

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

LUCIANO GABRIEL SILVA

**Modelos mesoscópicos para DNA mediado por metal e na  
presença de solventes que simulam o meio molecular**

Belo Horizonte  
2023

Luciano Gabriel Silva

**Modelos mesoscópicos para DNA mediado por metal e na presença de solventes que simulam o meio molecular**

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciências.

Orientador: Gerald Weber

Belo Horizonte

2023

Dados Internacionais de Catalogação na Publicação (CIP)

S586m Silva, Luciano Gabriel.  
Modelos mesoscópicos para DNA mediado por metal e na presença de solventes que simulam o meio molecular / Luciano Gabriel Silva. – 2023.  
166 f. : il.

Orientador: Gerald Weber.  
Tese (doutorado) – Universidade Federal de Minas Gerais,  
Departamento de Física.  
Bibliografia: f. 125-140.

1. Biofísica. 2. DNA. I. Título. II. Weber, Gerald. III. Universidade Federal de Minas Gerais, Departamento de Física.

CDU – 577.3 (043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

### ATA DE DEFESA DE TESE

**ATA DA SESSÃO DE ARGUIÇÃO DA 417ª TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA, DEFENDIDA POR LUCIANO GABRIEL SILVA** orientado pelo professor Gerald Weber, para obtenção do grau de **DOUTOR EM CIÊNCIAS, área de concentração Física**. Às 9 horas de sete de julho de dois mil e vinte e três reuniu-se, por videoconferência, a Comissão Examinadora, composta pelos professores **Gerald Weber** (Orientador - Departamento de Física/UFMG), **Ubirajara Agero Batista** (Departamento de Física/UFMG), **Mariana Torquato Quezado de Magalhães** (Instituto de Ciências Biológicas/UFMG), **Elsó Drigo Filho** (Instituto de Biociências, Letras e Ciências Exatas/Unesp) e **Márcio Santos Rocha** (Departamento de Física/UFV), para dar cumprimento ao Artigo 37 do Regimento Geral da UFMG, submetendo o Mestre **LUCIANO GABRIEL SILVA** à arguição de seu trabalho de Tese de Doutorado, que recebeu o título de "**Modelos mesoscópicos para DNA mediado por metal e na presença de solventes que simulam o meio molecular**". O candidato fez uma exposição oral de seu trabalho durante aproximadamente 50 minutos. Após esta, os membros da comissão prosseguiram com a sua arguição, e apresentaram seus pareceres individuais sobre o trabalho, concluindo pela aprovação do candidato.

Belo Horizonte, 07 de julho de 2023.

Prof. Gerald Weber  
Orientador do estudante  
Departamento de Física/UFMG

Prof. Ubirajara Agero Batista  
Departamento de Física/UFMG

Prof. Elso Drigo Filho  
Instituto de Biociências, Letras e Ciências Exatas/Unesp

Prof. Márcio Santos Rocha  
Departamento de Física/UFV

Profa. Mariana Torquato Quezado de Magalhães  
Instituto de Ciências Biológicas/UFMG

**Candidato:** Luciano Gabriel Silva



Documento assinado eletronicamente por **Luciano Gabriel Silva, Usuário Externo**, em 07/07/2023, às 14:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gerald Weber, Professor do Magistério Superior**, em 07/07/2023, às 14:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Márcio Santos Rocha, Usuário Externo**, em 07/07/2023, às 15:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ubirajara Agero Batista, Servidor(a)**, em 07/07/2023, às 15:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes, Professora do Magistério Superior**, em 12/07/2023, às 17:19, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Elsó Drigo Filho, Usuário Externo**, em 17/07/2023, às 14:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2452529** e o código CRC **32DE6106**.

## Agradecimentos

Desejo expressar minha gratidão aos meus pais, Gilmar e Eliana, pelo incentivo e apoio inestimáveis ao longo de minha jornada acadêmica, bem como pelos valiosos ensinamentos de vida, carinho e amor. Em particular, gostaria de agradecer a meu pai, cuja assistência na resolução de problemas matemáticos durante meus anos no ensino fundamental despertou em mim o amor pelas ciências exatas.

Gostaria também de expressar minha gratidão a meu irmão, Rafael, pelas conversas significativas que tivemos. Suas palavras e reflexões foram essenciais para minha tomada de decisões em diversas situações que enfrentei.

Embora meus tios, Roberto e Sirlene, atualmente não residam mais em Belo Horizonte (BH), considero-os como meus "pais" nesta cidade e desejo agradecer-lhes profundamente. Eles foram as primeiras pessoas de BH com quem tive contato e sempre me ofereceram auxílio quando necessário. Sentar-me e conversar com eles era um prazer indescritível.

Gostaria de estender meus agradecimentos a todas as pessoas com quem convivi em Belo Horizonte, as quais me ajudaram tanto nos momentos de lazer quanto nos momentos de estudo. Além disso, sou grato a outros amigos que fiz ao longo de toda minha trajetória acadêmica. Em especial, gostaria de mencionar meus amigos Alex Azevedo, Mateus Leal, Luis Felipe Araújo, Daniel Batista, Fabiano Ishiy (Zaparoli) e Isaac Martins. Nossa convivência em BH foi repleta de momentos inesquecíveis que permanecerão para sempre em minha memória.

Expresso minha gratidão aos professores Maurício Veloso e Klaus Krambrock, da Universidade Federal de Minas Gerais (UFMG), que foram meus primeiros mentores na pesquisa. Durante esse período, tive a oportunidade de conhecer diversos amigos e colegas que me auxiliaram em meu desenvolvimento pessoal e profissional. Em especial, desejo agradecer a Elias Vieira, Frederico Dias Brandão, Bruno Cordeiro, Henrique Fernandes e Daniele Cristina Ferreira. Essa foi uma época extremamente gratificante da qual sempre me recordarei.

Quero manifestar minha gratidão ao meu orientador e amigo, Gerald Weber, por ter reacendido minha paixão pela ciência. Agradeço por todo o aprendizado e apoio ao longo do trabalho, assim como pelas críticas construtivas que contribuíram para o meu desenvolvimento pessoal e profissional. Além disso, sou grato pela assistência e apoio que recebi durante momentos difíceis ao longo de meu doutorado.

Gostaria também de expressar minha gratidão aos meus amigos e colegas de laboratório. Em particular, a Mateus Leal, Maria Izabel, Pâmella Miranda e Izabela Ferreira, cujas contribuições me auxiliaram na compreensão dos conceitos biológicos, na compreensão das equações do modelo e na elaboração dos códigos necessários para a construção de alguns dos gráficos da seção de resultados.

Gostaria de agradecer a Luis Henrique, Camila Cacique e Daniel Batista (novamente), que moraram comigo na época e me ajudaram durante um problema psicológico que enfrentei durante o doutorado.

Também expresso meu agradecimento a todos os demais professores da UFMG, pelos ensinamentos valiosos tanto no âmbito acadêmico quanto no pessoal.

Por fim, desejo expressar minha gratidão às agências de fomento CNPq, CAPES e FAPEMIG pela assistência financeira ao longo desses anos de estudo. Em especial, à CAPES pela bolsa de doutorado.

## Resumo

Ácidos nucleicos, em especial DNA e RNA, têm importância biológica e tecnológica inegável. Aqui, desenvolvemos dois projetos relacionados à estabilidade de ácidos nucleicos. No primeiro projeto abordamos o efeito de íons metálicos,  $\text{Hg}^{2+}$  e  $\text{Ag}^+$ , que se ligam entre as bases do DNA (M-DNA). Esses íons têm potencial tecnológico para serem usados como biosensores. No segundo projeto, abordamos a questão do efeito de polietilenoglicol (PEG) na desnaturação do DNA. O PEG simula o efeito do confinamento molecular, representando macro- e micro-moléculas existentes na célula, o que é um importante problema em biologia molecular. Para o desenvolvimento destes projetos, usamos dados de desnaturação UV presentes na literatura e utilizamos modelos mesoscópicos, principalmente o modelo Peyrard-Bishop (PB), para calcular as interações presentes no sistema.

M-DNA é uma molécula de DNA em que a ligação de hidrogênio entre as bases é substituída por um íon metálico. No DNA natural, apenas os íons  $\text{Ag}^+$  e  $\text{Hg}^{2+}$  apresentam efeito estabilizador. No entanto, não está claramente estabelecido se estes íons o fazem por uma ligação covalente base-metal-base, ligação direta, ou se há outro mecanismo em questão, como alterações no efeito de empilhamento (*stacking*), ligações interplanares. Sabe-se que  $\text{Ag}^+$  estabiliza citosina-citosina (CC), enquanto  $\text{Hg}^{2+}$  estabiliza timina-timina (TT). Ou seja, os íons estabilizam pares de base do tipo *mismatch*, que normalmente são bastante instáveis. Em nossos resultados, vimos que  $\text{Hg}^{2+}$  estabiliza por meio de ligação base-metal-base próxima ao valor de um par CG, enquanto para  $\text{Ag}^+$  temos o mesmo tipo de estabilização, porém com intensidade próxima à de um par AT. Valores pequenos de empilhamento CC-CC para  $\text{Ag}^+$  e TT-TT para  $\text{Hg}^{2+}$  descartam interações interplanares. Com isto estabelecemos que modelo mesoscópico, a partir dos dados experimentais disponíveis, pode determinar conclusivamente a origem do efeito estabilizante do íon metálico.

Para compreender os efeitos intramoleculares de DNA e RNA, os experimentos *in vitro* são comumente realizados em solução de água e sal, livre de outros aditivos. Isto é muito diferente do conteúdo celular, que contém cerca de 20-40% de macro- e micro-moléculas. Portanto, não se sabe até que ponto os resultados destes experimentos correspondem ao que ocorre com essas moléculas no ambiente celular. Para estudar estes efeitos, costuma-se adicionar polietilenoglicol (PEG), que está disponível numa variedade de massas moleculares e, por isto, oferece uma maneira simples e controlada para simular experimentalmente o efeito da presença de micro- e macro-moléculas. PEGs com massas moleculares menores do que 1000 (PEG200, por exemplo) se comportam como micro-moléculas e desestabilizam o DNA, sendo que com massa maior (PEG2000, por exemplo) agem como macro-moléculas e estabilizam o DNA. Muitos autores modificaram o modelo PB adicionando um termo no potencial de Morse para descrever a interação dos nucleotídeos com solvente, no entanto, estes modelos nunca foram validados com dados experimentais. Por ser um

campo recente, apenas agora dispomos de dados experimentais suficientes para estudar se os modelos de fato descrevem o efeito que as moléculas de água ( $H_2O$ ) tem sobre os nucleotídeos. Neste estudo, comparamos situações de solução aquosa na presença ou ausência de PEG200 em DNA (baixa concentração de sal,  $Na^+=100$  mM) e RNA (baixa concentração de sal,  $Na^+=100$  mM, e alta concentração de sal,  $Na^+=1000$  mM). Nossos resultados mostram que o RNA é mais hidratado do que o DNA. Na presença de PEG200 a diferença de energia entre os pares AT e CG no DNA, e AU e CG no RNA, aumentam, porém no RNA a diferença de energia é bem menor. Grandes diferenças de energia entre os pares de base podem causar distorções na dupla hélice. Neste caso, o DNA na presença de PEG200 pode estar assumindo um outro tipo de conformação, hélice do tipo B, predominante em soluções com água, para uma hélice de tipo A, predominante em soluções com menos água disponível. Valores de energia AT na presença e ausência de PEG200 são praticamente iguais. Este resultado confirma que hélices de DNA homogêneas AT só existem no tipo B independente da mudança de hidratação do par de base. Dentre os modelos PB com barreira analisados, apenas o modelo Hamiltoniano Morse Solvente ( $HMS^-$ ) conseguiu prever as temperaturas de desnaturação de maneira equivalente, pois os parâmetros das qualidades das predições são pequenos. Na ausência de PEG200 e baixas concentrações de sal as energias de barreira são praticamente iguais. Ao contrário do que acontece em altas concentrações de sal, onde a barreira de CG é um pouco maior do que AU. Além de ser mais hidratado, o RNA concentra mais íons do que o DNA, por este motivo observamos uma redução da barreira de energia na presença de PEG200. O modelo  $HMS^-$  explica muito bem a interação nucleotídeo-solvente-nucleotídeo. Logo, encontramos um modelo PB modificado que explica uma parte da dinâmica dos nucleotídeos inexistente no modelo PB original.

**Palavras-chave:** Desnaturação UV, Nucleotídeos, DNA mediado por metal (M-DNA),  $Ag^+$ ,  $Hg^{2+}$ , ambiente macromolecular, PEG, hidratação, Modelo Peyrard-Bishop (PB), Modelo PB com barreira

## Abstract

Nucleic acids, in special DNA and RNA, have undeniable technological and biological importance. Here, we develop two projects linked to the stability of nucleic acids. In the first, we discuss the effects of metallic ions that bind between the base pairs of DNA (M-DNA), and which has technological potential to be used as biosensors. In the second, we address the effect of polyethylene glycol, that simulates the crowding effect in the denaturation of DNA, an important problem in molecular biology. For the development of these projects we use UV published melting. In addition, we use mesoscopic models, mainly the Peyrard-Bishop (PB), to calculate the interactions present in the system.

M-DNA is a DNA molecule where the hydrogen bond between the bases is changed due to the presence of a metallic ion. In natural DNA, only  $\text{Ag}^+$  and  $\text{Hg}^{2+}$  have a stabilizing effect. However, there is no clear understanding if these ions stabilize the DNA via a direct bond, covalent bond base-metal-base, or via another mechanism such as interplanar interactions, stacking between the nearest-neighbouring bases. We know that  $\text{Ag}^+$  stabilizes the cytosine-cytosine (CC) mismatch, while  $\text{Hg}^{2+}$  stabilize the thymine-thymine (TT) mismatch. In other words, the cations stabilize mismatches, who are otherwise quite unstable. Our results show that  $\text{Hg}^{2+}$  stabilizes the DNA via the pair base-metal-base which has a strength similar to a CG pair, while for  $\text{Ag}^+$  the stabilization process is the same, however with a strength somewhat weaker than an AT base pair. The small stacking interactions of CC-CC for  $\text{Ag}^+$  and TT-TT for  $\text{Hg}^{2+}$  does not support interplanar interactions which were hypothesized by some authors. Therefore, mesoscopic models can conclusively explain the origin of the metal ion stabilizing effect from experimental data available.

To understand the intramolecular effects in DNA and RNA, *in vitro* experiments are commonly performed in saline solutions, free of other compounds. This is very different from the cell environment that has a myriad of molecules, the cell volume is composed of 20-40% of micro- and macro-molecules. Therefore, it is not know how far the results of the experiments can explain what really happens on the molecular environment. To study these effects it is customary to add polyethylene glycol (PEG), which is available in a variety of molecular weights and therefore offers a simple and controlled way to experimentally simulate the effect of micro- and macro-molecules on the nucleotides. PEGs with molecular weights lower than 1000 (PEG200, for instance) behave like micromolecules and destabilize the DNA, while PEGs with molecular weights greater act as macromolecules and stabilize the DNA. Several authors proposed modifications to the PB model by adding a term in the Morse potential to describe the interaction of nucleotides with the solvent, water in particular, but these models have never been subject to validation with experimental data. Only recently have enough experimental data become available to study whether the models actually describe the effect that  $\text{H}_2\text{O}$  molecules have on nucleotides, giving us the opportunity to apply

the PB model in this type of system. In this thesis, we compare DNA (low salt concentration,  $[\text{Na}^+] = 100 \text{ mM}$ ) and RNA (low,  $[\text{Na}^+] = 100 \text{ mM}$ , and high,  $[\text{Na}^+] = 1000 \text{ mM}$ , salt concentration) aqueous solution in two situations: with PEG200 and without PEG200. Our results confirm that RNA is more hydrated than DNA. The energy difference between AT and CG, DNA, and AU e CG, RNA, increases when we add PEG200, but in RNA the energy difference is much smaller. Large energy differences between the base pairs cause distortions in the conformation of the double helix. In this case, DNA in the presence of PEG200 may be assuming another type of conformation. DNA B-type helix, characteristic in water solutions, changes to an A-type helix, characteristic in solutions with less water available. AT energy values with and without PEG200 are practically the same. This result confirms that homogeneous AT DNA helices assume only B-type conformation regardless of the hydration change. Among the barrier-modified PB models analysed, only the  $\text{HMS}^-$  was found to adequately predict melting temperatures. In the absence of PEG200 and low salt concentrations the barrier energies were found to be the same. Different to what happens at high salt concentrations, where the CG barrier is a little higher than the AU barrier. RNA concentrates more ions and it is more hydrated than DNA, for this particularly reason we observe a reduction in the barrier energy with PEG200. The  $\text{HMS}^-$  model was found to provide a satisfactory description of the nucleotide-solvent-nucleotide interaction. Therefore, we found a modified PB model that can explain the intermediate state in the duplex separation process, absent of the standard PB model.

**Keywords:** UV melting, Nucleotides, Metal mediated DNA (M-DNA),  $\text{Ag}^+$ ,  $\text{Hg}^{2+}$ , Crowding, PEG, hydration, Peyrard-Bishop Model (PB), PB barrier models

## Lista de Figuras

- 1 Monômero de DNA e RNA. a) Nucleotídeo desenhado em verde composto de três grupos fosfatos + pentose (ribose - RNA ou desoxirribose -DNA) + Base; b) Bases nitrogenadas. Fonte: o próprio autor . . . . . 28
- 2 Estrutura do polinucleotídeo a) Disposição dos monômeros na fita simples com destaque para o esqueleto em cinza sombreado (Fonte: o próprio autor); b) Modelo de dupla hélice proposto por Watson e Crick - DNA tipo B (figura retirada de Zaha et al. [1]); c) ligações de hidrogênio presentes nos nucleotídeos (Fonte: o próprio autor). . . . . 29
- 3 Definição da interação de empilhamento em anéis aromáticos. bolinhas azuis claro são qualquer átomo que realiza ligações químicas compartilhando elétrons e bolinhas verdes são átomos de carbono. a) Tipos de ligações covalentes e distribuição de carga em anéis aromáticos; b) Dois tipos de interações  $\pi - \pi$  presentes em anéis aromáticos: interações sanduíche (*face to face  $\pi$  stacking*) e paralelo deslocado (*offset stacking*). Fonte: o próprio autor. . . . . 30
- 4 Distribuição de cargas nas bases nitrogenadas. a) Potencial eletrostático na superfície das bases nitrogenadas. As escalas de cor estão em kcal/mol: vermelho (maior que 31.31), amarelo (entre 10.17 até 31.31), verde (entre -10.92 até 10.17), azul (mais negativo do que -10.92). Figura retirada de Veljković et al. [2] b) Representação simplificada da distribuição de carga molecular nas bases nitrogenadas derivada do potencial eletrostático (Fonte: o próprio autor); c) Tipos de dímeros presentes no DNA (Fonte: o próprio autor); d) Principais movimentos dos dímeros presentes na hélice de DNA. Cores (figuras b e d) - vermelho: nitrogênio, verde: carbono, azul: oxigênio e marrom: fósforo (Fonte: o próprio autor). . . . . 32
- 5 Os 10 tipos de próximos vizinhos possíveis no DNA representados com a distribuição de carga simplificada de cada base nitrogenada. Todos os átomos de hidrogênio, não desenhados na figura, possuem carga positiva. Cores - vermelho: nitrogênio, verde: carbono e azul: oxigênio (Fonte: o próprio autor). . . . . 32

6	<p>Interações das moléculas de água com moléculas responsáveis por estabilizar os nucleotídeos. a) Associação das moléculas de água formando um arranjo tetraédrico. A esfera vermelha representa o oxigênio, as esferas pretas o hidrogênio e as linhas tracejadas azuis às ligações de hidrogênio (Fonte: o próprio autor); b) Definição de conchas de hidratação mostradas no exemplo com o nucleotídeo adenina. Cores - vermelho: oxigênio, verde: carbono, azul: nitrogênio, branco: hidrogênio e roxo: fósforo. Figura retirada de Rozza et al. [3]; c) Interação dos cátions, esfera em cinza, com às moléculas de água (Fonte: o próprio autor). . . . .</p>	35
7	<p>Três tipos de hélices de DNA observadas em meio fisiológico. DNA tipo A, DNA tipo B e DNA tipo Z. Figura retirada de Zaha et al. [1]. . . . .</p>	37
8	<p>Distribuição de íons e águas no DNA e RNA a) Distribuição das moléculas de água ao redor dos pares de base r(G=C) e r(A-U), RNA, d(G=C) e d(A-T), DNA. Figura retirada de Auffinger and Westhof [4]; b) Distribuição dos íons de Na<sup>+</sup> nas hélices de DNA tipo B e RNA tipo A. Figura retirada de Gebala and Herschlag [5] . . . . .</p>	37
9	<p>Alguns tipos de emparelhamentos realizados pelos <i>mismatches</i>. Na imagem a) mostramos o par GG, na b) o par TT, na c) o par AA, na d) o par CC e na e) o par CC protonado. Figura retirada de Peyret et al. [6] . . . . .</p>	39
10	<p>Esquema de um experimento de UV <i>melting</i> e exemplo de espectro de absorção normalizado. Fonte: o próprio autor. . . . .</p>	40
11	<p>Tipos de sequências envolvidas no experimento de UV <i>melting</i>. a) Sequências auto-complementares (AC) com as fitas representadas em verde e b) Sequências não auto-complementares (NAC) com uma espécie de fita em azul e a outra em vermelho. Fonte: o próprio autor. . . . .</p>	41
12	<p>Curvas experimentais do experimento de desnaturação de DNA do duplexo 5'-CAAAAAAAG-3'. a) Curvas de desnaturação em função da temperatura em diferentes concentrações totais <math>C_t</math>. b) Gráfico de Van't Hoff onde cada ponto do gráfico representa a temperatura de desnaturação do gráfico a) e sua respectiva concentração <math>C_t</math>. Figura retirada de Howard [7] . . . . .</p>	43
13	<p>Modelo Peyrard-Bishop, S1 e S2 são as duas fitas representando o <i>backbone</i> do DNA com <math>N</math> pares de base, as linhas tracejadas verticais representam as ligações de hidrogênio. Fonte: o próprio autor. . . . .</p>	44
14	<p>Representação 2D do empilhamento e ligações de hidrogênio em DNA. As funções <math>w</math> mostradas representam o potencial de empilhamento e <math>V</math> as ligações de hidrogênio. Fonte: o próprio autor. . . . .</p>	48

15	a) comportamento gaussiano entre $Z_{\omega}(\Lambda)$ e $\omega$ para 4 sequências de 10 pares de base contendo 40% à 60% de pares de base C-G na temperatura de 370 K; b) Comportamento de $\omega_{max}$ em função da temperatura para outras 4 sequências com 10 pares de base; c) Temperaturas de <i>melting</i> experimentais ( $T_m$ ) em função dos valores de $\tau = (\omega_{max})^{1/2}$ para diversas sequências. Figuras retiradas de Weber et al. [8]. . . . .	51
16	a) Processo de minimização. O quadrado sombreado em amarelo representa o espaço onde o simplex é definido. O triângulo (1) representa uma das possíveis escolhas de parâmetros iniciais $P$ para iniciar a minimização da função $\chi^2$ . b) Alguns exemplos de figuras resultantes do processo de minimização. Fonte: o próprio autor. . . . .	54
17	Esquema de funcionamento do modelo NN. a) fitas auto-complementares (AC) e b) fitas não auto-complementares (NAC). Fonte: o próprio autor. . . . .	56
18	Representação teórica de uma fita de DNA com metais substituindo às ligações de hidrogênio. As esferas em verde representam os íons metálicos que substituem as ligações de hidrogênio por ligações covalentes mediadas por metal (figura retirada de Aich et al. [9]). . . . .	58
19	Configurações possíveis de $Ag^+$ e $Hg^{2+}$ que se ligam covalentemente as bases nitrogenadas do DNA. As duas figuras acima são do caso WC e as duas figuras abaixo são do caso WC reverso. R significa desoxirribose. Fonte: o próprio autor. . . . .	59
20	Configurações do par C-Hg-T. O duplexo em cinza do lado esquerdo é de conformação tipo B e a ligação com o Hg(II) é T(N3)-Hg(II)-(N3)C; O duplexo em azul do lado direito é de conformação tipo A e a ligação com o Hg(II) é T(N3)-Hg(II)-(N4)C. Figura extraída de Schmidt et al. [10] . . . . .	61
21	Estruturas de $Hg^{2+}$ encontradas por Liu et al. [11]. a) Sequência principal e complementar, os pontos vermelhos mostram a localização do íon metálico. b) A conformação detalhada do par T-Hg <sup>2+</sup> -T com os íons metálicos em vermelho. c) Visão esquemática mostrando em detalhes a interação dos <i>mismatches</i> T com o íon $Hg^{2+}$ . Todas as figuras foram retiradas de Liu et al. [11]. . . . .	61
22	Estruturas de M-DNA resolvidas experimentalmente. a) nanofio de Ag(I) com um átomo de Ag(I) entre as bases nitrogenadas. Figura extraída de Kondo et al. [12]. b) nanofio de Hg(II) as esferas em cinza são os átomos de Hg(II) e as esferas vermelhas são oxigênios da molécula de H <sub>2</sub> O. Figura extraída Ono et al. [13]. c) Duplexo de DNA com dois átomos de Hg(II) em vermelho entre os <i>mismatches</i> TT. Figura extraída de Yamaguchi et al. [14]. . . . .	62

