content/29/10/1345>.
- [75] Yong-Li Zhang, Wei-Mou Zheng, Ji-Xing Liu, and Y. Z. Chen. Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E*, 56(6):7100–7115, 1997.
- [76] Gerald Weber, Niall Haslam, Nava Whiteford, Adam Prügel-Bennett, Jonathan W. Essex, and Cameron Neylon. Thermal equivalence of DNA duplexes without melting temperature calculation. *Nat. Phys.*, 2:55–59, 2006. doi:[10.1038/nphys189](https://doi.org/10.1038/nphys189).
- [77] Pâmella Miranda, Luciana M. Oliveira, and Gerald Weber. Mesoscopic modelling of Cy3 and Cy5 dyes attached to DNA duplexes. *Biophys. Chem.*, 230C:62–67, 2017. ISSN 0301-4622. doi:[10.1016/j.bpc.2017.08.007](https://doi.org/10.1016/j.bpc.2017.08.007). URL <http://www.sciencedirect.com/science/article/pii/S0301462217302831>.
- [78] Izabela Ferreira, Tauanne D. Amarante, and Gerald Weber. DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration. *J. Chem. Phys.*, 143:175101, 2015. doi:[10.1063/1.4934783](https://doi.org/10.1063/1.4934783).
- [79] Erik de Oliveira Martins and Gerald Weber. An asymmetric mesoscopic model for single bulges in RNA. *J. Chem. Phys.*, 147:155102, 2017. doi:[10.1063/1.5006948](https://doi.org/10.1063/1.5006948).

- [80] Izabela Ferreira, Elizabeth A. Jolley, Brent M. Znosko, and Gerald Weber. Replacing salt correction factors with optimized RNA nearest-neighbour enthalpy and entropy parameters. *Chem. Phys.*, 521:69–76, may 2019. doi:[10.1016/j.chemphys.2019.01.016](https://doi.org/10.1016/j.chemphys.2019.01.016). URL <https://doi.org/10.1016%2Fj.chemphys.2019.01.016>.
- [81] G. Weber. Finite enthalpy model parameters from DNA melting temperatures. *Europhys. Lett.*, 96:68001, 2011. doi:[10.1209/0295-5075/96/68001](https://doi.org/10.1209/0295-5075/96/68001). URL <http://iopscience.iop.org/0295-5075/96/6/68001>.
- [82] Tauanne D. Amarante and Gerald Weber. Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model. *J. Chem. Inf. Model.*, 56(1):101–109, 2016. doi:[10.1021/acs.jcim.5b00571](https://doi.org/10.1021/acs.jcim.5b00571). URL <http://dx.doi.org/10.1021/acs.jcim.5b00571>.
- [83] Rodolfo Vieira Maximiano and Gerald Weber. Deoxyinosine mismatch parameters calculated with a mesoscopic model result in uniform hydrogen bonding and strongly variable stacking interactions. *Chem. Phys. Lett.*, 631–632:87–91, 2015. doi:[10.1016/j.cplett.2015.04.045](https://doi.org/10.1016/j.cplett.2015.04.045).
- [84] William H Press, Saul A Teukolsky, William T Vetterling, Brian P Flannery, et al. *Numerical recipes in C*, volume 2. Cambridge university press Cambridge, 1992.
- [85] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [86] Michael Quinten. *A Practical Guide to Optical Metrology for Thin Films*. John Wiley & Sons, 2012.
- [87] Alfredo Alves de Farias, Cibele Comini Cesar, and Jose Francisco Soares. *Introdução à estatística*. LTC, 2003.
- [88] Marcello Pagano and Kimberlee Gauvreau. *Princípios de bioestatística*. Cengage Learning, 2004.
- [89] <http://www.cqeacademy.com/cqe-body-of-knowledge/continuous-improvement/quality-control-tools/the-scatter-plot-linear-regression/>.
- [90] Tianbing Xia, John SantaLucia, Jr., Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, and Douglas H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochem.*, 37:14719–14735, 1998.
- [91] Gerald Weber, Jonathan W. Essex, and Cameron Neylon. Probing the microscopic flexibility of DNA from melting temperatures. *Nat. Phys.*, 5:769–773, 2009. doi:[10.1038/nphys1371](https://doi.org/10.1038/nphys1371).
- [92] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.
- [93] Gerald Weber. Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures. *Nucleic Acids Res.*, 41:e30, 2013. doi:[10.1093/nar/gks964](https://doi.org/10.1093/nar/gks964). URL <http://nar.oxfordjournals.org/content/41/1/e30>.

- [94] Marie Zgarbová, Michal Otyepka, Jiri Sponer, Filip Lankas, and Petr Jurečka. Base pair fraying in molecular dynamics simulations of DNA and RNA. *Journal of Chemical Theory and Computation*, 10(8):3177–3189, 2014. doi:[10.1021/ct500120v](https://doi.org/10.1021/ct500120v).