23	Profundidades médias do potencial de Morse $D$ calculadas para C-Ag <sup>+</sup> -C , T-Hg <sup>2+</sup> -T em azul e seus respectivos controles em vermelho. As barras de erro são os desvios padrões da última rodada de minimizações. Os valores de semente para os <i>mismatches</i> TT <sup><i>i</i></sup> e CC <sup><i>i</i></sup> são mostrados em cinza. As setas tracejadas indicam as mudanças devido a adição de cátions. A área sombreada mostra o intervalo mínimo e máximo para potenciais de Morse para todos os tipos de <i>mismatches</i> TT e CC da referência Oliveira et al. [15] . . . . .	67
24	Interações de empilhamento médias $k$ calculada para (a) T-Hg <sup>2+</sup> -T e (b) C-Ag <sup>+</sup> -C como pontos azuis. Os pontos vermelhos são os controles (a) TT-CTRL e (b) CC-CTRL. As barras de erro são os desvios padrões da última rodada de minimizações. A área sombreada mostra o intervalo mínimo e máximo para os empilhamentos dos <i>mismatches</i> TT e CC da referência Oliveira et al. [15] . . . . .	68
25	Deslocamento médio para as sequências das referências [12, 16–18]. Os pontos vermelhos são das CC-CTRL e os azuis são das C-Ag <sup>+</sup> -C. . . . .	69
26	Deslocamento médio para as sequências das referências [11, 14, 19–22]. Na parte (h) AT-TT da parte (g) foi substituído por TA-TT. Os pontos vermelhos são das TT-CTRL e os azuis são das T-Hg <sup>2+</sup> -T. . . . .	70
27	Estruturas possíveis de DNA mediado por metal Ag <sup>+</sup> , Hg <sup>2+</sup> e híbridos. Destacamos a região sombreada dos gráficos, onde são a dos <i>mismatches</i> mediados por um átomo metálico na estrutura. (a) Estruturas de DNA com o par AT nas pontas. O gráfico à direita mostra a região entre 0 e 2 Å ampliada. (b) Estruturas de DNA com o par CG nas pontas. O gráfico à direita mostra a região entre 0 e 2 Å ampliada. As linhas em preto das figuras do lado direito são de pares canônicos contendo 8 pares de base AT (A <sub>8</sub> ) e 8 pares CG (C <sub>8</sub> ) para ser a nossa referência. . . . .	71
28	Diferenças entre experimentos <i>in vivo</i> e <i>in vitro</i> . a) célula eucarionte, as regiões <i>X</i> e <i>Y</i> são discutidas no texto, e b) experimentos <i>in vitro</i> . . . . .	75
29	a) Estrutura química de alguns solutos utilizados para simular o ambiente molecular; b) comparação dos tamanhos dos PEGs de diferentes pesos moleculares, etilenoglicol e um pequeno duplexo de DNA. Figura extraída de Nakano et al. [23] . . . . .	76
30	Efeito de PEG na estabilidade do nucleotídeo. Mostramos o efeito de pequenos solutos e grandes solutos, esferas laranja, nos estados de fita dupla e fita simples na temperatura de desnaturação. a) PEGs com massas moleculares menores do que 1000 e b) PEGs com massas moleculares maiores do que 1000 . . . . .	77
31	Comparação entre potenciais de Morse (a) sem e (b) com barreira de potencial. Os elementos I, II e III são estados do sistema. I - fita dupla, II - intermediário e III - fita simples. . . . .	79

32	a)	Comparação entre o potencial de Morse PB (curva em preto) e o potencial de Morse-solvente - HMS (curvas em vermelho e azul) b) estiramento médio $\langle y \rangle$ em função da temperatura para o Hamiltoniano Morse-solvente — equação (11) substituindo o potencial de Morse $V(y_n)$ pelo potencial HMS da equação (34). Figuras retiradas de Weber [24]. . . . .	80
33		Comparação entre o potencial de Morse PB (linha tracejada em preto) e o potencial de Morse com barreira entrópica — PCLA (linha em preto). Os parâmetros dos potenciais foram escolhidos para reproduzir a mesma temperatura de desnaturação de uma sequência poly(A). Figura retirada de Peyrard et al. [25]. . . . .	81
34	a)	Comparação entre o potencial de Morse PB (linha tracejada em preto) e o potencial de Morse com potencial 'hump' — PCLJ (linha em preto). Os parâmetros dos potenciais foram escolhidos para reproduzir a mesma temperatura de desnaturação. b) Deslocamento médio dos pares de base no DNA descritos pelo modelo PB, figura de cima, e pelo modelo PB modificado com o potencial de 'hump', figura de baixo. Figuras retiradas de Peyrard et al. [26]. . . . .	83
35		Comparação entre o potencial de Morse PB (linha tracejada) e o potencial de Morse com potencial gaussiano — TRMF (linha sólida). Os parâmetros dos potenciais foram escolhidos para reproduzir a mesma temperatura de desnaturação. Figura retirada de Tapia-Rojo et al. [27]. . . . .	85
36		Profundidades médias de Morse $D$ calculadas para DNA e RNA em H <sub>2</sub> O como pontos vermelhos e em PEG200 como pontos azuis. Para comparação mostramos os parâmetros $D_{AT}$ , ou $D_{AU}$ , (linha inferior) e $D_{CG}$ (linha superior) de outras publicações do nosso grupo em H <sub>2</sub> O sob as mesmas concentrações de [Na <sup>+</sup> ] [28–30] como linhas tracejadas em cinza. (a) DNA-W/40 e DNA-P200/40 (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40. . . . .	91
37		Interações de empilhamento médias $k$ calculadas para DNA e RNA em H <sub>2</sub> O como pontos vermelhos e em PEG200 como pontos azuis. A região sombreada representa o intervalo de parâmetros de empilhamento de outras publicações do nosso grupo em H <sub>2</sub> O sob as mesmas concentrações de [Na <sup>+</sup> ] [28–30]. (a) DNA-W/40 e DNA-P200/40 (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40 . . . .	92
38		Comparação entre os potenciais do modelo PCLA, equação (35), e PCLA modificados, equação (39). PCLA-3 ( $p = 3$ ), PCLA-2 ( $p = 2$ ), PCLA-1 ( $p = 1$ ) e PCLA-0 ( $p = 0$ ). Os parâmetros de todas as curvas são: $D=83.51$ meV; $\lambda=0.11$ Å; $b=8.35$ meV/Å <sup><math>p</math></sup> , $b=8.35$ meV/Å <sup>3</sup> (PCLA); $c=4.4 \times 10^{-3}$ Å; $\alpha=9.39$ Å <sup>-1</sup> . . . . .	109

- 39 Modelo PCLA modificado. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $5^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a) PCLA-3, (b) PCLA-2, (c) PCLA-1, (d) PCLA-0. . 110
- 40 Modelo PCLA modificado. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $3^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a) PCLA-3, (b) PCLA-2, (c) PCLA-1, (d) PCLA-0. . 110
- 41 Modelo HMS-. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AU e CG em  $y_e = 0.3$  nm. Colorimos todos os pontos acima do valor máximo mostrado nas escalas de cores de  $\langle \chi^2 \rangle$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). os valores NA no mapa de calor são de cálculos que geraram valores infinitos. . . . . 115
- 42 Energias de barreira  $f_s D$  calculadas para DNA e RNA em H<sub>2</sub>O como pontos vermelhos e em PEG200 como pontos azuis. A seta tracejada indica a mudança na altura da barreira de potencial causada pela adição de PEG200 na solução. (a) DNA-W/40 e DNA-P200/40, (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40. 117
- 43 Interações de empilhamento médias  $k$  calculadas para DNA e RNA em H<sub>2</sub>O como pontos vermelhos e em PEG200 como pontos azuis. A região sombreada representa o intervalo de parâmetros de empilhamento dos resultados do modelo PB apresentados no capítulo 5.5.2. (a) DNA-W/40 e DNA-P200/40, (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40 . . . . . 118
- 44 Comparação entre os perfis de abertura ajustados pelo modelo PB, linha tracejada menos espaçada, e ajustados pelo modelo HMS<sup>-</sup>, linha tracejada. A região sombreada representa os pontos onde temos um par AT ou AU. No eixo  $x$  inferior apresentamos a sequência de DNA e no superior a de RNA usadas para calcular os perfis de abertura. (a)-(m) Perfis de abertura das 13 sequências com pares de base iguais na base de dados DNA-P200/40 e RNA-P200/40. (n) Perfis de abertura de uma sequência com 8 pares de base CG e (o) Perfis de abertura de uma sequência com 8 pares de base AT, ou AU para RNA. . . . . 120

- 45 Modelo HMS<sup>-</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $14^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm, equação (34). . . . . 148
- 46 Modelo HMS<sup>-</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $30^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm. . . . . 148
- 47 Modelo HMS<sup>+</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $40^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm. . . . . 150
- 48 Modelo HMS<sup>+</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $30^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm. . . . . 150
- 49 Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $20^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $d = 0.05$ , (b)  $d = 0.3$ , (c)  $d = 0.5$ , (d)  $d = 1$ . . . . . 152
- 50 Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $14^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $d = 0.05$ , (b)  $d = 0.3$ , (c)  $d = 0.5$ , (d)  $d = 1$ . . . . . 152

- 51 Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $G$  para AT e CG em diferentes distâncias de equilíbrio  $y_o$ . Colorimos todos os pontos maiores que  $6^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_o = 0.05$  nm, (b)  $y_o = 0.3$  nm, (c)  $y_o = 0.5$  nm, (d)  $y_o = 1$  nm. . . . . 154
- 52 Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $G$  para AT e CG em diferentes distâncias de equilíbrio  $y_o$ . Colorimos todos os pontos maiores que  $5^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_o = 0.05$  nm, (b)  $y_o = 0.3$  nm, (c)  $y_o = 0.5$  nm, (d)  $y_o = 1$  nm. . . . . 155
- 53 Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $E$  para AT e CG em diferentes energias  $D_h$ . Destacamos a região de intersecção entre os retângulos nas figuras. Esta região representa o intervalo de valores de  $E_{CG}$  e  $E_{AT}$  que reproduzem valores de profundidades de poço  $A$  similares ao modelo PB. Colorimos todos os pontos maiores que  $40^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$ . (a)  $D_h = 0$  meV, (b)  $D_h = 1$  meV, (c)  $D_h = 10$  meV. O gráfico abaixo da figura (b) é da região entre a intersecção dos retângulos ampliada. . . . . 157
- 54 Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $E$  para AT e CG em diferentes energias  $D_h$ . Destacamos a região de intersecção entre os retângulos nas figuras. Esta região representa o intervalo de valores de  $E_{CG}$  e  $E_{AT}$  que reproduzem valores de profundidades de poço  $A$  similares ao modelo PB. Colorimos todos os pontos maiores que  $30^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$ . (a)  $D_h = 0$  meV, (b)  $D_h = 1$  meV, (c)  $D_h = 10$  meV. O gráfico abaixo da figura (b) é da região entre a intersecção dos retângulos ampliada. . . . . 157

## Lista de Tabelas

1	Resumo dos pares de base e próximos vizinhos associados aos parâmetros analisados. Na coluna $N$ , apresentamos o número total de sequências extraídas da literatura de cada base de dados. Os sinais + e – ao lado do átomo metálico indicam, respectivamente, a presença e a ausência do átomo Ag ou Hg entre os <i>mismatches</i> . O número $n$ entre parênteses, nas colunas Morse e empilhamento, informa a quantidade de vezes que cada parâmetro aparece em nossa base de dados. . . . .	63
2	Parâmetros $k$ e $D$ iniciais utilizados para as minimizações das sequências CTRL. . . . .	64
3	Parâmetros $k$ e $D$ iniciais obtidos da otimização preliminar e utilizados na primeira rodada de minimizações de C-Ag <sup>+</sup> -C e T-Hg <sup>2+</sup> -T. . . . .	65
4	Comparação entre os $\langle \Delta T \rangle$ e $\langle \chi^2 \rangle$ antes da primeira e após a última minimização na presença e ausência de metal. Para ficar mais fácil de compararmos os resultados, normalizamos o valor de $\chi^2$ dividindo-o pelo número de sequências $N$ mostrados na tabela 1. . . . .	66
5	Parâmetros para o modelo HMS. . . . .	80
6	Parâmetros para o modelo PCLA com unidades redefinidas. . . . .	82
7	Parâmetros para o modelo PCLJ. . . . .	84
8	Parâmetros para o modelo TRMF. . . . .	85
9	Resumo dos pares de base e próximos vizinhos associados aos parâmetros analisados. Na coluna $N$ , apresentamos o número total de sequências extraídas da literatura de cada base de dados. O número $n$ entre parênteses, nas colunas Morse e empilhamento, informa a quantidade de vezes que cada parâmetro aparece em nossa base de dados. . . . .	87
10	Parâmetros $k$ e $D$ iniciais utilizados para as minimizações do modelo PB. . . . .	89
11	Comparação entre os $\langle \Delta T \rangle$ e $\langle \chi^2 \rangle$ antes da primeira e após a última minimização na presença e ausência de PEG200. Para ficar mais fácil de compararmos os resultados, normalizamos o valor de $\chi^2$ dividindo-o pelo número de sequências $N$ mostrados na tabela 9. . . . .	89
12	Parâmetros utilizados para iniciar o processo de minimização dos modelos com barreira. Os parâmetros de empilhamento, $k$ e $D$ , iniciais foram os ajustados pela terceira minimização do modelo PB, mostrados nas figuras 36a e 37a. . . . .	97
13	Parâmetros de energia de barreira utilizados para iniciar o processo de minimização dos modelos com barreira. . . . .	102

14	Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo HMS <sup>-</sup> . Mostramos os resultados para as sequências em H <sub>2</sub> O e PEG200 para quatro valores diferentes de $y_e$ . $\Delta\chi^2 = \chi^2_{(HMS^-)} - \chi^2_{(PB)}$ , onde $\chi^2_{(PB)}$ se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos $f_s$ , 2. Variar os termos $f_s$ e os termos do potencial de Morse $D$ e 3. Variar os termos $f_s$ e os termos de empilhamento $k$ . . . . .	105
15	Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo HMS <sup>+</sup> . Mostramos os resultados para as sequências em H <sub>2</sub> O e PEG200 para quatro valores diferentes de $y_e$ . $\Delta\chi^2 = \chi^2_{(HMS^+)} - \chi^2_{(PB)}$ , onde $\chi^2_{(PB)}$ se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos $f_s$ , 2. Variar os termos $f_s$ e os termos do potencial de Morse $D$ e 3. Variar os termos $f_s$ e os termos de empilhamento $k$ . . . . .	106
16	Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo PCLA. Mostramos os resultados para as sequências em H <sub>2</sub> O e PEG200 para quatro valores diferentes de $d$ . $\Delta\chi^2 = \chi^2_{(PCLA)} - \chi^2_{(PB)}$ , onde $\chi^2_{(PB)}$ se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos $b$ , 2. Variar os termos $b$ e os termos do potencial de Morse $D$ e 3. Variar os termos $b$ e os termos de empilhamento $k$ . . . . .	107
17	Parâmetros para o modelo PCLA modificado. . . . .	109
18	Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo TRMF. Mostramos os resultados para as sequências em H <sub>2</sub> O e PEG200 para quatro valores diferentes de $y_o$ . $\Delta\chi^2 = \chi^2_{(TRMF)} - \chi^2_{(PB)}$ , onde $\chi^2_{(PB)}$ se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos $G$ , 2. Variar os termos $G$ e os termos do potencial de Morse $D$ e 3. Variar os termos $b$ e os termos de empilhamento $k$ . . . . .	112
19	Comparação entre os dois testes para ver qual o melhor processo de minimização para se trabalhar com o modelo PCLJ. $\Delta\chi^2 = \chi^2_{(PCLJ)} - \chi^2_{(PB)}$ , onde $\chi^2_{(PB)}$ se encontra na seção 5.5.2, tabela 11. 2. Variar apenas os termos $E$ e $D_h$ , 3. Variar os termos $E$ , $D_h$ e os termos de empilhamento $k$ . . . . .	114
20	Comparação entre os $\langle\Delta T\rangle$ e $\langle\chi^2\rangle$ encontrados pelos mapas de calor, figuras 45b e 46b, e após a última minimização na presença e ausência de PEG200. Para ficar mais fácil de compararmos os resultados, normalizamos o valor de $\chi^2$ dividindo-o pelo número de sequências $N$ mostrados na tabela 9. . . . .	116
21	Sequências de DNA na presença Ag <sup>+</sup> . . . . .	142
22	Sequências de DNA na presença Hg <sup>2+</sup> . . . . .	143
23	Sequências de DNA em PEG200. . . . .	144

24	Sequências de RNA em PEG200-20. . . . .	145
25	Sequências de RNA em PEG200-40. . . . .	146
26	Parâmetros iniciais utilizados para as minimizações do modelo HMS <sup>-</sup> . . . . .	149
27	Parâmetros iniciais utilizados para as minimizações do modelo HMS <sup>+</sup> . . . . .	151
28	Parâmetros iniciais utilizados para as minimizações do modelo PCLA. Os valores diferentes de <i>d</i> são relativos a posições de equilíbrio entre o solvente e os nucleotídeos. . . . . .	153
29	Parâmetros iniciais utilizados para as minimizações do modelo TRMF. . . . .	155
30	Parâmetros iniciais utilizados para as minimizações do modelo PCLJ. . . . .	159

# Sumário

<b>1</b>	<b>Introdução</b>	<b>25</b>
<b>2</b>	<b>Estrutura DNA e RNA</b>	<b>27</b>
2.1	Experimento de UV <i>melting</i>	38
<b>3</b>	<b>Modelos teóricos</b>	<b>44</b>
3.1	Modelo Peyrard-Bishop	44
3.1.1	Modelo PB para sequências heterogêneas	47
3.1.2	Equivalência térmica	50
3.1.3	Otimização dos parâmetros	53
3.2	Modelo de próximos vizinhos	55
<b>4</b>	<b>DNA mediado por metal</b>	<b>58</b>
4.1	Revisão bibliográfica	58
4.2	Coleta e preparação dos dados	62
4.3	Procedimento de minimização	64
4.4	Resultados	66
4.5	Conclusões e Perspectivas	73
<b>5</b>	<b>Nucleotídeos em ambiente macromolecular</b>	<b>74</b>
5.1	Revisão bibliográfica	74
5.2	Nucleotídeos em ambiente macromolecular	74
5.3	Modelos PB com barreira de potencial	78
5.4	Coleta e preparação dos dados	86
5.5	Modelo Peyrard-Bishop (PB)	88
5.5.1	Procedimento de otimização	88
5.5.2	Resultados	89
5.5.3	Conclusões e Perspectivas	93
5.6	Modelos Peyrard-Bishop (PB) com barreira de potencial	94
5.6.1	Análise inicial	94
5.6.2	Otimizações finais	99
5.6.3	Análise Inicial - Parâmetros iniciais dos modelos com barreira	99
5.6.4	Análise Inicial - Otimizações de teste	101
5.6.5	Resultados	114
5.6.6	Conclusões e Perspectivas	121

<b>6 Conclusões</b> . . . . .	<b>123</b>
<b>Referências</b> . . . . .	<b>125</b>
<b>APÊNDICE A - Sequências de DNA e RNA utilizadas nos projetos</b> . . . . .	<b>141</b>
<b>APÊNDICE B - Procura das energias de barreira iniciais para os modelos HMS, PCLA, TRMF e PCLJ</b> . . . . .	<b>147</b>
<b>APÊNDICE C - Artigo Publicado</b> . . . . .	<b>160</b>

# 1 Introdução

DNA e RNA estão entre as moléculas mais importantes presentes no corpo humano. O DNA é responsável pelo código genético, que é único para cada indivíduo, e o RNA é responsável pela codificação da síntese de proteínas, onde cada proteína possui uma função específica para manter o bom funcionamento da célula. Em ocorrências naturais são compostos basicamente por quatro bases nitrogenadas Adenina (A), Timina(T), Citosina (C) e Guanina(G) no DNA e Adenina (A), Uracila(U), Citosina (C) e Guanina(G) no RNA. O DNA é estabilizado em forma de dupla hélice por duas forças: ligações de hidrogênio entre os nucleotídeos e o empilhamento das bases (do inglês *stacking*). O RNA é encontrado nas células como fitas simples, entretanto a formação de dupla hélice de RNA é possível em regiões onde existe o pareamento entre duas sequências [1]. Além disso, em certas situações dentro da célula, por exemplo síntese de proteínas, DNA e RNA se hibridizam formando híbrido DNA/RNA [31]. Logo, entender como estas interações, ligações de hidrogênio e empilhamento, são responsáveis por estabilizar estas estruturas (DNA, RNA e DNA/RNA) é extremamente importante para compreender as funções dos nucleotídeos em nosso organismo.

Experimentos de desnaturação UV (do inglês ultra-violet *melting*) são utilizados para verificar a estabilidade de nucleotídeos em soluções aquosas [32]. Micro e macro moléculas orgânicas [33–35] e íons metálicos [36–38] presentes nas soluções podem interferir na estabilidade dos nucleotídeos, dependendo de suas concentrações. Além disto, íons metálicos estabilizam o DNA ao interagir com o grupo fosfato ou com as bases nitrogenadas [38, 39]. No entanto, o experimento de desnaturação UV não fornece informações acuradas sobre as interações moleculares presentes nos nucleotídeos. Diante disto, montamos uma base de dados para duas situações específicas, DNA mediado por metal [40] e nucleotídeos em ambiente molecular [41], com os resultados de desnaturação UV presentes na literatura, e utilizamos modelos mesoscópicos [24, 41–43], principalmente o modelo Peyrard and Bishop [43], para calcular os parâmetros das interações presentes no sistema. A qualidade dos parâmetros será avaliada comparando às temperaturas de desnaturação calculadas com às experimentais [8, 28].

No DNA mediado por metal (M-DNA) o íon metálico se liga entre às bases nitrogenadas, no qual apenas  $\text{Hg}^{2+}$  e  $\text{Ag}^+$  formam uma estrutura estável naturalmente. De acordo com a literatura, o íon  $\text{Hg}^{2+}$  estabiliza o *mismatch* TT [14, 21, 44] e o íon  $\text{Ag}^+$  [16, 37, 45] estabiliza o *mismatch* CC. O fato de que  $\text{Hg}^{2+}$  e  $\text{Ag}^+$  se ligam a dois *mismatches* distintos possibilitam o seu uso como detectores de *mismatches* [46]. Além disso, a existência de vários íons metálicos ao longo do DNA pode tornar essa estrutura útil para produção de dispositivos [40] como, por exemplo, o nanofio [12].

Nucleotídeos em ambiente macromolecular é o estudo de DNA, RNA ou híbrido DNA/RNA em soluções que simulam o conteúdo de uma célula [41]. Você pode se perguntar porque não extrair diretamente uma porção da célula e realizar as análises. O motivo é que o interior da célula passa por constantes mudanças, devido a reações físico-químicas, dificultando a validação dos experimentos [47]. Um soluto<sup>1</sup>, que não interage com os nucleotídeos, é utilizado para evitar este grau extra de complexidade. O soluto mais utilizado é o polietileno glicol (PEG) [33–35] disponível em várias massas moleculares, simulando, respectivamente, micro- e macro-moléculas presentes no interior das células. PEGs com massas moleculares menores do que 1000 se comportam como micro-moléculas e interagem com as moléculas de água, essenciais para formação do duplexo, desestabilizando os nucleotídeos. Enquanto PEGs com massas maiores se comportam como macro-moléculas confinando os nucleotídeos em um espaço menor, por conta disto, possuem o efeito contrário. [48, 49].

Este trabalho está dividido da seguinte forma: No capítulo 2 apresentamos as estruturas do DNA e RNA descrevendo a composição química destes nucleotídeos, as interações intermoleculares, os efeitos de sal e água e tipos de pareamentos diferentes de AT, ou AU para RNA, e CG denominados *mismatches*. Ao final deste capítulo apresentamos o experimento de desnaturação UV. Um importante experimento que analisa a estabilidade dos duplexos. No capítulo seguinte (capítulo 3) apresentamos os modelos mesoscópicos utilizados na análise dos dados experimentais. Falaremos especificamente de dois modelos mesoscópicos: o modelo Peyrard-Bishop (PB) e o modelo de próximos vizinhos (NN), do inglês *Nearest-Neighbors*. Nos capítulos 4 e 5 abordaremos os casos específicos em DNA mediado por metal e nucleotídeos em ambiente macromolecular nas quais realizamos a revisão bibliográfica de ambos os casos. Na capítulo 5 adicionamos uma parte que trata sobre diferentes modificações do Modelo PB para interpretar as interações dos nucleotídeos com às moléculas de água. Por fim, ainda nos capítulos 4 e 5 discutimos nossos resultados, conclusões e perspectivas futuras para os dois trabalhos.

---

<sup>1</sup>Na literatura a palavra utilizada é *cosolute*: um soluto que existe em uma solução junto com outro soluto. Apesar da tradução da palavra *cosolute* ser *cosoluto*, não encontramos essa palavra em nenhum dicionário de português.

## 2 Estrutura DNA e RNA

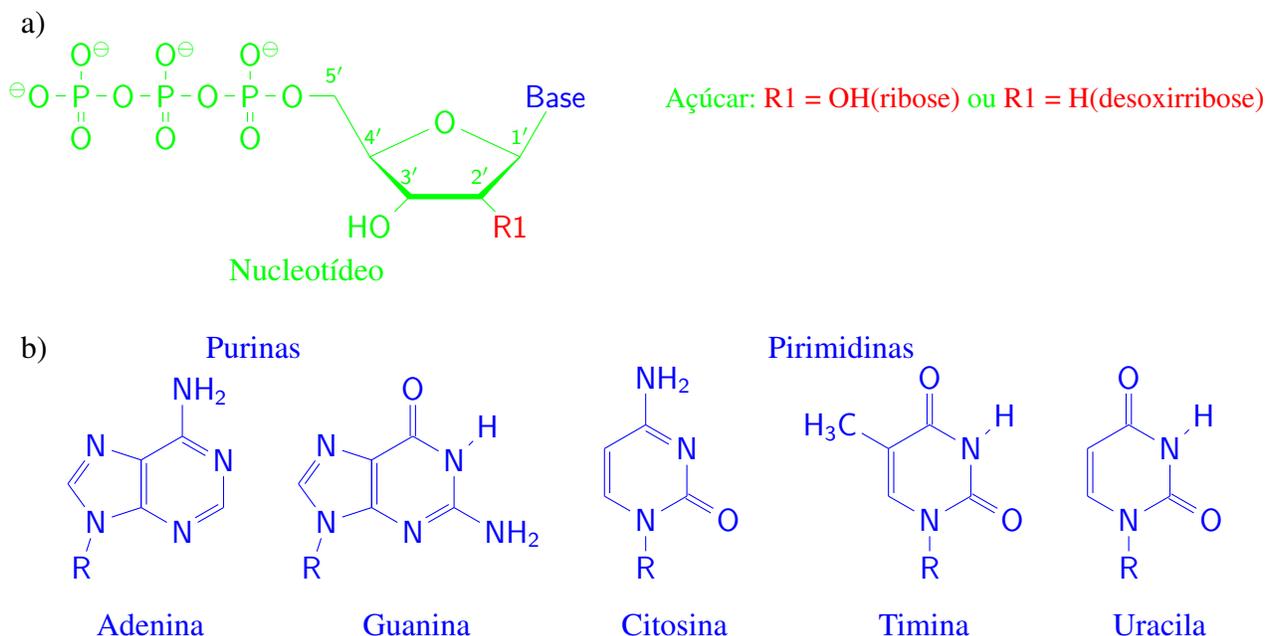
As composições químicas do DNA e RNA são bastante similares. Ambas as estruturas são compostas por blocos chamados nucleotídeos (figura 1a). Cada nucleotídeo é constituído por uma base nitrogenada, um grupo açúcar, e de um até três grupos fosfatos ( $\text{PO}_4^-$ ) [31, 50]. O que diferencia às estruturas são o grupo açúcar, no DNA temos a desoxirribose, enquanto que no RNA encontramos a ribose, e uma base nitrogenada existente apenas no DNA e uma outra base presente apenas no RNA [31, 50].

Há cinco tipos de bases nitrogenadas divididas em purinas e pirimidinas, figura 1b. As purinas são estruturas compostas por dois anéis aromáticos e as pirimidinas são formadas por um anel aromático. Assim, adenina (A) e guanina (G) são purinas e citosina (C), timina(T) e uracila(U) são pirimidinas. Do gene extraído da célula encontramos na molécula de DNA A, T, C e G, e já na molécula de RNA temos o mesmo conjunto de bases presentes no DNA, com exceção do T substituído por U.

A fita simples de DNA ou RNA é formada a partir do empilhamento dos monômeros. O empilhamento possui direcionalidade do carbono  $5'$ , início da cadeia, para o carbono  $3'$ , final da cadeia, figura 2a. Ambos os carbonos estão indicados na figura 1a. O conjunto, pentose+açúcar, é chamado esqueleto (*backbone*, em inglês) e se repete por toda a extensão do polinucleotídeo. Dessa forma, o que diferencia os nucleotídeos é a base nitrogenada. Assim, a fita de DNA que possui as bases ATTCG, por exemplo, é escrita como  $5'$ -ATTCG- $3'$ . Por esta notação sabemos exatamente qual o sentido de empilhamento dos monômeros presentes na estrutura.

Experimentos e observações de Erwin Chargaff [50, 51] em células animais constataram que o DNA possui sempre a mesma quantidade de purinas (C e T) e pirimidinas (A e G) e observaram que a razão entre C/G e A/T é sempre igual a um, independente da organização das bases na estrutura [38]. Além disso, Maurice Wilkins e Rosalind Franklin [50, 52] tinham suspeita que a estrutura era helicoidal. Assim, em 1953, com base nestas análises, Watson e Crick (WC) [53] propuseram o modelo de dupla hélice do DNA na qual C e G formam um par de base composto por três ligações de hidrogênio e A e T formam o outro par de base composto por duas ligações de hidrogênio. As fitas são enroladas de forma que não há solvente em seu interior. Com isso, o esqueleto fica em contato direto com solvente enquanto as bases nitrogenadas ficam confinadas no interior da hélice com apenas uma pequena parte exposta ao solvente, figura 2b.

A estrutura canônica de DNA proposta por Watson e Crick é a mais comumente encontrada



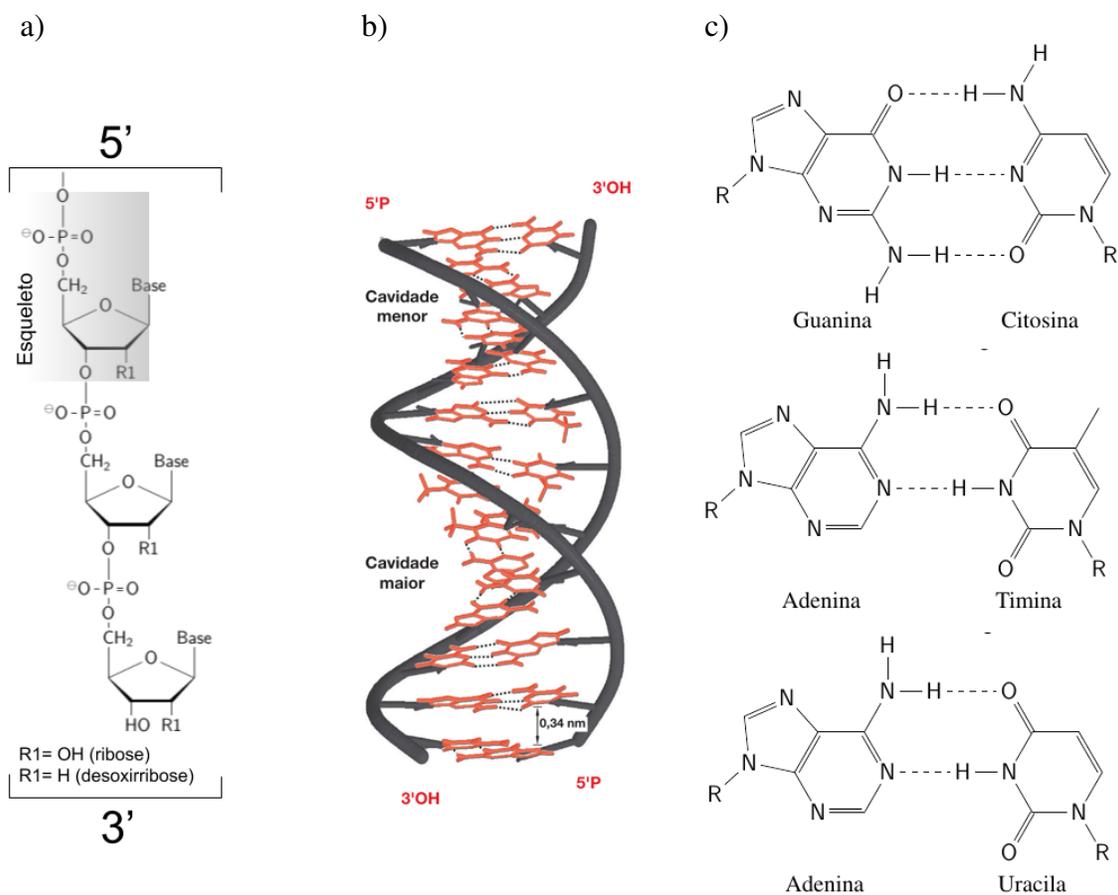
**Figura 1**

Monômero de DNA e RNA. a) Nucleotídeo desenhado em verde composto de três grupos fosfatos + pentose (ribose - RNA ou desoxirribose -DNA) + Base; b) Bases nitrogenadas. Fonte: o próprio autor

no meio fisiológico [3, 50] chamada de hélice WC do tipo b. A direcionalidade da fita complementar é inversa a da fita codante e a hélice se enrola no sentido da mão direita, figura 2b. O simples fato de as duas hélices serem anti-paralelas é extremamente importante no processo de replicação que copia a fita codante em um sentido e a fita complementar no sentido oposto [54]. Outro detalhe que observamos na figura 2b é a existência de duas cavidades denominadas cavidade maior (*major groove*, em inglês), e cavidade menor (*minor groove*, em inglês). A cavidade maior está mais exposta ao solvente do que a menor. Mais detalhes de como o solvente se organiza nas duas cavidades será discutido nos parágrafos posteriores.

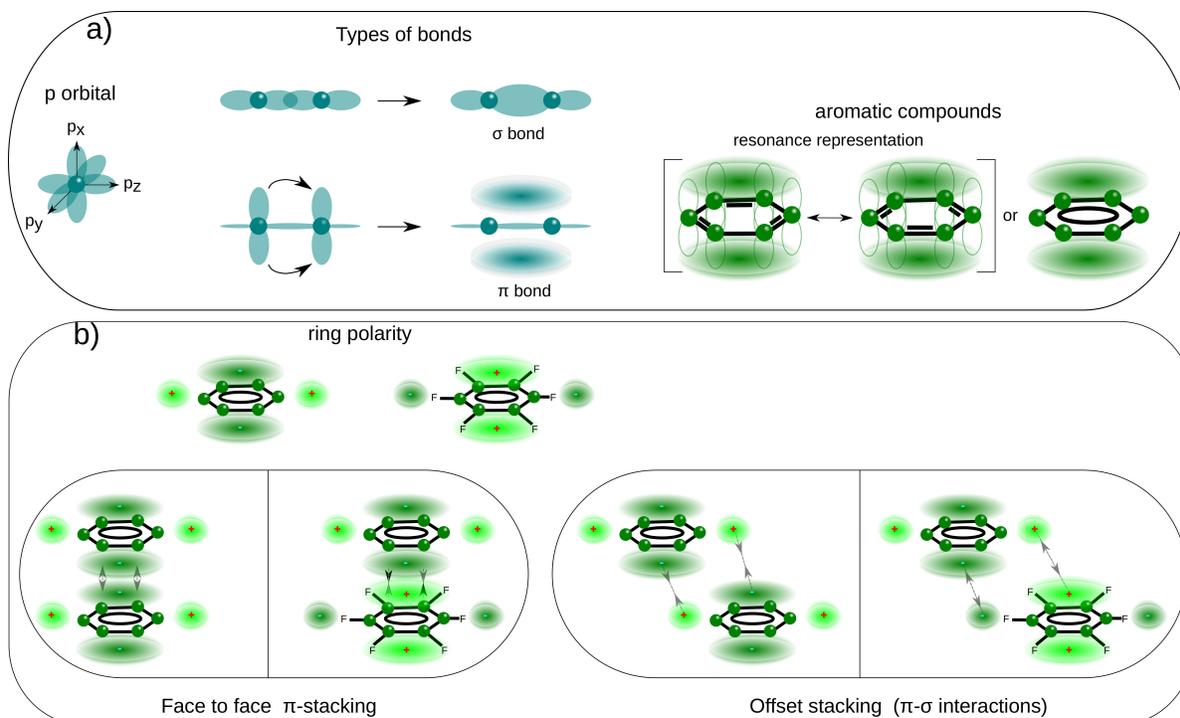
DNA é estabilizado como um duplexo na célula, por outro lado o RNA é comumente encontrado na célula como uma fita simples de diferentes tamanhos [31]. No entanto, em certas situações o RNA se hibridiza tanto com RNA quanto com DNA [31, 38, 50]. A formação de hélice dupla de RNA ocorre em regiões onde existe pareamento entre duas sequências. Neste caso, a sua estrutura assemelha-se ao DNA tipo A, figura 8a. A estrutura do DNA tipo B não é formada no RNA pareado devido ao impedimento espacial representado pelo grupamento hidroxílico na posição 2' da ribose [1], figura 1a. Os tipos de hélice de DNA serão discutidas nos parágrafos adiante.

**Interações intermoleculares** Os nucleotídeos são estabilizados principalmente por duas interações intermoleculares: ligações de hidrogênio, figura 2c, e empilhamento entre os pares de base,



**Figura 2**

Estrutura do polinucleotídeo a) Disposição dos monômeros na fita simples com destaque para o esqueleto em cinza sombreado (Fonte: o próprio autor); b) Modelo de dupla hélice proposto por Watson e Crick - DNA tipo B (figura retirada de Zaha et al. [1]); c) ligações de hidrogênio presentes nos nucleotídeos (Fonte: o próprio autor).



### Figura 3

Definição da interação de empilhamento em anéis aromáticos. bolinhas azuis claro são qualquer átomo que realiza ligações químicas compartilhando elétrons e bolinhas verdes são átomos de carbono. a) Tipos de ligações covalentes e distribuição de carga em anéis aromáticos; b) Dois tipos de interações  $\pi - \pi$  presentes em anéis aromáticos: interações sanduíche (*face to face  $\pi$  stacking*) e paralelo deslocado (*offset stacking*). Fonte: o próprio autor.

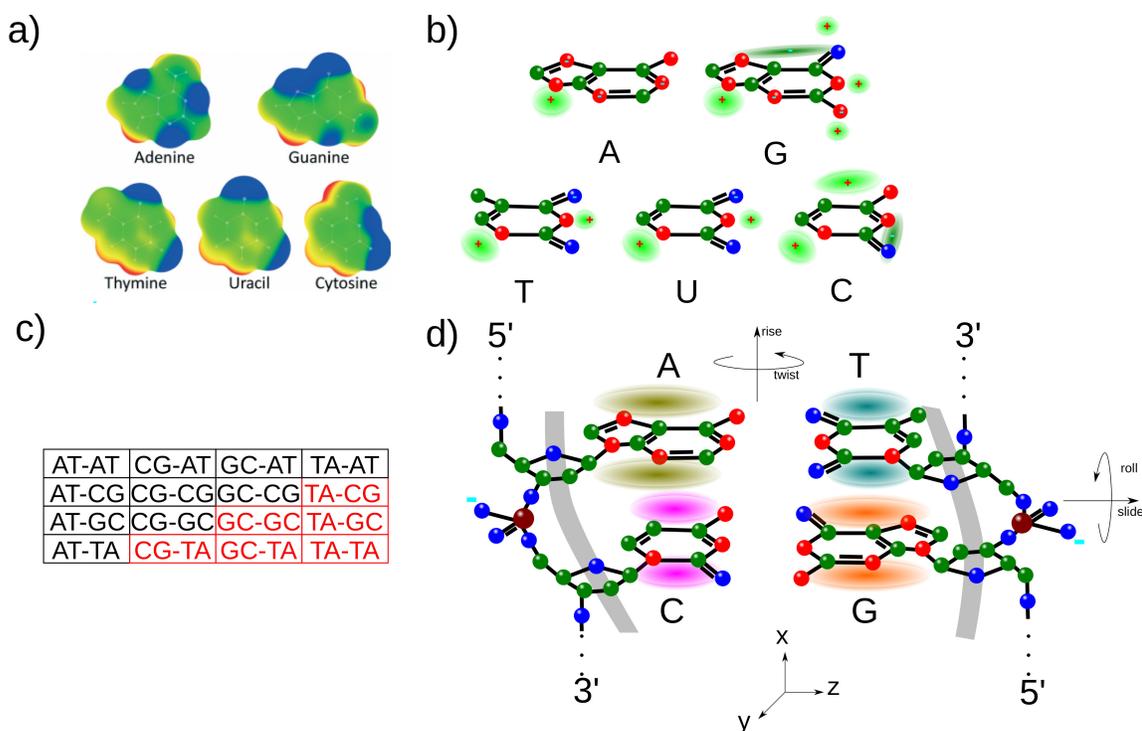
figura 5. Ligações de hidrogênio são definidas como interações intermoleculares que ocorrem entre os elementos mais eletronegativos, Fluor (F), Oxigênio (O) ou Nitrogênio (N), com átomos de hidrogênio (H). A estrutura do DNA e RNA apresenta N e O, elementos eletronegativos, e H. Foi observado que o par de base CG possui três ligações de hidrogênio, enquanto que o par AT, ou AU, possui apenas duas ligações de hidrogênio. Apesar de ambas serem de intensidade fraca quando comparadas com as ligações covalente, o par CG possui uma intensidade maior do que o par AT, ou AU para RNA em meio fisiológico devido a presença de uma ligação de hidrogênio a mais no par CG [31, 50]. Outra importante interação responsável pela estabilidade do duplexo está relacionada ao empilhamento entre as bases. Mas, antes de falar sobre a interação de empilhamento vamos definir o que significa essa interação, característica de anéis aromáticos.

Há dois tipos de ligações covalentes que os átomos realizam para formar moléculas: ligações  $\sigma$  e ligações  $\pi$ . As ligações  $\sigma$  acontecem quando há sobreposição direta dos orbitais atômicos, enquanto as ligações  $\pi$  ocorrem quando há sobreposição lateral dos orbitais, figura 3a [55]. A quantidade de elétrons desemparelhados na camada de valência dos átomos determina se a ligação covalente é simples, dupla ou tripla. As ligações simples são sempre do tipo  $\sigma$ , enquanto as ligações

duplas ou triplas ocorrem somente após os átomos já terem formado uma ligação simples. Logo, se houver ligações duplas ou triplas na estrutura, haverá sempre uma sobreposição lateral dos orbitais para formar a segunda ou terceira ligação, ou seja, ligações  $\pi$  [55–57].

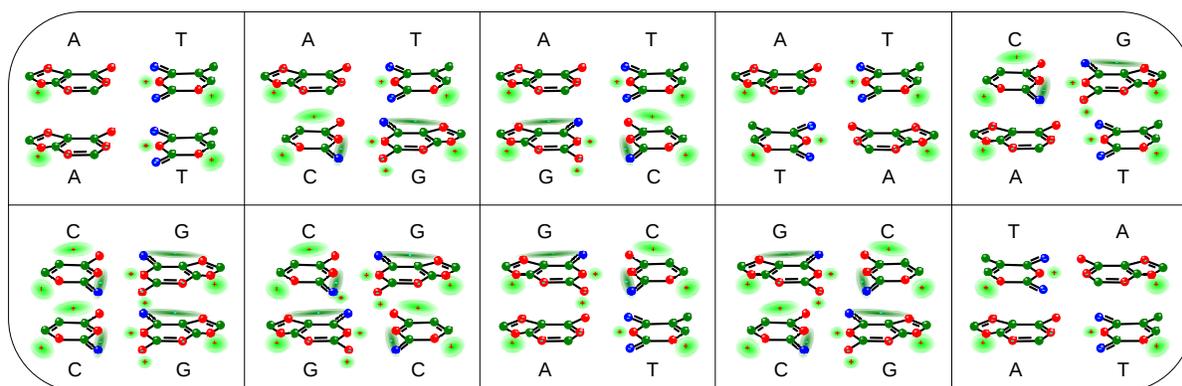
Estruturas aromáticas possuem uma característica interessante: a carga fica delocalizada por todo o anel aromático [57]. Isso é mais fácil de se observar em uma estrutura mais simples do que nas bases nitrogenadas do DNA, pois há elementos fora do anel que realizam ligações  $\pi$ . Um exemplo é o benzeno, figura 3a, onde cada átomo de carbono na estrutura realizam três ligações  $\sigma$  e uma ligação  $\pi$ . Na qual às ligações duplas se alternam entre os carbonos de duas maneiras, como pode ser visto nas duas estruturas de ressonância da figura 3a. Uma forma de mostrar que a carga fica delocalizada em toda estrutura é trocar as duplas ligações por um círculo no centro do hexágono, figura 3a. Essa figura significa que os elétrons no orbital  $p_y$  se distribuem em todos os carbonos presentes no anel aromático. Os átomos de carbono no benzeno são mais eletronegativos que os átomos de hidrogênio que se localizam fora do hexágono tornando a carga delocalizada no anel negativa enquanto os átomos fora do anel, átomos de hidrogênio, possuem carga positiva, figura 3b. Se trocarmos os átomos de hidrogênio por átomos de fluor a polaridade do anel se inverte, pois o fluor é mais eletronegativo que o carbono deslocando a carga negativa para fora do anel [58]. Por conta das polaridades dos átomos dentro e fora do anel as estruturas aromáticas realizam interações moleculares. Na figura 3b apresentamos dois tipos de interações entre os anéis aromáticos presentes no DNA. As interações  $\pi - \pi$  tipo sanduíche ocorrem quando um anel aromático encontra-se exatamente sob ou sobre um outro anel aromático. Nesta configuração temos uma interação entre os orbitais  $\pi$  de cada anel aromático e  $\sigma$  entre os elementos fora do anel. Um segundo tipo de interação  $\pi - \pi$  muito importante nos nucleotídeos é a paralelo deslocado, onde temos a interação entre os orbitais  $\pi$  do anel com às ligações  $\sigma$  fora do anel. Devido a diferenças de polaridade dentro e fora do anel às interações  $\pi - \pi$  sanduíche e paralelo deslocado são mais ou menos favoráveis [2, 59]. Homo-dímeros possuem uma interação atrativa quando estão na configuração paralelo deslocado, em contraste com a configuração do tipo sanduíche, que é repulsiva. Já nos heterodímeros, como o benzeno-hexafluorbenzeno, observamos o contrário, figura 3b. Os anéis aromáticos das bases nitrogenadas do DNA possuem elementos com diferentes eletronegatividades. Deste modo, a carga não se distribui de maneira tão uniforme nos anéis como no benzeno e no hexafluorbenzeno, figura 4a e 4b.

Há quatro tipos de bases nitrogenadas no DNA que formam pares AT, TA, CG ou GC. Na fita simples, os monômeros se organizam de forma que uma base se localize exatamente sobre ou sob a base adjacente. Para o DNA, representamos os dímeros separados por um traço entre o par de base de um ponto do duplexo e o seu próximo vizinho. Portanto, temos 16 dímeros, ou próximos vizinhos, presentes no DNA, figura 4c. No entanto, alguns desses dímeros são equivalentes, restando apenas



#### Figura 4

Distribuição de cargas nas bases nitrogenadas. a) Potencial eletrostático na superfície das bases nitrogenadas. As escalas de cor estão em kcal/mol: vermelho (maior que 31.31), amarelo (entre 10.17 até 31.31), verde (entre -10.92 até 10.17), azul (mais negativo do que -10.92). Figura retirada de Veljković et al. [2] b) Representação simplificada da distribuição de carga molecular nas bases nitrogenadas derivada do potencial eletrostático (Fonte: o próprio autor); c) Tipos de dímeros presentes no DNA (Fonte: o próprio autor); d) Principais movimentos dos dímeros presentes na hélice de DNA. Cores (figuras b e d) - vermelho: nitrogênio, verde: carbono, azul: oxigênio e marrom: fósforo (Fonte: o próprio autor).



#### Figura 5

Os 10 tipos de próximos vizinhos possíveis no DNA representados com a distribuição de carga simplificada de cada base nitrogenada. Todos os átomos de hidrogênio, não desenhados na figura, possuem carga positiva. Cores - vermelho: nitrogênio, verde: carbono e azul: oxigênio (Fonte: o próprio autor).

10, que estão destacados em preto na figura 4c. Isto pode ser facilmente visualizado na figura 4d, onde temos uma porção da hélice de DNA com o dímero AT-CG, onde 5'-A/C-3' se localiza na fita codante e 3'-T/G-5' na fita complementar. Se rodarmos o dímero em 180° em torno do eixo y, a fita codante se transforma na fita complementar e encontramos o dímero GC-TA, onde 5'-G/T-3' é a fita codante e 3'-C/A-5' a fita complementar. Ao comparar os dois dímeros, temos a mesma representação na qual apenas modificamos a referência da fita. Logo, AT-CG=GC-TA diminuindo o número de próximos vizinhos. Este mesmo processo pode ser aplicado aos demais dímeros, figura 4c. A redução do número de próximos vizinhos presentes no DNA só acontece porque as fitas são antiparalelas e o esqueleto é idêntico, por exemplo no híbrido DNA/RNA as 16 combinações são distintas, pois em uma fita temos a desoxirribose e na outra fita a ribose.

Na figura 5, apresentamos os dez tipos de próximos vizinhos presentes no DNA, com a distribuição de carga simplificada de cada base nitrogenada. As interações  $\pi - \pi$  tipo sanduíche ocorrem na fita simples de DNA e são chamadas de *intrastrand* [60, 61]. Por outro lado, as interações  $\pi - \pi$  paralelo deslocado acontecem entre a fita codante e a fita complementar, onde as bases se localizam em planos diferentes. A soma das ligações de hidrogênio com as interações  $\pi - \pi$  paralelo deslocado é chamada de *interstrand* [60, 61]. Um fato interessante é que, caso a polaridade do anel seja inversa à dos átomos que se localizam fora do anel, as ligações de hidrogênio são favorecidas [62].

A distribuição de cargas no anel é aproximadamente igual em todas as bases nitrogenadas, figura 4a. Portanto, a interação *interstrand* é repulsiva. Ao empilharmos um terceiro par de bases nos dímeros mostrados na figura 5, o par que se localiza no meio se movimenta para cima e para baixo devido à repulsão eletrostática com os primeiros vizinhos. Esse deslocamento é chamado de subida, *rise* em inglês. Para o par de bases na extremidade da fita não há uma base acima para mantê-lo na estrutura. Assim, o que mantém o nucleotídeo na fita simples é a ligação fosfodiéster, figura 4d.

O deslizamento, do inglês *slide*, das bases ocorre devido à interação de *interstrand* entre a fita codante e a fita complementar. Seguindo a mesma linha de raciocínio que aplicamos para o movimento de subida explicado anteriormente, onde inserimos um terceiro par de base nos dímeros mostrados na figura 5. Podemos ter bases que favorecem ou não às ligações de hidrogênio por conta das diferentes polaridades entre o anel e os elementos fora do anel. Nesta situação, as bases se movimentam para frente e para traz no sentido do eixo z, figura 4d.

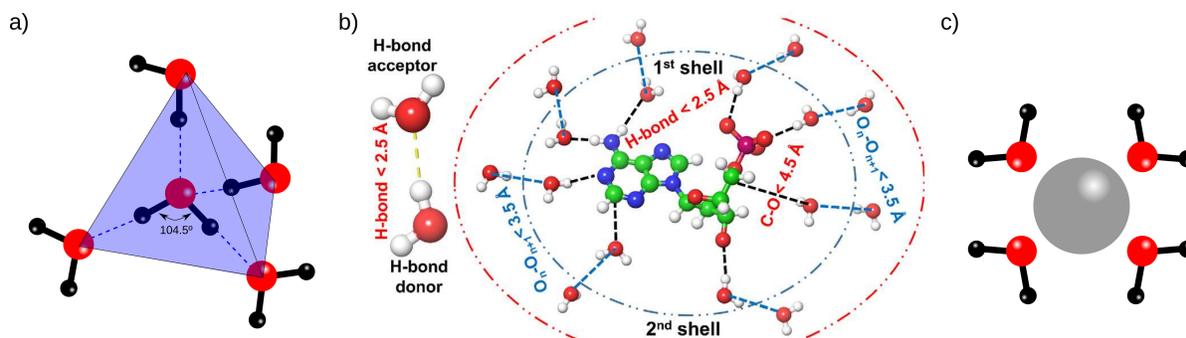
Os movimentos de torção, do inglês *twist*, e rolamento, do inglês *roll* que ocorrem no eixo x e z, respectivamente, figura 4d, acontecem porque a carga não se distribui uniformemente dentro e fora do anel, figura 4a. Logo, a base nitrogenada torce e rola para minimizar a energia eletrostática entre os dímeros na estrutura do DNA. Entretanto, por causa da estrutura química das bases nitrogenadas

temos restrições estéricas tanto para a torção quanto para o rolamento [62]. Por exemplo, a presença do grupo metil na timina restringe o movimento de rolamento. O metil é repelido pela pentose, figura 4d. Um outro exemplo acontece quando temos uma pirimidina sob ou sobre uma outra pirimidina, figura 5. Por conta do tamanho dos anéis na estrutura temos uma restrição no movimento de rolamento. Já a restrição na torção ocorre entre as bases e o esqueleto do DNA.

Ambas as interações intermoleculares são de extrema importância para a manutenção da dupla hélice de DNA. No entanto, a forte repulsão eletrostática entre as fitas, causada pelo grupo fosfato, interfere consideravelmente em sua estabilidade. Tanto às ligações de hidrogênio quanto as interações de empilhamento são afetadas [5, 63, 64]. Além disso, o simples fato de o par CG realizar mais ligações de hidrogênio do que o par AT, por si só, pode causar enormes diferenças de energia entre o par AT e CG no duplexo. Uma consequência é a mudança na conformação global da dupla fita [3, 4, 65]. Logo, o meio onde os nucleotídeos se encontram tem um papel fundamental na redução tanto da repulsão eletrostática do grupo fosfato quanto a diferença de energia entre o par AT e o par CG.

**Efeitos da água e do sal** Na célula existem moléculas de água, cátions e biomoléculas [31]. Os cátions são íons metálicos que são metais alcalinos,  $\text{Na}^+$  e  $\text{K}^+$ , alcalinos-terrosos,  $\text{Mg}^+$  e  $\text{Ca}^+$ , ou sais orgânicos como por exemplo amônia (em baixa concentração devido a sua toxicidade) [50]. As biomoléculas são compostas por lipídios, carboidratos, proteínas e ácidos nucleicos [31, 50]. As moléculas de água e os cátions são os principais responsáveis em reduzir a diferença de energia entre as ligações de hidrogênio e a repulsão eletrostática do grupo fosfato na dupla hélice, respectivamente [4].

Nas moléculas de água o átomo de oxigênio liga covalentemente a dois átomos de hidrogênio. Por causa da repulsão eletrostática entre os dois pares de base não ligantes no subnível p do oxigênio, a água possui uma geometria angular, cujo, o ângulo da ligação O-H é  $104.5^\circ$ , figura 6a. A simetria angular da molécula de água confere a ela um momento de dipolo na direção do elemento mais eletronegativo, neste caso, o átomo de oxigênio. A polaridade da água permite que ela atraia outras águas eletrostaticamente estabelecendo ligações de hidrogênio entre o oxigênio de uma molécula com o hidrogênio de outra molécula. São consideradas ligações de hidrogênio quando as moléculas de água se encontram a uma distância menor do que  $2.5 \text{ \AA}$  [3]. Os oxigênios da água atuam como aceitadores (a) de carga, enquanto os hidrogênios são doadores (d) de carga 6b. Há dois pares de elétrons não ligantes no oxigênio e dois átomos de hidrogênio na água. Isso permite que cada molécula de água forme até quatro interações moleculares com outras águas, 6a. No estado líquido não temos uma forma regular como no estado sólido tornando às ligações de hidrogênio instáveis, constantemente formando e quebrando estas interações [50]. A elevada tensão superficial



**Figura 6**

Interações das moléculas de água com moléculas responsáveis por estabilizar os nucleotídeos. a) Associação das moléculas de água formando um arranjo tetraédrico. A esfera vermelha representa o oxigênio, as esferas pretas o hidrogênio e as linhas tracejadas azuis às ligações de hidrogênio (Fonte: o próprio autor); b) Definição de conchas de hidratação mostradas no exemplo com o nucleotídeo adenina. Cores - vermelho: oxigênio, verde: carbono, azul: nitrogênio, branco: hidrogênio e roxo: fósforo. Figura retirada de Rozza et al. [3]; c) Interação dos cátions, esfera em cinza, com às moléculas de água (Fonte: o próprio autor).

e a viscosidade da água, quando comparada com a maioria dos líquidos orgânicos, podem ser atribuídas à ligação de hidrogênio que ocorre entre as moléculas de água. No entanto, como as ligações de hidrogênio na água líquida sofrem flutuações constantes, ela possui maior mobilidade e menor viscosidade em comparação com outras moléculas poliméricas [50].

O caráter polar da água permite que ela interaja com outras substâncias, por exemplo, compostos iônicos (sais, como sal de cozinha), outras moléculas polares não iônicas (álcoois, como o etanol) e moléculas parcialmente polares (sabão) [50]. Compostos apolares são insolúveis em água, pois não há uma atração entre as moléculas. Neste caso, observamos duas fases na solução quando misturamos líquidos polares e apolares, por exemplo, uma solução de água e óleo. Moléculas que interagem bem com a água são chamadas hidrofílicas e moléculas que não interagem com a água são chamadas hidrofóbicas. Ao entrar em contato com a água os sais iônicos se separam facilmente entre ânions e cátions em um processo chamado dissociação. O oxigênio da água, com polaridade negativa, solvata o cátion, figura 6c, enquanto os hidrogênios da água solvatam o ânion formando uma casca de hidratação ao redor destes íons.

Moléculas polares e parcialmente polares também são solvatadas pelas moléculas de água. DNA possui componentes polares e apolares em sua estrutura. Os átomos de nitrogênio, oxigênio e hidrogênios no DNA interagem hidrofílicamente com às moléculas de água, enquanto os átomos de carbono interagem hidrofobicamente, figura 6b. Segundo Rozza et al. [3], as moléculas de água ao redor do DNA podem ser divididas em três classes de interações com os nucleotídeos:

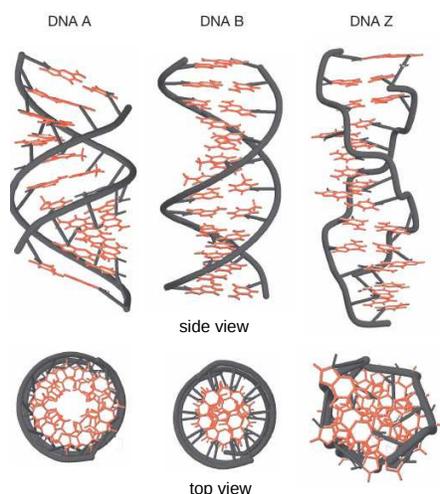
- *Bulk waters*, ou águas que não interagem com o DNA. Estas moléculas de água se comportam como a água líquida. Portanto não são afetadas pelo DNA na solução;

- Águas de superfície, ou águas biológicas. Elas interagem principalmente com o esqueleto do DNA;
- Moléculas de água interiores. Estas águas são encontradas nas cavidades maior e menor do DNA e interagem fortemente com os pares de base.

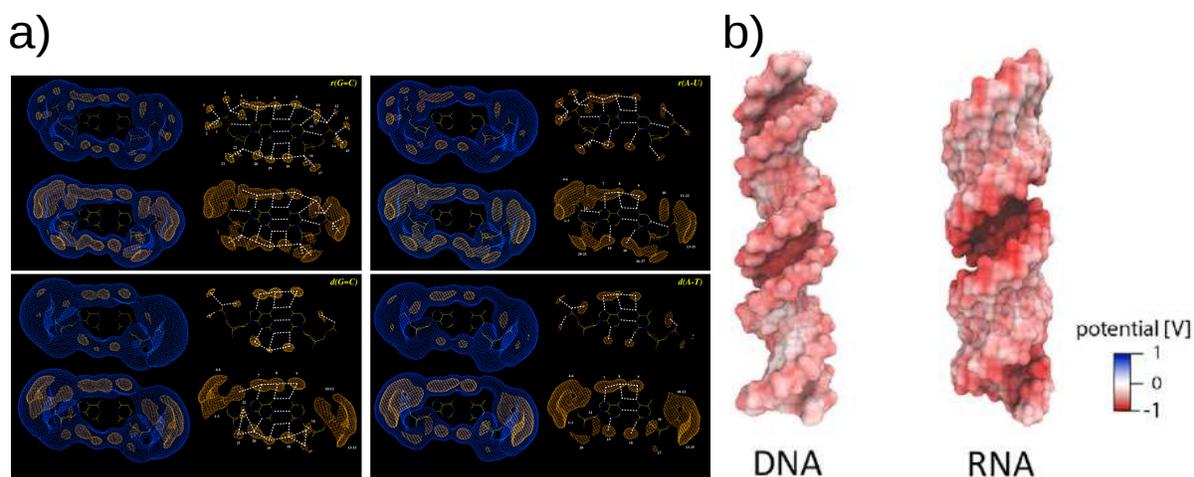
Análises experimentais mostram que 40% das moléculas de água em torno do DNA estão a uma distância de  $0.5 \text{ \AA}$  [65]. Essas moléculas de água interagem mais com as bases nitrogenadas (moléculas de água interiores). No entanto, também existem moléculas de água que interagem com a pentose e com o grupo fosfato (águas de superfície) [3]. Para melhor entender a interação das moléculas de água com os nucleotídeos podemos separar as moléculas de água em cascas de hidratação até atingirmos as águas que não interagem com o DNA, *Bulk waters*. Estudos experimentais [65–68] e teóricos [3, 4, 69, 70] utilizam esta abordagem para separar o efeito de águas próximas de águas distantes do DNA. Só lembrando que as ligações de hidrogênio ocorrem quando a distância entre os elementos mais eletronegativos, nitrogênio e oxigênio presentes nos nucleotídeos, e o hidrogênio é menor do que  $2.5 \text{ \AA}$ . Deste modo podemos definir as cascas de hidratação, figura 6b. A 1° casca contém todas as moléculas de água distantes até  $2.5 \text{ \AA}$  dos nucleotídeos, já a 2° casca de hidratação contém moléculas de água na distância entre  $2.5 \text{ \AA}$  e  $5 \text{ \AA}$ , figura 6b, a 3° casca de hidratação entre  $5 \text{ \AA}$  e  $7.5 \text{ \AA}$  e assim por diante. Podemos definir o número de cascas de hidratação que desejarmos. No entanto, as águas a partir da 2° casca possuem interação mais fraca com os nucleotídeos, pois elas estão mais distantes do DNA. Ou seja, as águas nas demais cascas interagem mais fracamente com o nucleotídeo e eventualmente atingiremos as águas que não interagem com o DNA, *Bulk waters*.

A distribuição das moléculas de água na primeira casca de hidratação difere das outras cascas. A partir da segunda casca em diante, a forma predominante de água é tetracoordenada, com duas interações doadoras e duas aceptoras (2d/2a, figura 6a). Na 1° casca o número de águas doadoras e aceptoras é menor, devido à interação das moléculas de água com os nucleotídeos e também pela diferença de tamanhos entre as cavidades maior e menor do DNA [3].

O nível de hidratação do DNA determina qual tipo de conformação se encontra a dupla hélice. As moléculas de água estão constantemente ao redor do duplexo e suas propriedades dinâmicas são cruciais para preservar a estabilidade estrutural e facilitar o papel biológico da forma dominante [3]. Elevada hidratação é observada em torno da hélice de mão direita tipo B, figura 7, enquanto uma diminuição no nível de hidratação, causada pela presença de sal ou solventes orgânicos, pode levar a uma transição para formas alternativas, como DNA tipo A e DNA tipo Z [1, 3, 4]. Hélices do tipo A são formadas também por fitas duplas de RNA e mistas de DNA/RNA. Híbridos DNA/RNA são formados durante o processo de transcrição e fitas duplas de RNA são encontradas em genomas de alguns vírus e, na célula, nos RNAs transportadores e ribossômicos. [1] O DNA tipo Z ocorre em

**Figura 7**

Três tipos de hélices de DNA observadas em meio fisiológico. DNA tipo A, DNA tipo B e DNA tipo Z. Figura retirada de Zaha et al. [1].

**Figura 8**

Distribuição de íons e águas no DNA e RNA a) Distribuição das moléculas de água ao redor dos pares de base r(G=C) e r(A-U), RNA, d(G=C) e d(A-T), DNA. Figura retirada de Auffinger and Westhof [4]; b) Distribuição dos íons de  $\text{Na}^+$  nas hélices de DNA tipo B e RNA tipo A. Figura retirada de Gebala and Herschlag [5]

regiões curtas do DNA em alta concentração salina, onde existem sequências de citosinas e guaninas seguidas [71].

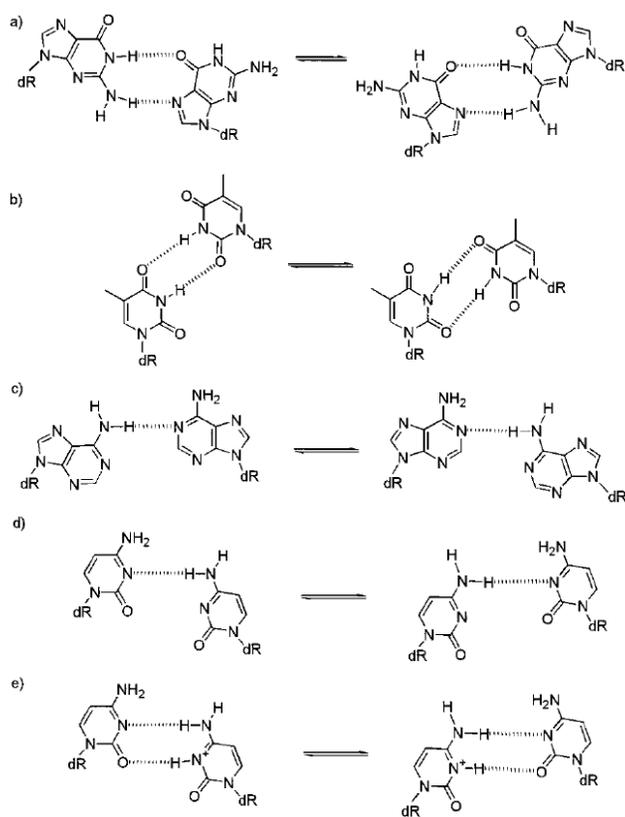
Estudos teóricos e experimentais revelaram que o duplexo de RNA concentra mais cargas e é mais hidratado do que o duplexo de DNA [3–5, 64, 69, 70], figura 8. Isso ocorre por causa do grupo açúcar, que no DNA é a desoxirribose e no RNA é a ribose, e pela diferença entre timina e uracila, metilação do carbono C(5) na timina inexistente na uracila, figura 8a. Outro aspecto determinante é a conformação da hélice do RNA que é predominantemente do tipo A, enquanto o DNA é do tipo B. As diferentes topologias dos ácidos nucleicos conferem distribuições de carga únicas para DNA e RNA, figura 8b. Para o RNA as cargas se encontram mais próximas, resultando em uma energia potencial maior [64].

**Mismatches** Em algumas situações específicas, o pareamento dos pares de base no DNA é diferente de AT e CG. O emparelhamento distinto do convencional é chamado de *Mismatch*. *Mismatches*, no DNA genômico, podem surgir de vários fatores, incluindo erros durante a replicação do DNA, exposição a produtos químicos mutagênicos, radiação ionizante ou desaminação espontânea da citosina metilada [15, 72]. Quando esses *Mismatches* ocorrem, passam por uma minuciosa inspeção e correção por meio de uma ampla gama de mecanismos de reparo. Essa tarefa crucial de corrigir esses *Mismatches* é atribuída a um conjunto de enzimas especializadas chamadas enzimas de reparo de *Mismatches* [72]. No entanto, em determinadas situações, apesar dos melhores esforços, o processo de reparo pode falhar, potencialmente levando à ocorrência de mutações. Existem oito pares de bases não correspondentes no DNA: AA, AC, AG, CC, CT, GG, GT e TT, que podem influenciar a estabilidade e as propriedades estruturais da hélice de DNA. Os pares de bases não correspondentes alteram a conformação das ligações de hidrogênio em comparação com os pares de bases canônicos, bem como as interações de empilhamento. Simulações de Dinâmica Molecular e experimentos de Ressonância Magnética Nuclear mostraram que os *Mismatches* não afetam significativamente a conformação global da dupla hélice DNA, como é o caso dos pares AA e TT [72, 73]. No entanto, em uma conformação local, a influência dos *Mismatches* varia de casos de fraca ligação (par CC) a casos de forte ligação (par GG) [72]. Além disso, outros fatores afetam os *Mismatches*, como valores de pH, concentração de sal e os pares de base adjacentes ao *Mismatch*, ou seja, eles são dependentes do contexto [15, 74]. Na figura 9 apresentamos exemplos de pareamentos realizados por *Mismatches*.

Na seção seguinte, apresentaremos o experimento de desnaturação UV (do inglês *UV melting*), explicando como são realizadas as análises experimentais. Este importante experimento avalia a estabilidade térmica do duplexo.

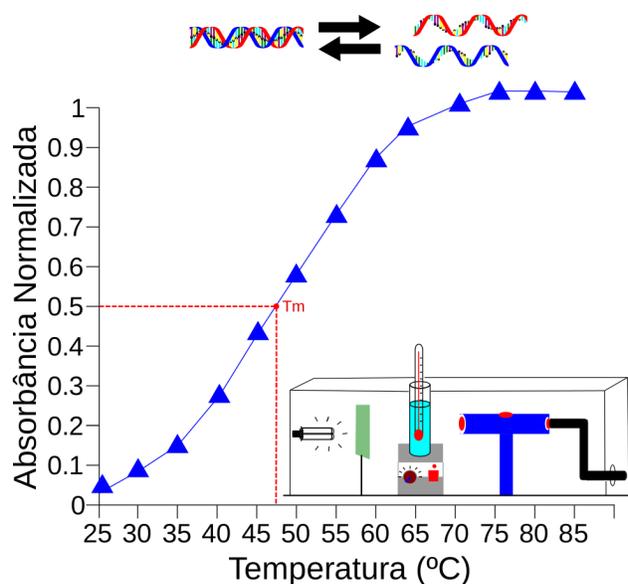
## 2.1 Experimento de UV *melting*

O experimento UV *melting* consiste de um banho térmico contendo DNA, sal e tampão, um espectrômetro de UV na faixa de 260–270 nm, que é a faixa de absorção da molécula de DNA, equipado com controlador de temperatura (figura 10) [32, 75]. Em baixas temperaturas, há apenas fitas duplas. A temperatura é aumentada gradativamente quebrando as ligações de hidrogênio e separando a fita de DNA em fitas simples. Durante o processo, essa transição é observada no espectro (linha azul na figura 10). A temperatura de desnaturação, ou temperatura de *melting* ( $T_m$ ), é a temperatura na qual a concentração de fitas duplas é igual à concentração de fitas simples. De acordo com o exemplo da figura 10,  $T_m \approx 47.5^\circ\text{C}$  (ponto de absorbância normalizado igual a 0.5 - linha tracejada em vermelho na figura 10).  $T_m$  é usado para medir a estabilidade térmica do DNA, quanto maior  $T_m$ , maior é a estabilidade da estrutura. Ademais, este processo é reversível. Ou seja,



### Figura 9

Alguns tipos de emparelhamentos realizados pelos *mismatches*. Na imagem a) mostramos o par GG, na b) o par TT, na c) o par AA, na d) o par CC e na e) o par CC protonado. Figura retirada de Peyret et al. [6]

**Figura 10**

Esquema de um experimento de UV *melting* e exemplo de espectro de absorção normalizado. Fonte: o próprio autor.

ao resfriarmos a solução, as fitas antes separadas voltam a se unir, formando o duplexo de DNA novamente. Este processo é chamado de hibridização ou *annealing*.

O tampão é usado para controlar o pH da solução. Já o sal é utilizado para favorecer a formação da fita dupla, pois a fita simples de DNA possui grupos fosfatos carregados negativamente que se repelem, e conseqüentemente atrapalham a formação da dupla hélice [76].

Na temperatura de desnaturação temos a seguinte equação química: fita dupla  $\rightleftharpoons$  fita codante + fita complementar. Para facilitar a análise, definiremos o estado de fita dupla como  $X \cdot Y$ , o da fita codante como  $X$  e o da fita complementar como  $Y$ . Dessa forma a relação de equilíbrio do sistema passa a ser descrita da seguinte forma



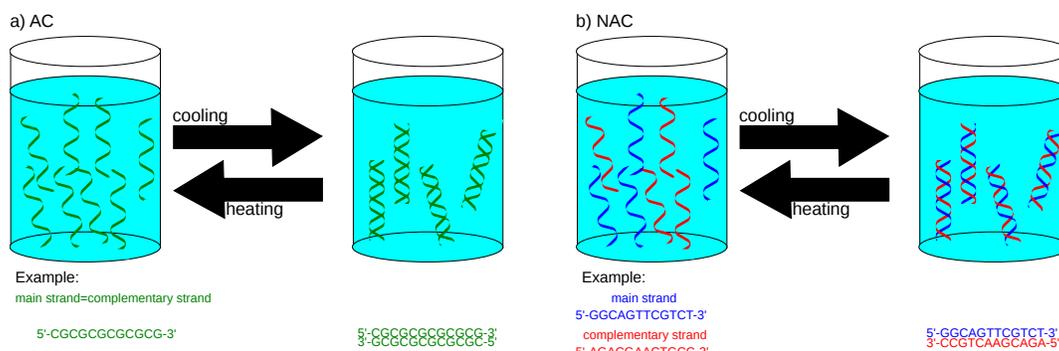
A constante de equilíbrio de dissociação  $K_d$  para a equação (1) é

$$K_d = \frac{[X][Y]}{[X \cdot Y]} \quad (2)$$

. Normalmente a constante de equilíbrio  $K_d$  para dissociação é descrita em termos da concentração total,  $C_t$ , de fitas na solução. O conceito de  $C_t$  é facilmente entendido quando pensamos em todas as hélices da solução dissociadas, isto é

$$C_t = [X]_{\text{tot}} + [Y]_{\text{tot}} \quad (3)$$

Entretanto, em meio ao processo de desnaturação podemos ter frações de hélices dissociadas e



**Figura 11**

Tipos de sequências envolvidas no experimento de UV *melting*. a) Sequências auto-complementares (AC) com as fitas representadas em verde e b) Sequências não auto-complementares (NAC) com uma espécie de fita em azul e a outra em vermelho. Fonte: o próprio autor.

ligadas na solução e a identidade dessas sequências é importante no fenômeno [77].

Existem dois tipos de sequências utilizadas no experimento: sequências auto-complementares (AC) e sequências não auto-complementares (NAC), figura 11. Se o conjunto de sequências for formado por apenas um dos tipos mencionados anteriormente não há distinção entre os sistemas na análise termodinâmica, porém se o conjunto de sequências analisado é composto por uma mistura dos dois tipos, a análise muda um pouco. Simplesmente porque sequências AC são aquelas em que a fita codante é idêntica a fita complementar (invertendo a fita codante temos a fita complementar), ou seja, a probabilidade de hibridização é maior, diferente das sequências NAC (fita codante  $\neq$  fita complementar) [78].

Por exemplo, uma concentração de 2  $\mu\text{M}$  de fitas simples AC produz 1  $\mu\text{M}$  de fitas duplas após o resfriamento da solução. A probabilidade de as fitas se hibridizarem na solução é 100%, pois se qualquer uma das fitas verdes, figura 11, encontrar qualquer outra fita verde na solução, as fitas hibridizarão. Para obtermos a mesma concentração de fitas duplas para sequências NAC, utilizamos 1  $\mu\text{M}$  de fita codante e 1  $\mu\text{M}$  de fita complementar, porém a probabilidade de hibridização cai para 50%. Neste caso, a hibridização acontecerá somente quando uma fita de cor azul encontrar uma fita de cor vermelha, figura 11. Apesar da concentração de duplexos ser a mesma nos dois casos, a população de fitas simples no caso AC é maior do que no caso NAC quando observamos apenas um tipo de fita presente na solução, como pode ser visto na figura 11. Logo, a velocidade de reação é maior no caso AC.

Note que na figura 11 temos 4 duplexos de DNA em ambos os casos, entretanto o número de fitas simples é diferente. No caso AC temos 8 fitas simples enquanto que no caso NAC temos 4 fitas simples principais (fitas vermelhas, por exemplo) e 4 fitas complementares (fitas azuis). Portanto, existe uma diferença de fator 2 entre os tipos de sequências fazendo a velocidade de hibridização AC

maior do que NAC. Assim, ao compararmos conjuntos de sequências AC e NAC uma multiplicação (ou divisão) por 2 é necessária.  $C_t$  foi definido na equação (3) como a concentração total de fitas simples presente na solução. Ao comparar os dois casos temos que  $C_t^{NAC} = 2C_t^{AC}$ . Logo, a constante de equilíbrio é diferente nestes dois casos particulares.

**Fitas NAC:** Neste contexto, a fração de duplexos dissociados é dado por

$$f = \frac{[X]}{C_t} = \frac{[Y]}{C_t}, \quad (4)$$

e a concentração total de fitas em uma dada temperatura é descrita por  $C_t = [X] + [Y] + 2[X \cdot Y]$ . Com um pouco de manipulação algébrica encontramos a relação entre a constante de equilíbrio, equação (2), concentração total e fração de duplexos dissociados

$$K_d = f^2 \frac{C_t}{2(1-f)}. \quad (5)$$

Na temperatura de desnaturação a fração de duplexos dissociados é 1/2, logo a constante de equilíbrio  $K_d = C_t/4$  para sequências NAC.

**Fitas AC:** Neste caso, temos apenas um tipo de fita simples na solução  $[X] = [Y]$ . Portanto, a fração de duplexos dissociados da equação (4) acontece apenas para  $[X]$  ou para  $[Y]$ . Escolhemos trabalhar com o caso  $[X]$ . Isso leva a concentração total de fitas em uma dada temperatura para  $C_t = [X] + 2[X \cdot X]$ . Realizando um tratamento semelhante ao que fizemos para o caso NAC encontramos a constante de equilíbrio, equação (2), reescrita em termos da concentração total e fração de duplexos dissociados para as sequências AC da forma

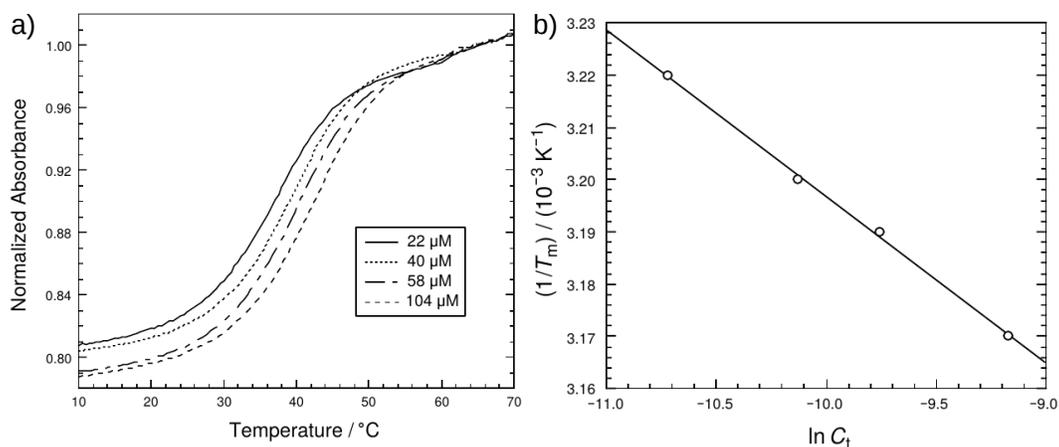
$$K_d = 2f^2 \frac{C_t}{(1-f)}. \quad (6)$$

Na temperatura de desnaturação a fração de duplexos dissociados é 1/2, logo a constante de equilíbrio  $K_d = C_t$  para sequências AC.

Definida as constantes de equilíbrio para as sequências AC e NAC podemos fazer o tratamento termodinâmico utilizando a equação de Van't Hoff

$$\Delta G = -RT \ln K_d. \quad (7)$$

A equação (7) pode ser reescrita em termos da variação de entalpia,  $\Delta H$ , e entropia,  $\Delta S$ , do sistema através da equação  $\Delta G = \Delta H - T\Delta S$ . Substituindo as constantes de equilíbrio  $K_d$  definidas para



**Figura 12**

Curvas experimentais do experimento de desnaturação de DNA do duplexo 5'-CAAAAAAAG-3'. a) Curvas de desnaturação em função da temperatura em diferentes concentrações totais  $C_t$ . b) Gráfico de Van't Hoff onde cada ponto do gráfico representa a temperatura de desnaturação do gráfico a) e sua respectiva concentração  $C_t$ . Figura retirada de Howard [7]

cada tipo de sequência, equações (5) e (6), na equação de Van't Hoff, encontramos a dependência da temperatura de desnaturação com as concentrações totais  $C_t$

$$T_m = -\frac{\Delta G}{R \ln(C_t/n)} \text{ ou } T_m = \frac{\Delta H}{\Delta S - R \ln(C_t/n)} \quad (8)$$

onde,  $R$  é a constante universal dos gases ideais,  $C_t$  é a concentração total de fitas e  $n = 1$ , caso as sequências sejam AC, ou  $n = 4$ , caso as sequências sejam NAC.

Através de uma regressão linear utilizando o método de Van't Hoff, podemos extrair os parâmetros termodinâmicos de entalpia  $\Delta H$  e entropia  $\Delta S$  ou  $\Delta G$ , conforme as equações (8) a uma temperatura fixa. A visualização do método de Van't Hoff pode ser vista na figura 12 no caso de uma dupla hélice de DNA.

No próximo capítulo, falaremos sobre os modelos mesoscópicos utilizados para a análise dos dados experimentais. Neste capítulo, introduziremos dois modelos: o modelo Peyrard-Bishop e o modelo de próximos vizinhos.

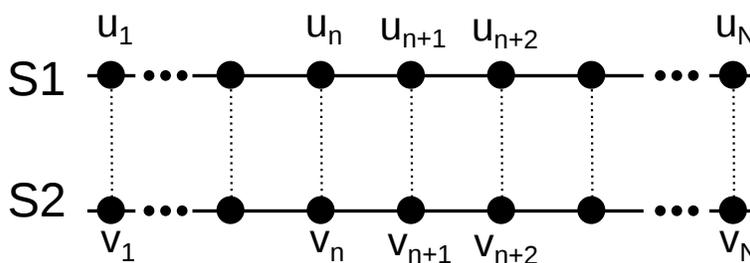
### 3 Modelos teóricos

#### 3.1 Modelo Peyrard-Bishop

A série de modelos de Peyrard-Bishop (PB) teve início em 1989 com os pesquisadores que dão nome ao modelo, Peyrard and Bishop [43]. Este modelo bi-dimensional, figura 13, busca descrever estatisticamente a dinâmica do DNA através de potenciais simples. Por exemplo, o empilhamento entre as bases (do inglês *stacking*) presentes em uma única fita é descrito por um potencial harmônico. Já a interação de hidrogênio entre a fita codante e a fita complementar é descrita pelo potencial de Morse. O modelo consiste de duas cadeias de moléculas que interagem entre si como mostrado na figura 13. S1 e S2 são os nomes dados às fitas simples de DNA. As linhas contínuas e tracejadas entre os pontos na figura 13 representam o empilhamento e as ligações de hidrogênio, respectivamente.  $u_n$  e  $v_n$  são os deslocamentos das bases de suas posições de equilíbrio na direção da ligação de hidrogênio que conecta as fitas S1 e S2. Por simplicidade, os nucleotídeos foram reduzidos a massas pontuais de valor  $m$ , pois as bases A, T, C e G possuem massas moleculares muito próximas. Ademais, a constante elástica de acoplamento entre as bases possui o mesmo valor  $k$  ao longo das fitas S1 e S2. Com base nestas informações o Hamiltoniano do sistema é

$$H = \sum_n \frac{1}{2} m (\dot{u}_n^2 + \dot{v}_n^2) + \frac{1}{2} k [(u_n - u_{n-1})^2 + (v_n - v_{n-1})^2] + D(e^{-a(u_n - v_n)} - 1)^2. \quad (9)$$

O Hamiltoniano do modelo PB consiste basicamente de três termos. O 1º termo é a energia cinética, o 2º termo é o empilhamento entre as bases representado pelo potencial harmônico e o último termo é a interação de hidrogênio representado pelo potencial de Morse. As constantes  $D$  e  $a$  do potencial de Morse são, respectivamente, a profundidade e a largura do poço de potencial. O valor de  $D$  nos informa o quão forte os pares de base estão ligados, ou seja, se  $D$  for alto mais energia será necessária para separar o par de bases. As variáveis da equação (9) estão acopladas, e isso dificulta o



**Figura 13**

Modelo Peyrard-Bishop, S1 e S2 são as duas fitas representando o *backbone* do DNA com  $N$  pares de base, as linhas tracejadas verticais representam as ligações de hidrogênio. Fonte: o próprio autor.

cálculo da função partição. Portanto, transformaremos as variáveis  $u_n$   $v_n$  em  $x_n$   $y_n$ , equação (10), para desacoplarmos estas variáveis.

$$x_n = (u_n + v_n)/\sqrt{2} \text{ e } y_n = (u_n - v_n)/\sqrt{2} \quad (10)$$

$x_n$  e  $y_n$  representam, respectivamente, os movimentos em fase e fora de fase das fitas. Agora a distância entre as fitas passa a ser determinado apenas por  $y_n$ . Através da transformação canônica da equação (10) reescrevemos o Hamiltoniano da equação (9)

$$H = \underbrace{\sum_n \frac{p_{x_n}^2}{2m} + \frac{1}{2}k(x_n - x_{n-1})^2}_{H(x)} + \underbrace{\sum_n \frac{p_{y_n}^2}{2m} + \frac{1}{2}k(y_n - y_{n-1})^2 + D(e^{-\frac{y_n}{\lambda}} - 1)^2}_{H(y)} \quad (11)$$

A equação (11) pode ser dividida em dois termos:  $H(x)$  referente à variável  $x_n$  e  $H(y)$  referente à variável  $y_n$ .  $p_{x_n}$  e  $p_{y_n}$  são os momentos lineares das novas variáveis  $x_n$  e  $y_n$ . Já a constante  $\lambda = 1/a\sqrt{2}$ .

O sistema é analisado no contexto de equilíbrio (número de fitas duplas = número de fitas simples). Nesta condição, a temperatura do sistema é fixa na temperatura de desnaturação,  $T_m$ , existindo apenas troca de energia no sistema. A mudança de energia significa a troca entre o estado de fita simples para o estado de fita dupla, ou vice-versa. Isso caracteriza o ensemble canônico. Para uma molécula composta por  $N$  pares de base, como mostrados na figura 13, a função partição  $Z$  é

$$Z = \int_{-\infty}^{+\infty} \prod_{n=1}^N dp_{x_n} dp_{y_n} dx_n dy_n e^{-\beta H(p_{x_n}, p_{y_n}, x_n, y_n)} = \left(\frac{8\pi^3 m^2}{k\beta^3}\right)^{N/2} Z_y, \quad (12)$$

onde  $\beta = 1/k_b T$ , sendo  $k_b$  a constante de Boltzmann e  $T$  a temperatura. A integração nos momentos,  $p_{x_n}$  e  $p_{y_n}$ , e na coordenada  $x_n$  são de forma gaussiana, cujo valor é  $(2\pi m/\beta)^{N/2}$  para os momentos e  $(2\pi/k\beta)^{N/2}$  para  $x_n$ . No entanto, a integração na coordenada  $y_n$  não é de maneira direta, por conta do potencial de Morse, equação (13). Para resolver esta integral, utilizamos o método de integral de transferência (IT). Mas antes de utilizarmos IT, reescreveremos o integrando  $Z_y$  em um formato mais prático para usufruir deste método. Primeiramente, substituiremos o integrando  $Z_y$  por  $K(y_n, y_{n-1})$ .

$$Z_y = \int_{-\infty}^{+\infty} \prod_{n=1}^N dy_n e^{-\beta[\frac{1}{2}k(y_n - y_{n-1})^2 + D(e^{-\frac{y_n}{\lambda}} - 1)^2]} = \int_{-\infty}^{+\infty} \prod_{n=1}^N dy_n K(y_n, y_{n-1}) \quad (13)$$

A função  $K(y_n, y_{n-1})$  da equação (13) é chamado de *kernel* ou núcleo da função.  $K(y_n, y_{n-1}) \neq$

$K(y_{n-1}, y_n)$ , porém se escrevermos o potencial de Morse  $V(y_n)$  como uma média

$$V(y_l, y_{l-1}) = \left[ \frac{D}{2}(e^{-\frac{y_l}{\lambda}} - 1)^2 + \frac{D}{2}(e^{-\frac{y_{l-1}}{\lambda}} - 1)^2 \right]$$

Quando  $l = n$ , o *kernel*  $K(y_n, y_{n-1})$  se torna simétrico, pois se invertermos  $y_n$  com  $y_{n-1}$  o integrando não se altera. Isso nos permite escrever a função partição  $Z_y$  da seguinte maneira

$$Z_y = \int_{-\infty}^{+\infty} dy_1 dy_2 dy_3 \dots dy_{N-1} dy_0 K(y_1, y_0) K(y_2, y_1) K(y_3, y_2) \dots K(y_0, y_{N-1}) \quad (14)$$

No limite de  $n$  grande impomos condições periódicas nesta função de forma que o último par de base interaja com o primeiro, equação (14). Agora estamos em condições de usar o método IT que é definido da seguinte forma

$$\int_{-\infty}^{+\infty} K(x, y) \phi(y) dy = \eta \phi(x), \quad (15)$$

Onde  $K(x, y)$  deve respeitar as seguintes condições:

1. Simétrica

$$K(x, y) = K(y, x)$$

2. Seja quadrado integrável

$$\|K(x, y)\|^2 = \int \int [K(x, y)]^2 dx dy < \infty$$

3. Pode ser expandida como

$$K(x, y) = \sum_n \eta_n \phi_n(x) \phi_n(y)$$

Contanto que as autofunções sejam ortonormais ( $\int \phi_n(x) \phi_m(x) dx = \delta_{nm}$ ) e formem um conjunto completo ( $\sum_{n=1}^{\infty} \phi_n(x) \phi_n(y) = \delta(x - y)$ ).

Comparando as equações (14) e (15) em conjunto com as propriedades das autofunções  $\phi_n$ , finalmente encontramos a função de partição  $Z_y = \sum_{n=1}^{\infty} \eta_n^N$ . O Kernel  $K(y_n, y_{n-1})$  da equação (13) foi definido como  $e^{-\beta E(y_n, y_{n-1})}$ , onde  $E(y_n, y_{n-1})$  é soma do potencial de Morse  $V(y_n)$  com o potencial harmônico  $w(y_n, y_{n-1})$ . Note que a função  $K(y_n, y_{n-1})$  pode ser dividida em um produto de exponenciais

$$K(y_n, y_{n-1}) = e^{-\beta} e^{E(y_n, y_{n-1})}$$

Comparando esta última equação com a equação (15) percebemos que o autovalor  $\eta_n$  na verdade é

$\eta_n = e^{-\beta\epsilon_n}$ . No qual  $\epsilon_n$  é a energia no estado  $n$  do sistema.

$$Z_y = \sum_{n=0}^{\infty} e^{-\beta\epsilon_n N} \quad (16)$$

Os outros termos da função partição são apenas constantes. Logo, toda a termodinâmica do sistema se restringe a resolvermos a função partição  $Z$  na coordenada  $y$ . Através da função partição  $Z$ , encontramos o estiramento médio das fitas  $\langle y_l \rangle$ . O sub-índice  $l$  indica a posição do par de base ou sítio no nucleotídeo. O cálculo de  $\langle y_l \rangle$  é feito de maneira análoga ao da função partição, equações (13)–(16). Este valor é extremamente importante, pois fornece informações de como o duplexo se abre e se transforma em fita simples.

$$\langle y_l \rangle = \frac{1}{Z} \int_{-\infty}^{+\infty} \prod_{n=1}^N dy_n y_l K(y_n, y_{n-1}) = \frac{1}{Z} \sum_{n=0}^{\infty} \langle \phi_n(y) | y_l | \phi_n(y) \rangle e^{-\beta\epsilon_n N}, \quad (17)$$

onde

$$\langle \phi_n(y) | y_l | \phi_n(y) \rangle = \int \phi_n(y) y_l \phi_n(y) dy_n \quad (18)$$

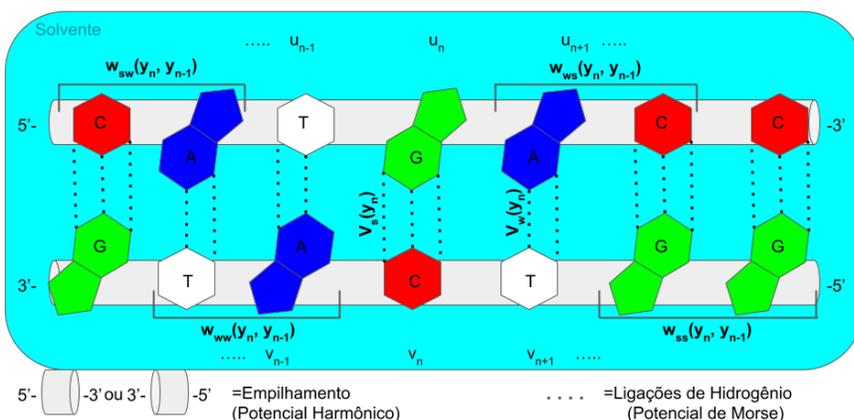
A fita é homogênea, com isso  $\langle y_l \rangle$  independe da posição do sítio e neste caso podemos fazer  $\langle y_l \rangle = \langle y \rangle$ . No limite termodinâmico ( $N \rightarrow \infty$ ), o autovalor de menor energia  $\epsilon_0$  domina a somatória. Assim, as equações (16) e (17) podem ser simplificadas como

$$Z_y = \sum_{n=0}^{\infty} e^{-\beta\epsilon_n N} \quad \text{e} \quad \langle y \rangle = \langle \phi_0(y) | y | \phi_0(y) \rangle \quad (19)$$

As autofunções  $\phi_n$  e os autovalores  $\epsilon_n$  são calculados numericamente. O modelo PB foi um marco histórico na construção de modelos teóricos para análise da estrutura do DNA. O modelo PB passou por várias modificações com passar dos anos [8, 24, 28, 79] para conseguir reproduzir os resultados experimentais da melhor forma possível. Na próxima parte discutiremos a aplicação do modelo PB para sequências heterogêneas.

### 3.1.1 Modelo PB para sequências heterogêneas

O modelo PB descrito na seção anterior vale somente para sequências homogêneas de DNA, tais como sequências compostas somente por pares CG ou somente por pares AT. No entanto, qualquer sequência de DNA de interesse biológico ou tecnológico não é homogênea. Zhang et al. [79] demonstrou que o modelo PB pode ser facilmente estendido para abranger sequências heterogêneas. Os autores desenvolveram o método abordagem de matriz de transferência estendida (AMTE) com o auxílio do método IT. Os detalhes deste método serão explicados em sequência.

**Figura 14**

Representação 2D do empilhamento e ligações de hidrogênio em DNA. As funções  $w$  mostradas representam o potencial de empilhamento e  $V$  as ligações de hidrogênio. Fonte: o próprio autor.

O Kernel  $K$  da equação (13) é a soma do potencial de Morse com o potencial harmônico. No DNA canônico existem dois tipos de potenciais de Morse  $V(y_n)$ , AT e CG, que indexaremos como  $V_s(y_n)$  para os pares de base CG e  $V_w(y_n)$  para os pares AT. O sub-índice  $s$  significa forte (do inglês *strong*), enquanto que o  $w$  significa fraco (do inglês *weak*) [79], pois CG faz três ligações de hidrogênio enquanto que AT faz apenas duas ligações (figura 14). Como teste, os autores assumiram valores diferentes para a constante  $D$  do potencial de Morse, onde CG (42 meV) > AT (38 meV).

Chamaremos o potencial harmônico da equação (13) de  $w(y_n, y_{n-1})$  para evitarmos de escrever esta equação. Por simplicidade desprezaremos em primeiro momento a ordem com que as bases se empilham na estrutura e assumiremos que a interação de empilhamento entre os pares de base seja  $ss$ ,  $sw$ ,  $ws$  ou  $ww$  de valor  $k$  (4.2 eV/nm<sup>2</sup>). Empilhamentos  $ss$  formam dímeros com todas as combinações possíveis de pares CG ou GC, enquanto os do tipo  $ww$  formam dímeros com todas as combinações possíveis de pares AT ou TA. Para o caso  $sw$  o primeiro par de base do dímero é do tipo CG ou GC e o segundo é um par de base AT ou TA, ao todo temos quatro possibilidades. O caso  $ws$  é o inverso do caso  $sw$ . Portanto, teremos quatro tipos de interações de empilhamento na estrutura chamados de  $w_{ss}(y_n, y_{n-1})$ ,  $w_{sw}(y_n, y_{n-1})$ ,  $w_{ws}(y_n, y_{n-1})$  e  $w_{ww}(y_n, y_{n-1})$ . Um exemplo para cada tipo de empilhamento na estrutura do DNA é mostrado na figura 14.

Anteriormente tínhamos apenas um *kernel*  $K$ , com a nova notação temos agora quatro *kernels*  $K$ 's distintos na equação (13)  $K_{ss}(y_n, y_{n-1})$ ,  $K_{sw}(y_n, y_{n-1})$ ,  $K_{ws}(y_n, y_{n-1})$  e  $K_{ww}(y_n, y_{n-1})$ . Por causa da imposição de empilhamento  $k$  igual entre todos os pares de base  $K_{sw}(y_n, y_{n-1}) = K_{ws}(y_n, y_{n-1})$ .

Para sequências de DNA pequenas, até 3 pares de base, os autores demonstraram que é possível calcular tanto  $Z_y$  quanto  $\langle y_n \rangle$  com os quatro, ou três, diferentes *kernels* utilizando o método PB descrito na seção anterior. Além disso, os resultados encontrados concordam com os dados experimentais de desnaturação em DNA. Por exemplo, sequências de DNA com mais pares de base

CG na estrutura possuem temperatura de desnaturação mais altas [80]. No entanto, para sequências maiores utilizar, o método PB gasta muito tempo de cálculo, isso significa que o custo computacional é muito grande.

Engenhosamente, Zhang et al. [79] utilizaram o fato de que o *kernel*  $K$  pode ser expandido em qualquer conjunto de bases ortonormais desde que a expansão respeite a condição 2 da IT, equação (15). Em outras palavras, os *kernels*  $K$ 's de cada par de base do modelo PB, equação (14), passam a ser produtos de matrizes  $K^{(i)}(y_n, y_{n-1})$ . O sobrescrito  $i$  indica a posição entre pares de base vizinhos, cujo valor se encontra entre  $1 < i < N$  ( $N =$  total de pares de base). Na sequência de DNA da figura 14 temos CG-AT ( $i = 1$ ); AT-TA ( $i = 2$ ) ...; CG-CG ( $i = N - 1$ ); CG-CG ( $i = N$ ) - neste caso  $N = 7$ . Esse método foi chamado de abordagem de matriz de transferência estendida, AMTE. Logo, a expansão da função partição  $Z_y$  da equação (14) se torna

$$Z_y = \int_{-\infty}^{+\infty} dy_1 dy_2 dy_3 \dots dy_{n-1} dy_0 K^{(1)}(y_1, y_0) K^{(2)}(y_2, y_1) K^{(3)}(y_3, y_2) \dots K^{(N)}(y_0, y_{n-1}), \quad (20)$$

onde o número acima de  $K$  distingue os empilhamentos  $w(y_n, y_{n-1})$  entre as bases vizinhas. Novamente fazemos uso da condição de contorno periódica (último par de base interage com o primeiro,  $N = 0$ ). O número total de configurações possíveis entre próximos vizinhos no DNA canônico são 16, porém 6 destas combinações são equivalentes restando apenas 10 (AT-AT; AT-CG; AT-GC; AT-TA; CG-AT; CG-CG; CG-GC; GC-AT; GC-CG; TA-AT), ou seja, 10 tipos de *kernels*  $K$ 's.

Antes de resolvermos a função partição da equação (20) pelo método AMTE, é importante ressaltar que a condição 2 do método TI, equação (15), não é respeitada no modelo PB, pois a técnica IT diz que todos os potenciais tem de ser não-ligados para  $K(x, y)$  ser quadrado integrável. No entanto, o potencial de Morse do modelo PB é ligado e conseqüentemente a condição 2 do método TI, equação (15), não é satisfeita. Uma das soluções propostas pelos autores foi reduzir o espaço de integração de  $[-\infty, \infty]$  para um espaço  $[a, b]$  aonde a função exista. Ou seja, discretizar o problema, e posteriormente expandir  $a \rightarrow -\infty$  e  $b \rightarrow \infty$ .

A solução proposta por Zhang et al. [79] resolve parcialmente o problema, pois ao aumentarmos o limite  $b$  a temperatura de desnaturação aumenta. Contudo nosso grupo resolveu o problema adicionando um pequeno ângulo de torção na fita ( $\theta = 0.01$  rad) [81]. Logo, o potencial harmônico  $w(y_n, y_{n-1})$ , equação (13), passa a ser escrito como

$$w(y_n, y_{n-1}) = \frac{1}{2}k(y_n - y_{n-1})^2 = \frac{1}{2}k[y_n^2 + y_{n-1}^2 - 2(y_n)(y_{n-1})\cos\theta].$$

Agora podemos afirmar que a função  $K^{(i)}(y_n, y_{n-1})$  é definida no espaço  $[a, b; a, b]$ , e sua norma exista. Desse modo podemos escolher qualquer conjunto de bases ortonormais  $\phi_n(y)$  para expandir

$K^{(i)}(y_n, y_{n-1})$

$$K^{(i)}(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} C_{nm}^{(i)} \phi_n(x) \phi_m(y), \quad (21)$$

onde

$$C_{nm}^{(i)} = \int \int dx dy K^{(i)}(x, y) \phi_n^*(x) \phi_m^*(y). \quad (22)$$

Na prática a equação (21) é truncada nas  $M$  primeiras bases. Apesar de sacrificarmos a precisão dos resultados, ganhamos velocidade no cálculo computacional. Neste trabalho o *kernel*  $K$  da função de partição, equação (15), foi calculado entre  $y = -0.1$  e  $y = 20$  nm em uma matriz de tamanho 400. Desta matriz pegamos apenas os 10 maiores autovalores ( $M = 10$ ), pois mais termos não geram uma diferença tão significativa nos resultados. Substituindo a equação (21) na equação (20), e utilizando a completeza e ortonormalidade definidas na seção anterior (condição 3 da equação (15)) encontramos

$$Z_y = \sum_{n=0}^M C_{nn}^{(1)} C_{nn}^{(2)} \dots C_{nn}^{(N)} = Tr \left( \prod_{i=1}^N C^{(i)} \right) \quad (23)$$

A ortonormalidade impõe  $m = n$ , ou seja, a soma é feita apenas nos elementos diagonais das matrizes de cada *kernel*  $K^{(i)}$ . Com a função partição  $Z$  calculamos o estiramento médio do sistema. Primeiramente definimos a matriz  $Y$  com elementos

$$Y_{nm} = \langle \phi_n(y) | y | \phi_m(y) \rangle = \int dy \phi_n(y) y \phi_m(y)$$

Logo, o estiramento médio no sítio  $l$  será dado por

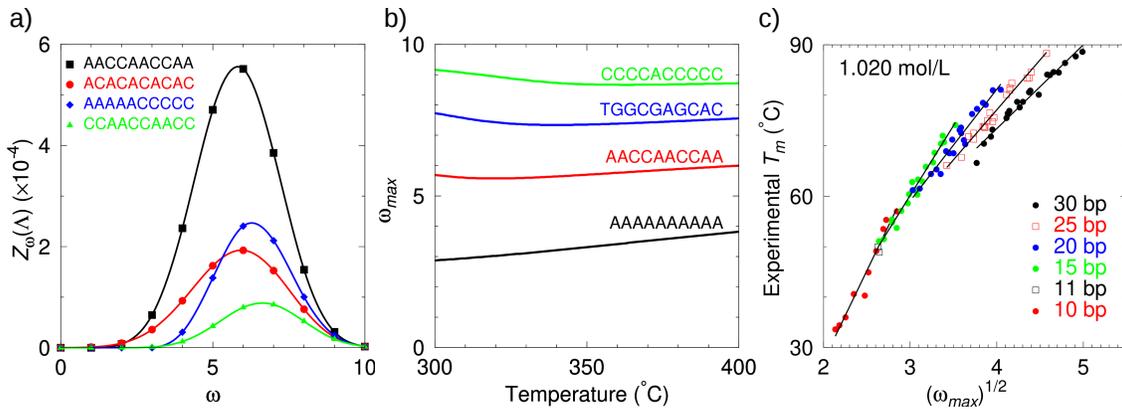
$$\langle y_l \rangle = \frac{1}{Z} Tr [C^{(1)} \dots C^{(i-1)} Y C^{(i)} \dots C^{(0)}] \quad (24)$$

As interações de empilhamento dos próximos vizinhos são diferentes, além claro dos potenciais de Morse. Portanto, cada sítio  $l$  terá um valor de estiramento médio diferente. Para ficar mais claro o argumento anterior resolveremos a equação (24) no caso  $l = 1$  (segundo par de base). Neste caso a matriz  $Y$  ocupará a posição entre as matrizes  $C^{(1)}$  e  $C^{(2)}$

$$\langle y_1 \rangle = \frac{1}{Z} Tr [C^{(1)} Y C^{(2)} C^{(3)} \dots C^{(N-1)} C^{(0)}],$$

### 3.1.2 Equivalência térmica

O método usado para tratar o modelo PB em sequências heterogêneas pode ser utilizado para determinar um índice de equivalência, que pode ser correlacionado com as temperaturas de desnaturação medidas experimentalmente [8, 28]. Weber et al. [8] utilizou das propriedades do traço



**Figura 15**

a) comportamento gaussiano entre  $Z_\omega(\Lambda)$  e  $\omega$  para 4 seqüências de 10 pares de base contendo 40% à 60% de pares de base C-G na temperatura de 370 K; b) Comportamento de  $\omega_{max}$  em função da temperatura para outras 4 seqüências com 10 pares de base; c) Temperaturas de *melting* experimentais ( $T_m$ ) em função dos valores de  $\tau = (\omega_{max})^{1/2}$  para diversas seqüências. Figuras retiradas de Weber et al. [8].

para separar as matrizes  $C_{nm}^{(i)}$  da equação (23) em uma soma de duas matrizes

$$C^{(i)} = \Lambda + \Delta^{(i)}, \quad (25)$$

onde  $i$  indica a posição entre pares de base vizinhos no duplexo. A matriz  $\Lambda$  é composta apenas por elementos diagonais  $\eta_n$ ; onde  $\eta_n$  são os autovalores,  $e^{-\beta\epsilon_n}$  - equação (16), calculados para seqüências homogêneas CG. Já a matriz  $\Delta^{(i)}$  é a diferença entre qualquer outro próximo vizinho e o vizinho CG-CG ( $\Delta^{(i)} = C^{(i)} - \Lambda$ ). Em particular,  $\Delta^{(i)}$  para CG-CG é zero [8]. Com isso, a equação (23) é reescrita como

$$\begin{aligned} Z_y &= Tr[(\Lambda + \Delta^{(1)})(\Lambda + \Delta^{(2)})\dots(\Lambda + \Delta^{(N-1)})(\Lambda + \Delta^{(N)})] \\ &= \sum_{\omega=0}^N Z_\omega(\Lambda) = \sum_{\omega=0}^N Tr[M(\Lambda^\omega)], \end{aligned} \quad (26)$$

onde  $M(\Lambda^\omega)$  são termos contendo  $\omega$  multiplicações da matriz  $\Lambda$ . Se a cadeia for puramente CG  $\omega = N$ , caso  $\omega < N$  existirá pontos no duplexo diferente do par CG. Portanto, o grau  $\omega$  da matriz  $\Lambda$  da equação (26) representa uma medida de ordenamento na cadeia de nucleotídeos. A figura 15a apresenta os valores calculados de  $Z_\omega(\Lambda)$  em função de  $\omega$  para quatro tipos de duplexos com proporções de CG entre 40% e 60% na estrutura.

O perfil da curva na figura 15a é claramente gaussiano. Além disso, a forma e o pico máximo da curva ( $\omega_{max}$ ) mudam de acordo com a posição dos pares de base presentes no duplexo. Isso demonstra que  $Z_\omega(\Lambda)$  seria perfeito para comparar seqüências. No entanto, a função de partição possui uma forte dependência com a temperatura impossibilitando seu uso desta maneira. Porém,

$\omega_{max}$  apresenta pouca variação com a temperatura (figura 15b), sendo um parâmetro interessante para diferenciar as sequências de nucleotídeos.

Sequências com mesmo tamanho, mas com pares de base em posições distintas apresentam temperaturas de desnaturação próximas. A figura 15c mostra a dependência da temperatura de desnaturação experimental ( $T_m$ ) em função de  $\omega_{max}$  calculado para cada sequência. É perfeitamente claro pela figura 15c que sequências com  $\omega_{max}$  próximos apresentam temperaturas de desnaturação similares. Ademais esta dependência é aproximadamente linear com  $(\omega_{max})^{1/2}$ . Com base nestas observações foi definido  $\omega_{max}$  como índice de equivalência térmica e  $\tau = (\omega_{max})^{1/2}$  como índice de *melting*.

O índice de *melting* é usado para prever as temperaturas de desnaturação. Portanto, para cada sequência  $i$  experimental, calculamos um valor  $\tau_i(P)$ , onde  $P$  é o conjunto de parâmetros de Morse,  $D$ , e empilhamento,  $k$ . Deste  $\tau_i(P)$  encontramos um valor de temperatura  $T'_i(P)$  através dos coeficientes da regressão linear do conjunto de pontos  $(\tau_i(P), T_i)$ , onde  $T_i$  são as temperaturas experimentais. Este conjunto de pontos é dividido em cadeias de mesmo tamanho, como mostrado na figura 15c. Neste trabalho utilizamos dois métodos de predição de temperaturas ( $T'_i(P)$ ):

#### 1. Regressão simples

$$T'_i(P) = a_0 + a_1\tau_i(P) \quad (27)$$

#### 2. Regressão que leva em conta o tamanho da sequência

$$T'_i(P) = a_0(N) + a_1(N)\tau_i(P), \quad (28)$$

onde

$$a_k = b_{0,k} + b_{1,k}N^{1/2} \quad (29)$$

Os coeficientes  $a_k$  das equações (27) e (28), onde  $k = 0$  ou  $1$ , são os coeficientes linear ( $a_0$ ) e angular da reta ( $a_1$ ).  $P$  é o conjunto de parâmetros de Morse e empilhamento utilizados para calcular a função partição [equações (16), (23) e (26)] da sequência  $i$ . Para o DNA o número máximo de parâmetros utilizados seriam 12 (2 parâmetros de Morse e 10 parâmetros de empilhamento). Por fim,  $N$  na equação (28) é o tamanho do nucleotídeo.

### 3.1.3 Otimização dos parâmetros

O processo de otimização consiste em minimizar a função mérito

$$\chi_j^2 = \sum_{i=1}^N [T'_i(P_j) - T_i]^2, \quad (30)$$

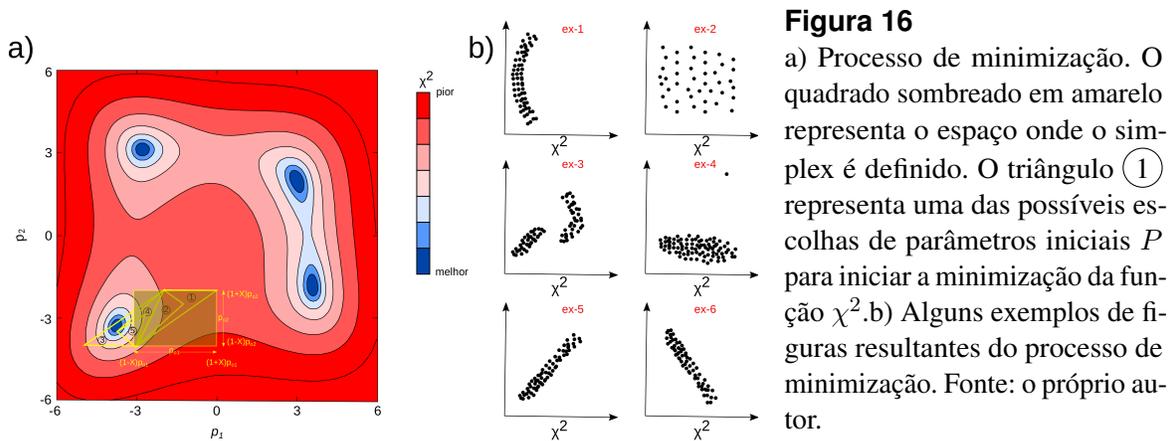
onde  $T'_i(P_j)$  é a temperatura de desnaturação calculada pela equação (27), ou (28),  $T_i$  é a temperatura de desnaturação experimental e  $N$  é o número total de temperaturas de desnaturação na base de dados. A procura pelo mínimo é feita através do método Nelder-Mead (*downhill simplex*) [82], explicado abaixo, e este procedimento é realizado  $j$  vezes. Para cada  $j$ , um novo conjunto de parâmetros de potenciais de Morse,  $D$ , e empilhamentos,  $k$ , são utilizados, e conseqüentemente um novo  $\chi^2$  é encontrado.

A escolha dos novos parâmetros é feita de maneira aleatória com base nos valores iniciais inseridos no programa. Para isso estabelecemos um intervalo de variação. Por exemplo, 10% do valor inicial  $P$ , onde  $P$  é o conjunto de parâmetros de Morse,  $D$ , e empilhamento,  $k$ . Ou seja, o processo de minimização pode se iniciar por qualquer valor contido no intervalo  $[0.9P, 1.1P]$ .

O algoritmo de minimização inicia-se por um *simplex* (triângulo  $N$ -dimensional). Sua forma varia de acordo com a dimensão analisada. Em uma dimensão o *simplex* é uma reta, em duas dimensões é um triângulo e em três dimensões é um tetraedro [83]. No caso geral o *simplex* possui uma dimensão a mais do que o espaço de parâmetros analisado. Em nosso caso particular, o *simplex* é um triângulo  $P$ -dimensional.

Na figura 16a apresentamos como o processo de minimização funciona na prática para um espaço de parâmetros bi-dimensional. A região entre -6 e 6 representa uma fração do domínio do espaço dos parâmetros  $p_1$  e  $p_2$  que possui a maior probabilidade de encontrar um mínimo global. Este domínio é característico de cada conjunto de dados experimentais, com isso a figura 16a representa apenas um dentre estes inúmeros casos. As regiões mais azuis da figura 16a são posições de mínimos locais. Apenas em uma destas regiões haverá um ponto azul-escuro representando o mínimo global. Há quatro possíveis regiões na figura 16a com possibilidades de se encontrar um mínimo global. Com base em referências experimentais e/ou teóricas selecionamos os parâmetros iniciais  $p_{o1}$  e  $p_{o2}$ . Em seguida definimos o incremento extra  $X$ , aonde o *simplex* dará início ao processo de minimização, quadrado em sombreado amarelo mostrado na figura 16a. O *simplex*, neste caso um conjunto de três pontos, é escolhido de forma aleatória dentro do quadrado sombreado. Na figura 16a mostramos uma destas possíveis escolhas, triângulo ①.

Os três vértices do triângulo dão origem a valores de  $\chi^2$  diferentes. O método *downhill simplex* compara estes três valores e descarta o pior deles. O vértice descartado sofre uma reflexão



dentro do domínio procurando um valor menor de  $\chi^2$  de forma a reduzir o tamanho do triângulo, triângulos ②-⑤ mostrados na figura 16a. Este procedimento acontece várias vezes até o triângulo se tornar tão pequeno dentro da região azul-escuro (figura 16a), posições de mínimo da função mérito, que o programa encerra o processo. Ao final do processo obtemos um valor de mínimo local ou, por sorte, global. Por isso, para contornar o problema de mínimos locais o procedimento acima é realizado  $j$  vezes.

Em cada minimização  $j$  geramos um ponto, onde a coordenada  $x$  é o mínimo de  $\chi^2$  encontrado e a coordenada  $y$  são os parâmetros de empilhamento,  $k$ , e Morse,  $D$ . Ao final da  $j$ -ésima minimização teremos  $j$  pontos. A partir destes  $j$  pontos fazemos um *plot* dos valores de  $\chi^2$  em função do parâmetro. Utilizamos estas figuras para identificar a presença de *outliers*. Pontos de dados que se diferenciam significativamente dos demais em um conjunto de dados. Eles são observações que estão muito distantes da maioria dos outros pontos. Na figura 16b mostramos alguns exemplos de figuras resultantes do processo de minimização. No gráfico ex-4 apresentamos um exemplo contendo um *outlier*.

Os parâmetros finais mostrados nas seções 4 e 5 são do menor  $\chi^2$  encontrados pelas  $j$  minimizações. Além do mais, avaliamos a influência do erro experimental nas nossas análises. Para isto, realizamos uma nova minimização variando as temperaturas experimentais, onde o intervalo  $X$  de variação das temperaturas é igual a média das incertezas registradas nos artigos experimentais de UV melting. Ao final avaliamos o quanto o valor do  $\chi^2$  se desviou do valor encontrado anteriormente. A incerteza mostrada nos gráficos e tabelas de cada parâmetro das seções 4 e 5 são os desvios padrões sobre todas as minimizações  $j$  referentes à influência do erro experimental em cada parâmetro. Adicionalmente, nós utilizamos também outras funções para determinar a qualidade das temperaturas preditas que é a diferença média entre as temperaturas experimentais e calculadas

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T'_i - T_i| \quad (31)$$

e o desvio quadrático médio

$$\Delta T_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N [T'_i - T_i]^2} = \sqrt{\frac{\chi^2}{N}} = \sqrt{\langle \chi^2 \rangle} \quad (32)$$

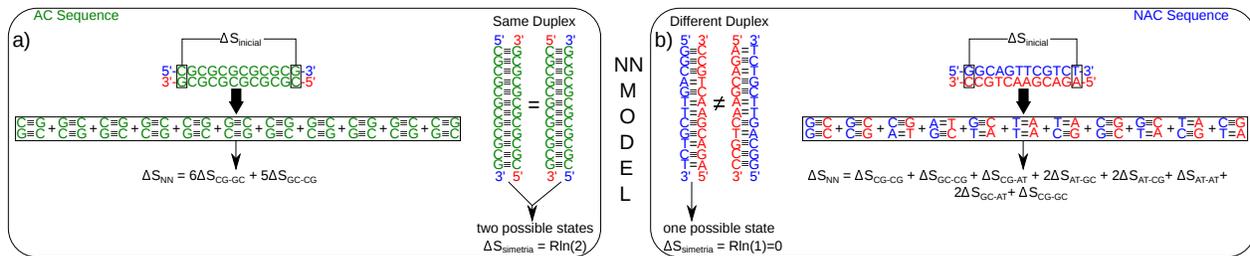
Na seção seguinte apresentamos o modelo de próximos vizinhos. Este modelo calcula temperaturas de desnaturação a partir das grandezas termodinâmicas de entalpia,  $\Delta H$ , entropia,  $\Delta S$ , e energia livre de Gibbs,  $\Delta G$ , totais dos duplexos.

## 3.2 Modelo de próximos vizinhos

O modelo de próximos vizinhos, do inglês *nearest neighbors* (NN), é baseado no conceito que a estabilidade da dupla fita é determinada por interações entre um par de bases e seu par de bases vizinho. Este modelo foi desenvolvido em 1971 por Tinoco et al. [42] e mais tarde aperfeiçoado por Breslauer et al. [84]. De acordo com o modelo NN, os valores termodinâmicos consistem de três termos. O primeiro termo é a mudança de energia livre para o início de formação do duplexo. As bases terminais estão parcialmente expostas ao solvente. Portanto, a formação do primeiro par de base afeta a estabilidade da dupla hélice. Esta energia livre de iniciação depende do par de base inicial, pois no DNA, por exemplo, CG faz três ligações de hidrogênio, enquanto AT faz apenas duas ligações. O segundo termo é a soma das mudanças de energias livres das interações entre o par de base e seu par de base vizinho. Há 10 tipos de NN canônicos para DNA/DNA ou RNA/RNA e 16 tipos de NN para o híbrido DNA/RNA. A redução do número de próximos vizinhos para hélices duplas de DNA ou RNA foi apresentado na capítulo 2, página 31, e uma representação das 16 possíveis combinações está na figura 4c. O terceiro termo é associado a diferença de simetria entre fitas auto-complementares (AC) e fitas não auto-complementares (NAC). Assim, a mudança de energia livre total ( $\Delta G_{37}$ ) para formação do duplexo em 37 °C é

$$\Delta G_{37(\text{total})} = \Delta G_{37(\text{inicial})} + \sum_i n_i \Delta G_{37,NN(i)} + \Delta G_{37(\text{simetria})} \quad (33)$$

onde,  $\Delta G_{37,NN(i)}$  é a mudança de energia livre padrão para os 10, DNA/DNA ou RNA/RNA, ou 16, DNA/RNA, possíveis NN. Por exemplo, o dímero AT-AT para o DNA seria  $\Delta G_{37,AT-AT}$ .  $n_i$  é o número de repetições de cada NN na sequência,  $\Delta G_{37(\text{inicial})}$  é o fator de iniciação e como dito anteriormente, depende de qual par de base inicia a cadeia de nucleotídeos.  $\Delta G_{37(\text{simetria})} = 0.43$  kcal/mol se o duplexo é AC e 0 se o duplexo é NAC. Todos os valores de  $\Delta G_{37}$  da equação (33) para soluções diluídas ( $\text{H}_2\text{O}$  + tampão), ou  $\Delta S$  e  $\Delta H$ , são tabelados e possuem um pequeno desvio em relação aos dados experimentais [78, 85, 86].

**Figura 17**

Esquema de funcionamento do modelo NN. a) fitas auto-complementares (AC) e b) fitas não auto-complementares (NAC). Fonte: o próprio autor.

O modelo NN é utilizado para calcular parâmetros termodinâmicos ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$  e  $T_m$ ) do duplexo. Sabendo que a concentração total de fitas,  $C_t$ , é diretamente proporcional a  $T_m$ , podemos normalizar  $C_t$  por duas maneiras:

- Usar os valores medidos de  $\Delta G$  ou  $\Delta H$  e  $\Delta S$ , fornecidos pelos artigos, para calcular a temperatura  $T_m$  na concentração  $C_t$  desejada.
- Estimar os valores de  $\Delta G$  ou  $\Delta H$  e  $\Delta S$  através da equação (33), e em seguida calcular a temperatura  $T_m$  na concentração  $C_t$  desejada.

Para isso utilizamos as equações (8) apresentadas na seção 2.1.

Na figura 17, apresentamos dois exemplos de como o modelo NN calcula o  $\Delta S_{total}$  de um duplexo para as fitas AC e NAC. O termo de simetria tem origem entrópica relacionado aos estados possíveis do sistema. Utilizando a relação definida por Boltzmann,  $\Delta S = R\ln\Omega$ , onde  $\Omega$  é o número de estados do sistema, podemos obter as variações de entropia  $\Delta S_{simetria}$  relativas à simetria do sistema. Para sequências AC,  $\Delta S_{simetria}$  é diferente de zero, pois a sequência codante é idêntica a complementar. Na figura 17, indexamos a sequência codante de azul e a sequência complementar de vermelho para mostrar de onde vem este valor de entropia. No caso AC, figura 17a, não importa qual fita é a codante e qual é a complementar, pois obtemos o mesmo duplexo e observamos dois estados possíveis para o sistema. Para NAC, figura 17b, a fita que definimos como codante importa, pois os duplexos são diferentes e, neste caso, existe apenas um estado.  $\Delta S_{inicial}$  é calculado para cada par de base CG ou AT presente no nucleotídeo, já  $\Delta S_{NN}$  é associado a cada dímero específico da sequência, figura 17. Ao somar todas as contribuições de entropia, chegamos a variação de  $\Delta S_{total}$  do duplexo. A mesma análise é aplicada para o cálculo do parâmetro  $\Delta H_{total}$ . Uma vez encontrados  $\Delta S_{total}$  e  $\Delta H_{total}$ , calculamos as temperaturas de desnaturação  $T_m$  por meio das equações (8).

A otimização dos parâmetros da equação (33) é feita de maneira similar ao apresentado na seção 3.1.3 para o modelo Peyrard and Bishop [43], discutido na seção anterior. No entanto, diferente do modelo PB, no modelo NN, os parâmetros otimizados são os parâmetros termodinâmicos  $\Delta H$

e  $\Delta S$  referente aos próximos vizinhos que compõem o conjunto de sequências estudado. Já as temperaturas preditas são calculadas pelas equações (8) apresentadas na seção 2.1.

Nos capítulos seguintes, apresentaremos os projetos estudados durante o doutorado. O primeiro projeto apresentado foi chamado de DNA mediado por metal, cujo, apresentaremos e discutiremos todas as análises realizadas neste trabalho.

## 4 DNA mediado por metal

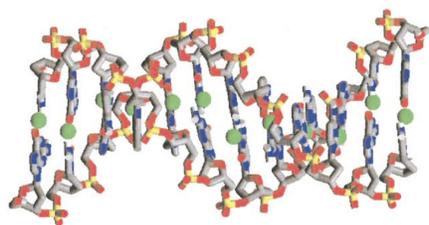
Neste capítulo, apresentaremos a revisão bibliográfica e os resultados do primeiro projeto, intitulado DNA mediado por metal. Na próxima subseção, apresentaremos uma revisão bibliográfica sobre esse tipo de DNA. Nessa parte incluímos estudos experimentais e teóricos. Após isto, apresentamos nossos resultados, conclusões e perspectivas futuras. No apêndice A, tabelas 21 e 22, estão as seqüências encontradas na literatura e utilizadas neste projeto.

### 4.1 Revisão bibliográfica

DNA mediado por metal (M-DNA) é um tipo de DNA na qual a ligação de hidrogênio é substituída por um íon metálico entre as bases nitrogenadas (figura 18). Além de melhorar a estabilidade térmica do duplexo, há uma melhora nas propriedades eletromagnéticas em relação ao DNA Watson e Crick tipo B (DNA WC tipo B), discutido na seção 2.

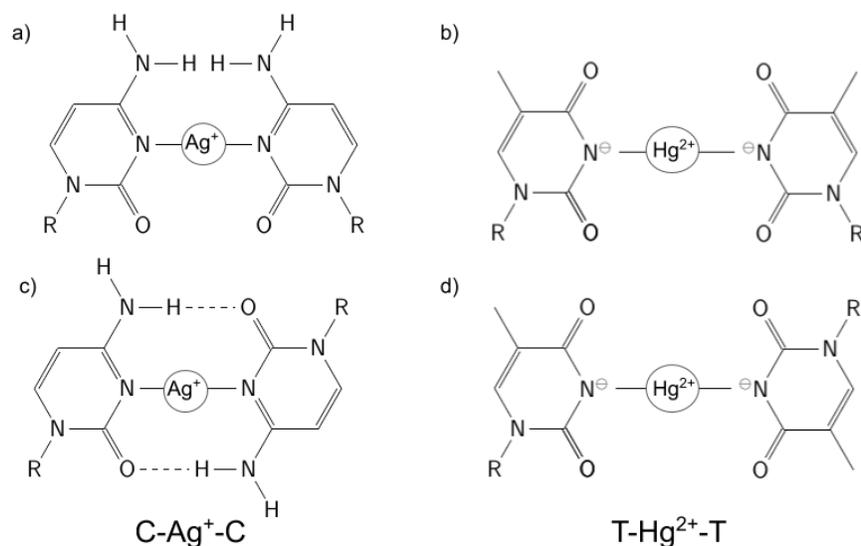
Após a definição da estrutura de dupla hélice do DNA, várias dúvidas surgiram sobre sua estabilidade. Dentre elas, como outros metais, além do sódio ( $\text{Na}^+$ ), afetam a estrutura do DNA, uma vez que a presença de sódio na célula é essencial para formação do duplexo de DNA. Assim, entre a década de 50 e 90, diversos estudos foram realizados de UV *melting* do DNA em solução com outros metais sob diferentes concentrações [36, 39, 87–90] tanto que, em 1962, Eichhorn [36] descobriu que alguns metais estabilizam e outros destabilizam o duplexo de DNA. A relação entre estabilidade e instabilidade se deve à concentração do metal na solução, e a partir de um certo limite, particular para cada íon metálico, a estabilidade do duplexo fica comprometida.

Experimentos com mercúrio ( $\text{Hg}^{2+}$ ) em soluções com alta concentração de AT e com prata ( $\text{Ag}^+$ ) em soluções com alta concentração de CG sugeriram que ambos se ligam entre as bases nitrogenadas [87, 88]. No entanto, isso era apenas uma proposta baseada nos experimentos de UV *melting*, e eles não tinham exatamente uma ideia de aonde o metal se encaixava no DNA. Até que



**Figura 18**

Representação teórica de uma fita de DNA com metais substituindo às ligações de hidrogênio. As esferas em verde representam os íons metálicos que substituem as ligações de hidrogênio por ligações covalentes mediadas por metal (figura retirada de Aich et al. [9]).



**Figura 19**

Configurações possíveis de  $\text{Ag}^+$  e  $\text{Hg}^{2+}$  que se ligam covalentemente as bases nitrogenadas do DNA. As duas figuras acima são do caso WC e as duas figuras abaixo são do caso WC reverso. R significa desoxirribose. Fonte: o próprio autor.

em 1963, para o  $\text{Hg}^{2+}$ , Sidney propôs que o nitrogênio N3 da timina fosse desprotonado, formando à ligação covalente T-Hg(II)-T [91]. Esta dúvida foi finalmente resolvida em 1985 por E. et al. [92], através de experimentos de ressonância magnética nuclear (NMR), onde descobriram que Hg(II) realmente substitui à ligação de hidrogênio [40] (figura 19).

Em soluções com pH alto, foi proposto que  $\text{Ni}^{2+}$ ,  $\text{Zn}^{2+}$  e  $\text{Co}^{2+}$  se ligam ao duplexo de DNA, formando M-DNA [9, 63, 93, 94]. Entretanto, estudos posteriores mostraram que isso não acontece [45, 95]. No caso do  $\text{Ag}^+$ , foi confirmado em 2008 por Ono et al. [17] que o  $\text{Ag}^+$  forma a ligação covalente C- $\text{Ag}^+$ -C (figura 19). Semelhante à T- $\text{Hg}^{2+}$ -T, o par C- $\text{Ag}^+$ -C causa pouco ou nenhum impacto na conformação da dupla hélice de DNA [95, 96].

Os pares de base TT e CC mencionados acima são chamados *mismatches*. *Mismatch* (MM) é um defeito do nucleotídeo (DNA ou RNA) que ocorre quando duas bases não complementares se alinham, formando um par de bases no nucleotídeo. MMs podem aparecer durante o processo de duplicação de DNA e através de influência de mutagênicos químicos, físicos ou biológicos [72]. Além do mais, os MMs produzem alterações na estrutura e estabilidade da dupla hélice, especialmente na proximidade do local do MM. No entanto, na presença dos íons metálicos  $\text{Ag}^+$  e  $\text{Hg}^{2+}$  os pares de bases CC e TT são estabilizados, por exemplo.

Em resumo, há metais que preferem:

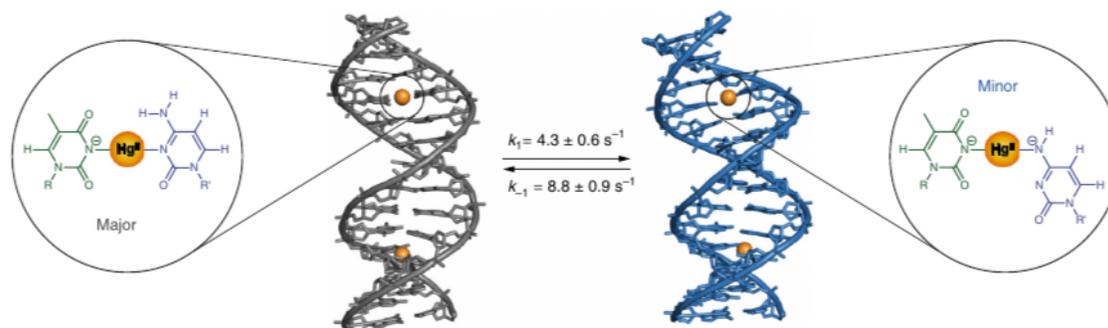
1. se ligar ao grupo fosfato (alcalinos e alcalinos-terrosos);

2. metais que se ligam entre as bases nitrogenadas (Ag(I) e Hg(II)), figura 18);
3. outros que se ligam em ambos os componentes com afinidades na ordem Co(II)~Ni(II) < Mn(II)~Zn(II) < Cd(II) < Pb(II) < Cu(II) [38, 39].

Assim, para as bases naturais (A, T, C e G), apenas  $\text{Hg}^{2+}$  e  $\text{Ag}^+$  se ligam ao DNA formando M-DNA, sem alterar a conformação do DNA WC tipo B [45, 97]. Com isso, a partir dos anos 2000, começou a se sintetizar bases nitrogenadas artificiais com propriedades físico-químicas semelhantes às naturais. Estas bases têm a propriedade de formar estruturas de coordenação tanto com as bases naturais quanto com as artificiais que estabilizam, além de  $\text{Hg}^{2+}$  e  $\text{Ag}^+$ , outros metais na estrutura do DNA [37, 40, 98–100]. Atualmente, os estudos da área se concentram em sintetizar uma molécula de DNA com um metal por par de bases, constituída inteiramente por bases naturais (ou Ag, ou Hg e ou ambos) ou por uma mistura artificial-natural junto com natural-natural [40, 101] (figura 18). Além de novos experimentos de M-DNA UV *melting* com CC e TT [14, 21, 22, 37, 44, 45, 95, 97, 102–113], foram realizados estudos de Hg(II) e Ag(I) com outros MMs [37, 45, 95, 102–104, 108–111], no qual apenas CT teve uma mudança apreciável na temperatura de desnaturação na presença de Hg(II) ou Ag(I). Embora, este aumento de temperatura seja bem menor quando comparado com o MM CC na presença de Ag(I) ou o MM TT na presença Hg(II), este resultado sugere a formação do par C-Ag(I)-T e C-Hg(II)-T WC, onde o metal M (M=Hg(II) ou Ag(I)) ocupa a posição T(N3)-M-(N3)C [37]. No caso específico para o Hg(II), Schmidt et al. [10] identificou esta ligação em um duplexo de DNA com dois MMs CT separados por 6 pares de base, e além disso, observou um segundo tipo de ligação C-Hg(II)-T na qual a hélice muda para a conformação WC tipo A (figura 20 dupla hélice em azul), onde o metal ocupa a posição T(N3)-Hg(II)-(N4)C. A razão entre as duas conformações é de 3:1, predominando a hélice de tipo B. Entretanto, em situações favoráveis à formação da hélice tipo A, como cristalização, desidratação e sequências ricas em pares de base GC, a coordenação T(N3)-Hg(II)-(N4)C passa a ser dominante. Quanto a C-Ag(I)-T, não há registros deste par de base por outras técnicas experimentais.

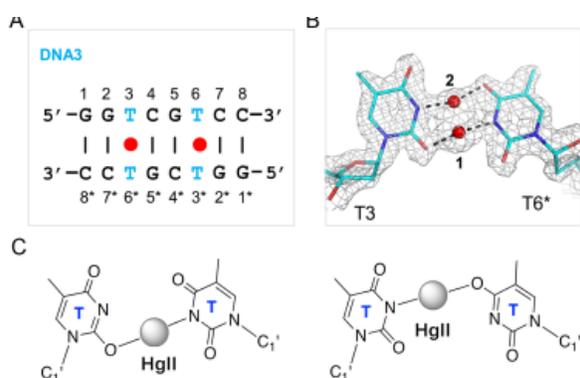
Kondo et al. [12] conseguiram sintetizar em 2017 um nanofio de DNA WC tipo B com um átomo de Ag(I) por par de base (figura 22b). A ligação C-Ag(I)-C é como mostrado na figura 19a. A distância Ag(I)-Ag(I) determinada foi em média de 3.2 Å e a distância entre o nitrogênio N3 da citosina e o Ag(I) foi de 2.3 Å.

Wang et al. [114] propuseram uma possível estrutura para C-Ag<sup>+</sup>-C com uma ligação N3-Ag<sup>+</sup>-O2, onde o nitrogênio N4 da citosina é desprotonado. No entanto, experimentos de NMR realizados por Dairaku et al. [16] encontraram uma estrutura semelhante ao da figura 19a. Experimentos de raio x em cristais de N<sup>1</sup>-hexilcitosina [115] levantaram mais questões em relação ao tipo de estrutura para C-Ag<sup>+</sup>-C. Os autores encontraram uma estrutura parecida com a da figura 19c,



**Figura 20**

Configurações do par C-Hg-T. O duplexo em cinza do lado esquerdo é de conformação tipo B e a ligação com o Hg(II) é T(N3)-Hg(II)-(N3)C; O duplexo em azul do lado direito é de conformação tipo A e a ligação com o Hg(II) é T(N3)-Hg(II)-(N4)C. Figura extraída de Schmidt et al. [10]



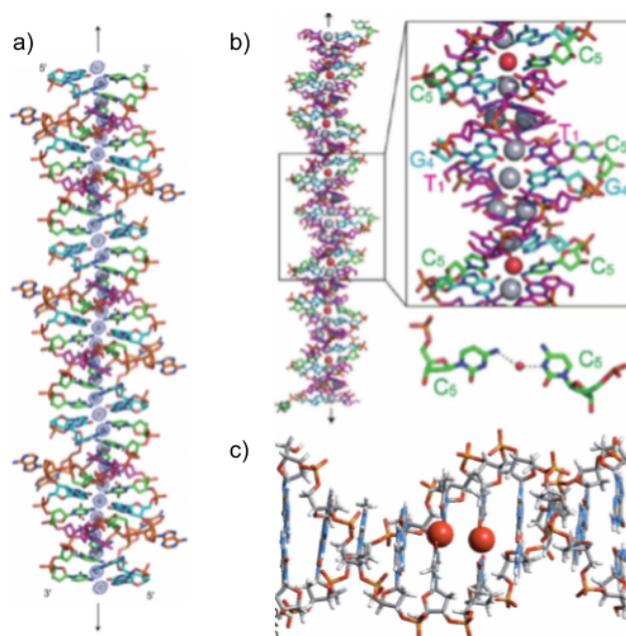
**Figura 21**

Estruturas de  $\text{Hg}^{2+}$  encontradas por Liu et al. [11]. a) Sequência principal e complementar, os pontos vermelhos mostram a localização do íon metálico. b) A conformação detalhada do par T-Hg<sup>2+</sup>-T com os íons metálicos em vermelho. c) Visão esquemática mostrando em detalhes a interação dos *mismatches* T com o íon  $\text{Hg}^{2+}$ . Todas as figuras foram retiradas de Liu et al. [11].

entretanto, devido ao ângulo de torção entre as citosinas apenas a ligação de hidrogênio O—H-N foi visualizada. Além disso, foram observadas interações interplanares entre os íons  $\text{Ag}^+$ . Logo, existem várias dúvidas em relação ao tipo de estrutura formada pelo par C- $\text{Ag}^+$ -C.

Experimentos de NMR e difração de raio X com dois MMs TT na presença Hg(II) (figura 22c) realizados por Yamaguchi et al. [14], confirmaram que a conformação da hélice é WC tipo B, onde a ligação T-Hg(II)-T é como mostrado na figura 19b. A distância Hg(II)-Hg(II) determinada foi de 3.3 Å, e a distância entre o nitrogênio N3 da timina e o Hg(II) foi de 2 Å. Resultados similares foram encontrados por outros estudos de NMR e raio X [19, 21, 22]. No entanto, experimentos de raio X realizados por Liu et al. [11] em um duplexo com dois pares TT separados por dois pares CG encontraram as estruturas O2-Hg<sup>2+</sup>-N3 e N3-Hg<sup>2+</sup>-O4 mostradas na figura 21.

Em 2019, Ono et al. [13] sintetizaram um nanofio de DNA com um átomo de Hg(II) por par de bases (figura 22d). No entanto, o nanofio de Hg(II) não contém nenhum par de base tipo WC, e as citosinas são coordenadas por uma molécula de H<sub>2</sub>O. Ademais, o nanofio de Hg(II) possui quatro átomos de mercúrio que formam um alinhamento tipo espiral que se repete ao longo da estrutura do M-DNA, completamente diferente do alinhamento unidimensional do nanofio de Ag(I) (figura 22b).



**Figura 22**

Estruturas de M-DNA resolvidas experimentalmente. a) nanofio de Ag(I) com um átomo de Ag(I) entre as bases nitrogenadas. Figura extraída de Kondo et al. [12]. b) nanofio de Hg(II) as esferas em cinza são os átomos de Hg(II) e as esferas vermelhas são oxigênios da molécula de H<sub>2</sub>O. Figura extraída Ono et al. [13]. c) Duplexo de DNA com dois átomos de Hg(II) em vermelho entre os *mismatches* TT. Figura extraída de Yamaguchi et al. [14].

Estudos teóricos de DNA mediado por metal se concentram principalmente em duas áreas: Teoria do Funcional Densidade (DFT) [16, 115–125] e Dinâmica Molecular [126, 127]. Nestes estudos, os autores encontraram estruturas para C-Ag<sup>+</sup>-C [16, 115, 117–122, 126, 127] e T-Hg<sup>2+</sup>-T [116, 125] semelhantes aos resultados experimentais apresentados. Outros metais como Cu<sup>2+</sup> [122], Au<sup>+</sup> [118, 122], Zn<sup>2+</sup> [123, 124], Co<sup>2+</sup> [124] e Fe<sup>2+</sup> [124] foram propostos para formar M-DNA com as bases naturais, mas experimentalmente isto não foi observado [95].

Como não há estudos teóricos utilizando modelos mesoscópicos de M-DNA com bases naturais, iremos focar inicialmente nesta parte. Por isso, não discutiremos em detalhes o M-DNA com bases artificiais neste texto. Logo, estudaremos a estabilidade térmica do M-DNA com bases naturais através do modelo estatístico de Peyrard-Bishop [43].

Nas seções seguintes, apresentaremos toda a metodologia empregada no estudo de DNA mediado por metal. Nosso foco foi no caso C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T, que possuem mais sequências disponíveis na literatura. Primeiramente, abordaremos como foi a coleta de dados e os procedimentos de minimização com o modelo PB. Em seguida, discutiremos os resultados. Por último, apresentaremos nossas conclusões e perspectivas futuras.

## 4.2 Coleta e preparação dos dados

**Notação** Usaremos a notação usual para os pares não-canônicos do DNA CC e TT. Os empilhamentos entre as bases serão representados por um (par de base 1)-(par de base 2). Como exemplo, utilizaremos o *mismatch* CC e o par canônico AT, correspondente ao dímero 5'-CA-3'/3'-CT-5'

Dados	<i>mismatch</i>	presença de metal	$N$	Morse ( $D$ )	empilhamento ( $k$ )		
C-Ag <sup>+</sup> -C	CC	+Ag	22	CC(24)	AT-CC(18) CC-CG(10)	CC-AT(12) CC-GC(6)	CC-CC(1)
CC-CTRL		-Ag	21	CC(23)	AT-CC(16) CC-CG(10)	CC-AT(12) CC-GC(6)	CC-CC(1)
T-Hg <sup>2+</sup> -T	TT	+Hg	34	TT(50)	AT-TT(26) TA-TT(14)	CG-TT(14) TT-TT(15)	GC-TT(16)
TT-CTRL		-Hg	24	TT(30)	AT-TT(21) TA-TT(9)	CG-TT(9) TT-TT(5)	GC-TT(11)
C-Ag <sup>+</sup> -C   T-Hg <sup>2+</sup> -T	CC e TT	+Ag e +Hg	57	CC(25) TT(51)	AT-CC(18) CC-CC(1) CC-TT(1) TA-TT(14)	AT-TT(27) CC-CG(10) CG-TT(14) TT-TT(15)	CC-AT(13) CC-GC(6) GC-TT(16)

**Tabela 1**

Resumo dos pares de base e próximos vizinhos associados aos parâmetros analisados. Na coluna  $N$ , apresentamos o número total de sequências extraídas da literatura de cada base de dados. Os sinais + e - ao lado do átomo metálico indicam, respectivamente, a presença e a ausência do átomo Ag ou Hg entre os *mismatches*. O número  $n$  entre parênteses, nas colunas Morse e empilhamento, informa a quantidade de vezes que cada parâmetro aparece em nossa base de dados.

será representado por CC-AT. Devido a condições de simetria o par CC-AT é igual ao par TA-CC reduzindo o número de parâmetros utilizados em nossas análises.

Para diferenciar entre a presença ou ausência de metal, Ag<sup>+</sup> ou Hg<sup>2+</sup>, no duplexo utilizaremos a seguinte notação: sequências na ausência de metal serão chamadas controle CC-CTRL e TT-CTRL; sequências na presença dos metais serão chamadas C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T, que é uma notação comum nesta área.

**Dados de temperatura de desnaturação** Extraímos sequências de DNA mediado por metal da literatura [10, 14, 17, 21, 37, 44, 45, 95, 97, 97, 102–111, 128, 129, 129]. Inicialmente, observamos os tipos de primeiros vizinhos presentes nestas sequências de M-DNA com os MMs: CC e TT (com e sem metais - M: Ag ou Hg), diferentes dos pares canônicos. Com isso, identificamos os parâmetros de *stacking*,  $k$ , para minimizarmos, além do potencial de Morse,  $D$ , entre as bases nitrogenadas dos MMs. Todas as sequências utilizadas apresentam um MM intermediado, ou não, por um átomo metálico. Na tabela 1, apresentamos a quantidade de sequências extraídas da literatura e os parâmetros a serem estudados próximos à região dos MMs.

Notamos que algumas sequências tinham concentrações de duplexos diferentes. Mas, como dito anteriormente, às regressões lineares, equações 27 e 28, exigem que as sequências dos duplexos possuam a mesma concentração. Dessa forma, utilizamos o modelo de próximos vizinhos

Parâmetro	CC-CTRL	Parâmetro	TT-CTRL
$D_{CC}$	2.28 meV	$D_{TT}$	5.06 meV
$k_{AT-CC}$	0.33 eV/nm <sup>2</sup>	$k_{AT-TT}$	4.32 eV/nm <sup>2</sup>
$k_{CC-AT}$	0.75 eV/nm <sup>2</sup>	$k_{CG-TT}$	1.78 eV/nm <sup>2</sup>
$k_{CC-CC}$	2.28 eV/nm <sup>2</sup>	$k_{GC-TT}$	0.60 eV/nm <sup>2</sup>
$k_{CC-CG}$	0.41 eV/nm <sup>2</sup>	$k_{TA-TT}$	2.68 eV/nm <sup>2</sup>
$k_{CC-GC}$	1.20 eV/nm <sup>2</sup>	$k_{TT-TT}$	4.37 eV/nm <sup>2</sup>

**Tabela 2**

Parâmetros  $k$  e  $D$  iniciais utilizados para as minimizações das sequências CTRL.

(subseção 3.2) para normalizar a concentração de todas as sequências. Como referência, usamos as concentrações de duplexos que mais se repetiam em cada conjunto de dados experimentais, separados de acordo com o MM observado.

Além das sequências e concentração de duplexos, retiramos dos artigos a temperatura de desnaturação e a concentração de sal (Na<sup>+</sup>). A concentração de sal mais comum nos experimentos é de 100 mM. No caso de valores diferentes, aplicamos a equação (22) do artigo [130] para normalizar a concentração de sal de todas às sequências para 100 mM.

### 4.3 Procedimento de minimização

Os procedimentos das minimizações para as sequências controle, CC-CTRL e TT-CTRL, foram iguais aos descritos na subseção 3.1.3. Como valores iniciais, utilizamos os da referência Oliveira et al. [15], mostrados na tabela 2. No entanto, o procedimento das sequências C-Ag<sup>+</sup>-C, T-Hg<sup>2+</sup>-T e C-Ag<sup>+</sup>-C | T-Hg<sup>2+</sup>-T foram um pouco diferentes das sequências CTRL.

**CTRL** Para as sequências controle encontramos um problema, onde alguns parâmetros tendiam para zero. Um dos desafios é que, uma vez que um parâmetro chega a zero, ele tende a permanecer nesse valor durante às minimizações subsequentes. Isso pode resultar em valores sub-ótimos para os demais parâmetros. Entretanto, não podemos simplesmente descartar esses valores zero, especialmente considerando que MMs possuem valores muito pequenos de potenciais de Morse  $D$  e empilhamento  $k$ . Para assegurar que os valores próximos a zero não sejam artefatos algorítmicos, identificamos os parâmetros que iam para zero e os substituímos pelo valor inicial da referência Oliveira et al. [15] Em seguida, repetimos todo o processo com uma mistura de parâmetros otimizados, não nulos, e aqueles que foram artificialmente alterados. Apenas três parâmetros do conjunto CC-CTRL (CC, AT-CC e CC-CC) e um parâmetro do conjunto TT-CTRL (TT-TT) apresentaram tal problema. Se após muitas rodadas de minimizações o parâmetro permanecia em zero, isto sugeria que esse era, de fato, seu valor otimizado. Após essa etapa, realizamos mais três rodadas de minimizações,

Parâmetro	C-Ag <sup>+</sup> -C	Parâmetro	T-Hg <sup>2+</sup> -T
$D_{CC}$	100.91 meV	$D_{TT}$	78.63 meV
$k_{AT-CC}$	1.86 eV/nm <sup>2</sup>	$k_{AT-TT}$	6.36 eV/nm <sup>2</sup>
$k_{CC-AT}$	8.55 eV/nm <sup>2</sup>	$k_{CG-TT}$	2.29 eV/nm <sup>2</sup>
$k_{CC-CC}$	0.03 eV/nm <sup>2</sup>	$k_{GC-TT}$	3.46 eV/nm <sup>2</sup>
$k_{CC-CG}$	0.92 eV/nm <sup>2</sup>	$k_{TA-TT}$	3.97 eV/nm <sup>2</sup>
$k_{CC-GC}$	0.42 eV/nm <sup>2</sup>	$k_{TT-TT}$	0.69 eV/nm <sup>2</sup>

**Tabela 3**

Parâmetros  $k$  e  $D$  iniciais obtidos da otimização preliminar e utilizados na primeira rodada de minimizações de C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T.

sempre usando os parâmetros da rodada anterior como valores iniciais. Os conjuntos CC-CTRL e TT-CTRL foram otimizados separadamente. Cada rodada de minimização foi realizada 1000 vezes. Posteriormente, uma quarta e última rodada de minimização foi realizada para avaliar a influência do erro experimental nos novos parâmetros, como descrito na subseção 3.1.3. Nós utilizamos uma incerteza de 0.6 °C para as sequências controle, valor obtido das médias das incertezas dos valores de temperatura de desnaturação das sequências extraídas da literatura.

**C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T** Na rodada preliminar, juntamos os conjuntos de sequências TT-CTRL com T-Hg<sup>2+</sup>-T e CC-CTRL com C-Ag<sup>+</sup>-C. Variamos apenas os parâmetros contendo um metal por MM enquanto mantivemos os parâmetros de controle fixos. Esta rodada preliminar possui o efeito prático que todas as temperaturas de desnaturação são forçadas a seguir os mesmos parâmetros de regressão das equações (28). Dessa maneira, os parâmetros de C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T assumem valores que possuem alguma referência com os parâmetros controle (tabela 3). Utilizamos esta abordagem apenas para obter parâmetros iniciais para C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T. Em seguida, realizamos três rodadas de otimizações de C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T separadas das sequências CTRL. Para a última rodada, seguimos o mesmo procedimento que a última rodada de otimização para as sequências controle (CTRL), assumimos uma incerteza de 0.6 °C para avaliar a influência do erro experimental em nossos resultados. Cada rodada de minimizações também foi executada 1000 vezes.

**C-Ag<sup>+</sup>-C | T-Hg<sup>2+</sup>-T** Após realizar as minimizações para C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T, ficamos interessados em saber como uma hélice de DNA se comportaria ao conter simultaneamente os metais Ag<sup>+</sup> e Hg<sup>2+</sup>. Infelizmente, encontramos apenas uma única sequência que continha ambos os metais [128]. Nesse caso, nossa abordagem foi estimar o parâmetro de empilhamento CC-TT, que representa a interação entre os dois metais. Para isso, combinamos os conjuntos C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T em um único conjunto e incluímos a sequência contendo o dímero CC-TT (tabela 1). Como realizado para os metais C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T, analisados de forma separada, o efeito prático consiste em forçar a regressão da sequência CC-TT a seguir os mesmos parâmetros de regressão das sequências

Conjunto de dados	regressão por tamanho da sequência			
	antes da primeira minimização		depois da última minimização	
	$\langle \Delta T \rangle$ (°C)	$\langle \chi^2 \rangle$ (°C <sup>2</sup> )	$\langle \Delta T \rangle$ (°C)	$\langle \chi^2 \rangle$ (°C <sup>2</sup> )
CC-CTRL	3.57	38.98	1.24	2.79
C-Ag <sup>+</sup> -C	3.33	26.13	1.86	5.05
TT-CTRL	6.36	95.39	1.79	6.01
T-Hg <sup>2+</sup> -T	3.80	35.71	1.35	2.77
C-Ag <sup>+</sup> -C   T-Hg <sup>2+</sup> -T	2.32	10.44	2.32	10.44

**Tabela 4**

Comparação entre os  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$  antes da primeira e após a última minimização na presença e ausência de metal. Para ficar mais fácil de compararmos os resultados, normalizamos o valor de  $\chi^2$  dividindo-o pelo número de sequências  $N$  mostrados na tabela 1.

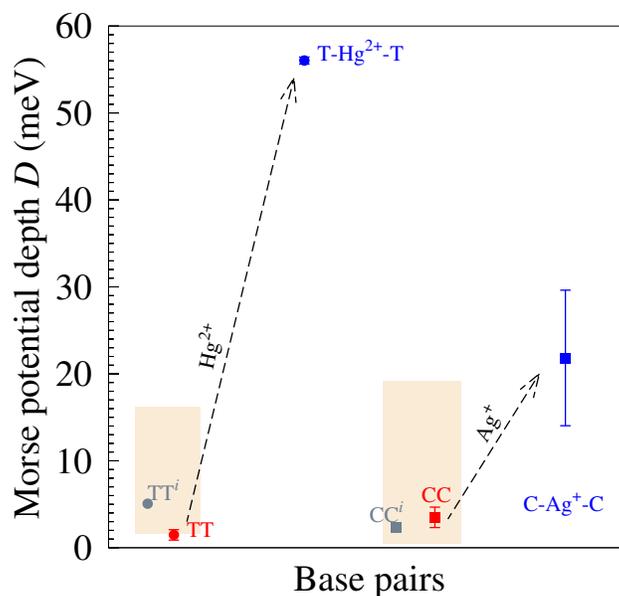
C-Ag<sup>+</sup>-C e T-Hg<sup>2+</sup>-T. Realizamos apenas uma rodada de minimização, executada 1000 vezes, mantendo os parâmetros de Morse,  $D$ , e empilhamento,  $k$ , ajustados para as sequências com os metais Ag<sup>+</sup> e Hg<sup>2+</sup>, com valores fixos aos encontrados na subseção 4.4 (figuras 23 e 24). Utilizamos um valor inicial genérico de 2.5 eV/nm<sup>2</sup> para o parâmetro de empilhamento CC-TT. O objetivo dessa otimização era procurar especificamente o empilhamento CC-TT que representasse a interação entre Ag<sup>+</sup> e Hg<sup>2+</sup>.

## 4.4 Resultados

As rodadas de otimização foram independentes para T-Hg<sup>2+</sup>-T e C-Ag<sup>+</sup>-C, assim como para suas respectivas sequências controle. No total, quatro rodadas com 1000 condições iniciais diferentes foram realizadas, como descrito na subseção 4.3. Os parâmetros de qualidade finais estão descritos na tabela 4. Os valores médios encontrados,  $\langle \Delta T \rangle$ , diferem entre 1.24 °C e 1.79 °C dos valores experimentais. Estes valores estão dentro da faixa de erro experimental dos aparelhos de medição, o que indica que nossos resultados estão muito bons.

Na figura 23 mostramos os potenciais de Morse otimizados. O parâmetro  $D$  de TT-CTRL é de alguma forma menor do que seu valor inicial  $TT^i$  que foi obtido da referência Oliveira et al. [15]. O parâmetro  $D$  de CC-CTRL é também menor do que o valor  $CC^i$ , apesar de maior que o valor de TT. No entanto, ambos os valores são muito menores quando comparados com os pares de base canônico AT e CG. Portanto, não é possível dizer com estes resultados que o par CC é mais estável que o par TT.

A adição de Hg<sup>2+</sup> entre o par TT aumenta enormemente o valor do potencial de Morse  $D$ . Esse aumento é muito maior do que qualquer configuração de TT conhecida [15] (área sombreada) e aproximadamente 4/5 do potencial de Morse  $D$  do par CG (70meV) [28]. Isto deixa evidente que T-Hg<sup>2+</sup>-T é estabilizado por um Hg<sup>2+</sup> entre os MM TT. Para entender este resultado, vamos



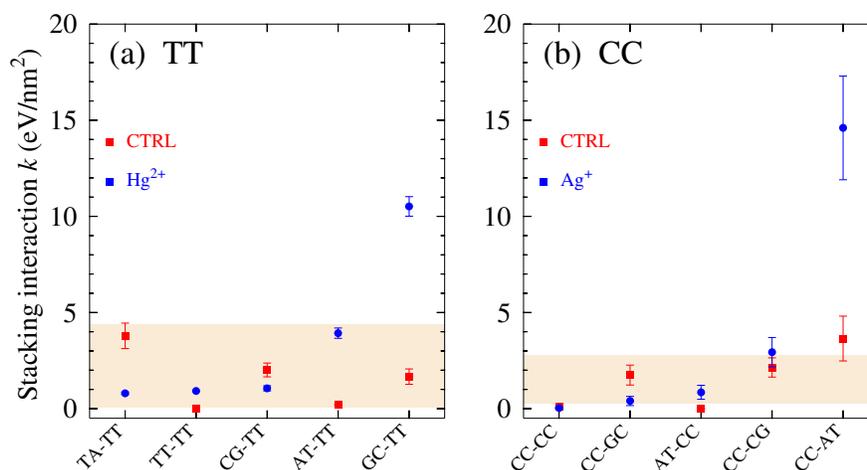
**Figura 23**

Profundidades médias do potencial de Morse  $D$  calculadas para  $C-Ag^+-C$ ,  $T-Hg^{2+}-T$  em azul e seus respectivos controles em vermelho. As barras de erro são os desvios padrões da última rodada de minimizações. Os valores de semente para os *mismatches*  $TT^i$  e  $CC^i$  são mostrados em cinza. As setas tracejadas indicam as mudanças devido a adição de cátions. A área sombreada mostra o intervalo mínimo e máximo para potenciais de Morse para todos os tipos de *mismatches* TT e CC da referência Oliveira et al. [15]

discutir brevemente  $T-Hg^{2+}-T$ . Foram propostas três configurações:  $N3-Hg^{2+}-N3$ ,  $O2-Hg^{2+}-N3$  e  $N3-Hg^{2+}-O4$ . A estrutura  $N3-Hg^{2+}-N3$  possui os átomos de oxigênio opostos um do outro, isto é suficiente para manter a força de repulsão entre eles não tão forte, o que explicaria o alto valor de  $D$  encontrado (figura 23). Entretanto, as estruturas  $O2-Hg^{2+}-N3$  e  $N3-Hg^{2+}-O4$  também são possíveis. Infelizmente, nosso modelo não permite identificar qual a estrutura, apenas aponta para a estrutura que apresenta a ligação mais forte entre os pares.

A adição de  $Ag^+$  também aumenta o potencial de Morse  $D$  do par CC (figura 23). Porém, este aumento é bem menor do que  $T-Hg^{2+}-T$ , aproximadamente  $2/3$  do potencial de Morse  $D$  do par AT (34 meV). Isso sugere uma estrutura similar à  $N3-Ag^+-N3$ . O valor mais baixo é por causa da proximidade dos grupos amina no  $C-Ag^+-C$ , resultando em grandes ângulos de torção. Outros tipos de estruturas foram sugeridos, como por exemplo, a configuração  $N3-Ag^+-O2$ , que possui uma ligação de hidrogênio adicional. No entanto, essa possibilidade é improvável devido ao valor moderado do potencial de Morse  $D$  encontrado.

A partir dos resultados do potencial de Morse  $D$ , figura 23, não há dúvida de que  $T-Hg^{2+}-T$  é estabilizado por uma ligação planar entre os pares de base, enquanto para  $C-Ag^+-C$ , não podemos fazer tal afirmação. Então o que explicaria o aumento na temperatura quando  $Ag^+$  é adicionado? As primeiras sequências apresentadas na tabela 21 do apêndice A dão pistas do que



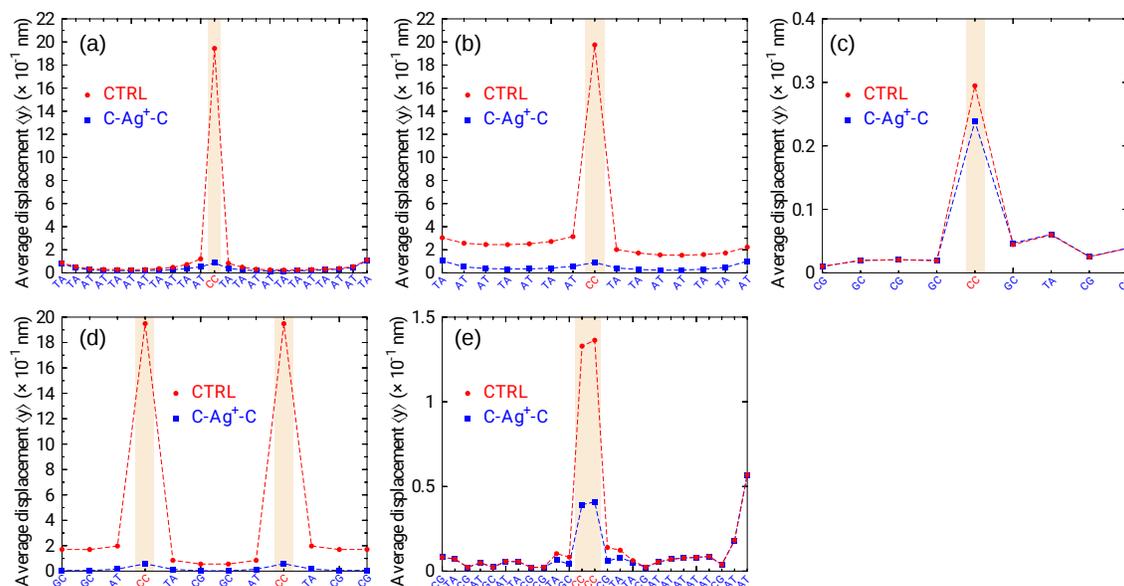
**Figura 24**

Interações de empilhamento médias  $k$  calculada para (a) T-Hg<sup>2+</sup>-T e (b) C-Ag<sup>+</sup>-C como pontos azuis. Os pontos vermelhos são os controles (a) TT-CTRL e (b) CC-CTRL. As barras de erro são os desvios padrões da última rodada de minimizações. A área sombreada mostra o intervalo mínimo e máximo para os empilhamentos dos *mismatches* TT e CC da referência Oliveira et al. [15]

pode estar acontecendo. Há 6 sequências de mesmo tamanho com um MM central CC, onde apenas os próximos vizinhos mudam. Para o trímero central ACA/TCT, há um aumento de 10 °C, enquanto para o CCG/GCC, o aumento é apenas 2 °C. Esta grande diferença indica que os próximos vizinhos têm um papel importante na estabilidade. A interação de próximos vizinhos é modelada pela constante  $k$ , equação (11), que está associada com a interação de empilhamento. Nossos resultados para as constantes de acoplamento  $k$  são mostradas na figura 24 tanto para T-Hg<sup>2+</sup>-T quanto para C-Ag<sup>+</sup>-C. De forma marcante, a interação de empilhamento mais alta é do próximo vizinho CC-AT, presente exatamente no trímero ACA/TCT. Interessantemente, o próximo vizinho AT-CC do trímero ACA/TCT possui uma interação muito fraca. A interação CC-CC, dois MMs consecutivos CC, é praticamente inexistente, o que indica que o aumento de estabilidade é devido às ligações planares entre as bases com o íon de Ag<sup>+</sup> no centro. Ligações interplanares foram previstas por DFT e observadas em estruturas cristalinas de N<sup>1</sup>-hexilcitosina. Portanto, era esperado um valor maior de  $k$  para o próximo vizinho CC-CC.

Para os parâmetros de empilhamento de T-Hg<sup>2+</sup>-T, mostrados na figura 24, observamos um empilhamento com valor muito alto do que os previamente encontrados para MM TT (área sombreada), que foi o do próximo vizinho GC-TT. AT-TT aumentou moderadamente em relação ao TT-CTRL, enquanto TA-TT é muito pequeno. Apesar de que TA-TT do TT-CTRL, sua contraparte, ser alto.

Na figura 25 apresentamos os perfis de abertura média de sequências obtidas a partir de estudos de ressonância magnética nuclear (NMR) ou espalhamento de raio X para CC-CTRL e

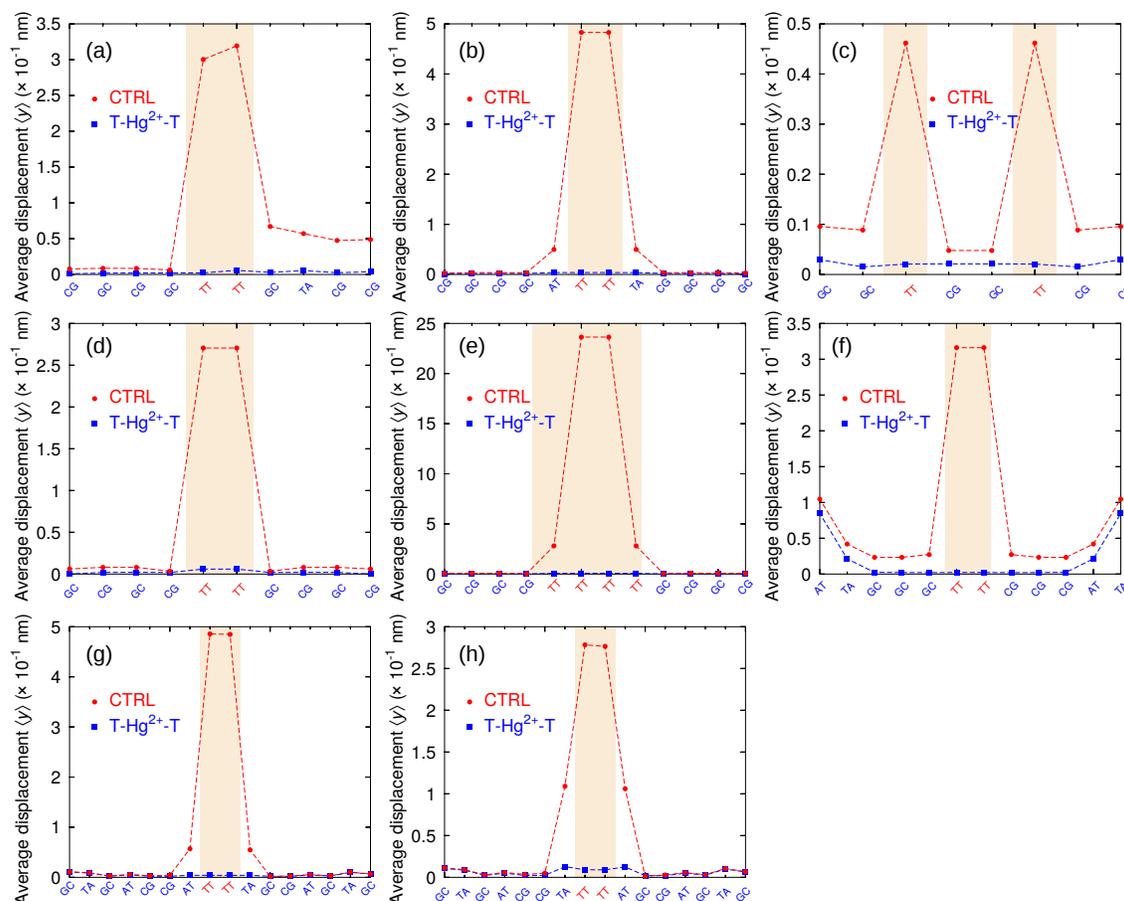


**Figura 25**

Deslocamento médio para as sequências das referências [12, 16–18]. Os pontos vermelhos são das CC-CTRL e os azuis são das C-Ag<sup>+</sup>-C.

C-Ag<sup>+</sup>-C [12, 16–18]. Utilizamos sequências diferentes das usadas para os cálculos dos parâmetros para validar os nossos resultados. Estes gráficos têm como característica mostrar às regiões do nucleotídeo que são mais, ou menos, estáveis. Valores altos de  $\langle y \rangle$  indicam pontos de menor estabilidade, onde a fita se abrirá antes de outros pontos no duplexo. Observamos que, em todos os gráficos da figura 25, a adição de metal (pontos azuis) estabiliza o nucleotídeo (região sombreada). Nas figuras 25a, 25b e 25d, onde os próximos vizinhos são os mesmos, do tipo AT-CC (por condições de simetria AT-CC = CC-TA), a estabilidade no ponto de MM é imensa com a adição de Ag<sup>+</sup> (variou aproximadamente de  $20 \times 10^{-1}$  nm, CC-CTRL, para  $1 \times 10^{-1}$  nm, C-Ag<sup>+</sup>-C). Na figura 25c a estabilidade é bem menor, com a abertura próxima à observada em CC-CTRL. Nesse caso, temos dois tipos de próximos vizinhos, GC-CC, equivalente a CC-CG, e CC-GC. Aparentemente, quando o próximo vizinho é deste tipo, a presença de um equivalente de Ag<sup>+</sup> não faz tanta diferença. Isto é explicado novamente pelas 6 primeiras sequências da tabela 21 do apêndice A. As primeiras três sequências, cujos próximos vizinhos são do tipo CC-AT ou AT-CC, possuem um aumento de temperatura,  $T_{exp}$ , com a adição do Ag<sup>+</sup> muito maior do que as sequências 4–6, onde os próximos vizinhos são do tipo CC-GC ou CC-CG. Na figura 25e mesmo com CC-CC aproximadamente zero, o valor de Morse  $D$  C-Ag<sup>+</sup>-C, maior do que CC-CTRL, é suficiente para estabilizar o nucleotídeo. Nossos resultados corroboram com os resultados de NMR e raio X [12, 16–18] de que a Ag<sup>+</sup> é inserida entre o par CC e, além disso, a ligação é direta, do tipo covalente entre os nitrogênios N(3) da citosina.

Na figura 26, apresentamos os perfis de abertura média das sequências estudadas por

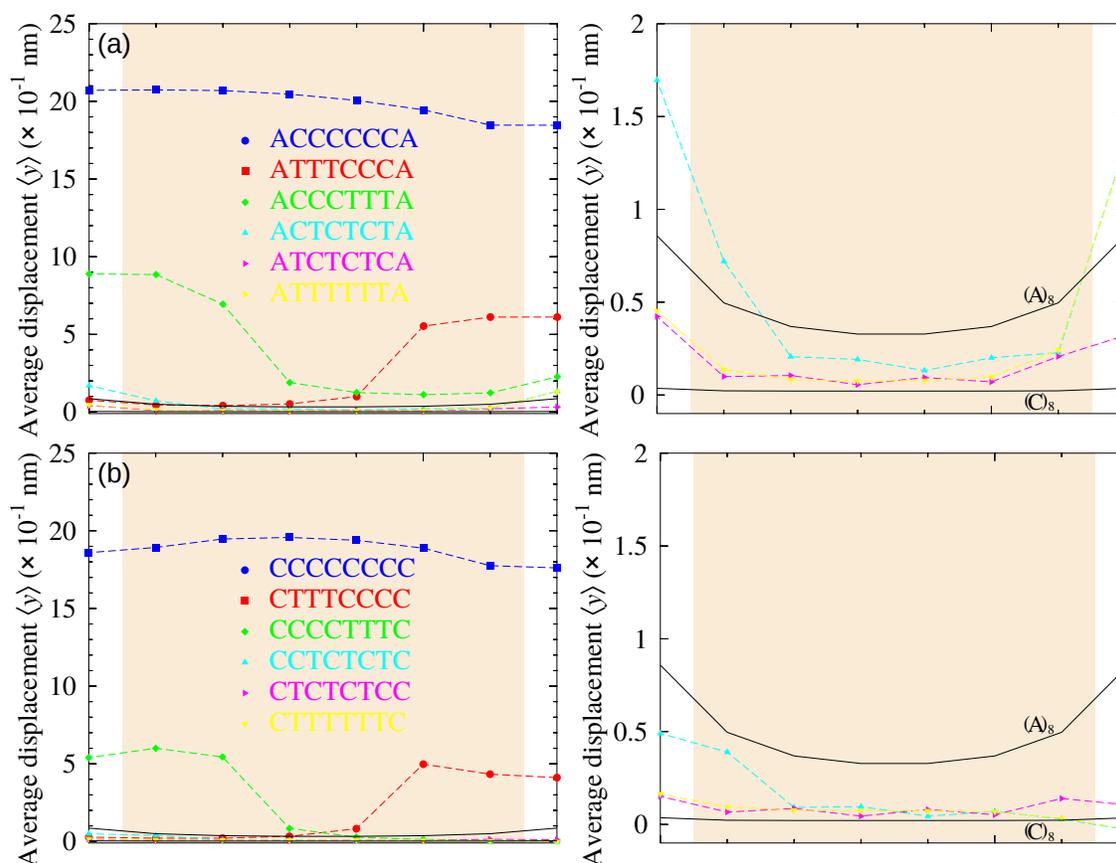


**Figura 26**

Deslocamento médio para as seqüências das referências [11, 14, 19–22]. Na parte (h) AT-TT da parte (g) foi substituído por TA-TT. Os pontos vermelhos são das TT-CTRL e os azuis são das T-Hg<sup>2+</sup>-T.

ressonância magnética nuclear ou espalhamento de raio X para TT-CTRL e T-Hg<sup>2+</sup>-T [11, 14, 19–22]. Como nas seqüências C-Ag<sup>+</sup>-C e CC-CTRL, a inserção do Hg<sup>2+</sup> entre o par TT estabiliza bastante o nucleotídeo. As seqüências CTRL mostram grandes deslocamentos na região do MM (área sombreada), enquanto a adição de Hg<sup>2+</sup> resulta em uma completa estabilização (deslocamentos médios muito pequenos). Essa estabilidade se deve principalmente ao alto valor do potencial de Morse  $D$  T-Hg<sup>2+</sup>-T encontrado. Nenhuma das seqüências estudadas por NMR ou espalhamento de raio X possui o próximo vizinho TA-TT. Para avaliar a influência do empilhamento TA-TT na abertura da fita, substituímos o par AT-TT da figura 25g por TA-TT, conforme ilustrado na figura 25h. Observamos apenas uma pequena perturbação por causa do valor muito menor do empilhamento, mas isso não é suficiente para causar uma grande desestabilização. Claramente, o potencial de Morse  $D$  do T-Hg<sup>2+</sup>-T é suficientemente forte para compensar o pequeno valor do empilhamento TA-TT.

Durante a realização deste estudo, despertou nosso interesse comparar o comportamento de uma fita contendo Ag<sup>+</sup> e Hg<sup>2+</sup> com os casos individuais de uma fita composta exclusivamente por



**Figura 27**

Estruturas possíveis de DNA mediado por metal  $\text{Ag}^+$ ,  $\text{Hg}^{2+}$  e híbridos. Destacamos a região sombreada dos gráficos, onde são a dos *mismatches* mediados por um átomo metálico na estrutura. (a) Estruturas de DNA com o par AT nas pontas. O gráfico à direita mostra a região entre 0 e 2 Å ampliada. (b) Estruturas de DNA com o par CG nas pontas. O gráfico à direita mostra a região entre 0 e 2 Å ampliada. As linhas em preto das figuras do lado direito são de pares canônicos contendo 8 pares de base AT ( $\text{A}_8$ ) e 8 pares CG ( $\text{C}_8$ ) para ser a nossa referência.

$\text{Ag}^+$  e outra fita composta exclusivamente por  $\text{Hg}^{2+}$ . No entanto, só encontramos uma sequência publicada que continha os dois metais  $\text{Ag}^+$  e  $\text{Hg}^{2+}$ , um ao lado do outro [128]. Dado que não possuíamos mais sequências, nossa única opção foi tentar estimar o parâmetro CC-TT, que representa a interação entre os dois metais. Para isso, combinamos os conjuntos de dados C- $\text{Ag}^+$ -C e T- $\text{Hg}^{2+}$ -T e adicionamos a sequência contendo o dímero CC-TT. Após a minimização, encontramos o valor de  $3.43 \text{ eV/nm}^2$  para CC-TT. Ao compararmos esse valor com os resultados obtidos nas figuras 24a e 24b, observamos que o valor de  $3.43 \text{ eV/nm}^2$  não excede a região sombreada. Logo, não observamos interação entre os metais. Outro fato interessante é que a hélice se torna mais rígida na junção entre os dois metais na estrutura do DNA. Para comparar os três casos de DNA, plotamos os gráficos da figura 27.

A figura 27, consiste de seis tipos de hélice de DNA contendo metais entre os *mismatches*

no meio da estrutura, com pares canônicos AT na figura 27a e CG na figura 27b nas extremidades. A escolha de pares canônicos nas pontas, ao invés de uma hélice composta exclusivamente por metais, se deve ao fato de que o efeito dos metais no centro da estrutura é diferente do efeito nas extremidades. Funai et al. [102] avaliaram o efeito da adição dos metais  $\text{Ag}^+$  e  $\text{Hg}^{2+}$  em nucleotídeos contendo um *mismatch* CC e um TT nas pontas e no centro do DNA. Embora observamos uma grande diferença de temperatura de desnaturação quando os *mismatches* se encontram no centro, o mesmo efeito não é observado com os *mismatches* nas pontas. No caso da prata, não há diferença entre as temperaturas de desnaturação na ausência ou presença de  $\text{Ag}^+$ .

A ordem de estabilidade das estruturas contendo os três casos: C- $\text{Ag}^+$ -C, T- $\text{Hg}^{2+}$ -T e C- $\text{Ag}^+$ -C | T- $\text{Hg}^{2+}$ -T, mostra que T- $\text{Hg}^{2+}$ -T é o mais estável seguida pela hélice híbrida, C- $\text{Ag}^+$ -C e T- $\text{Hg}^{2+}$ -T, e o caso menos estável é a hélice contendo puramente C- $\text{Ag}^+$ -C no centro do duplexo, conforme ilustrado na figura 27. Nossos resultados concordam com os achados de Ono and Miyake [128], que registraram temperaturas de desnaturação de duplexos contendo os três casos. Dependendo da combinação de CC e TT (linhas vermelho, verde, ciano e roxo, figura 27) no duplexo, encontramos casos onde a estabilidade é da ordem ou até um pouco melhor à do caso T- $\text{Hg}^{2+}$ -T (linha amarela, figura 27). Comparando com os dois extremos de pares canônicos, ( $A_8$ ) e ( $C_8$ ), podemos observar que a ordenação das bases TCTCTC e TTTTTT no centro resulta em hélices de DNA mais estáveis do que uma hélice composta por oito pares de bases AT.

A robustez dos resultados apresentados são encorajadores e devem nos permitir aplicar o método em outras configurações, como o par C- $\text{Hg}^{2+}$ -T. Um conjunto limitado de temperaturas de desnaturação estão disponíveis, veja Ono et al. [37], Funai et al. [45], Torigoe et al. [95], Funai et al. [102], Han et al. [108], Mata et al. [109], Schmidt et al. [110], Torigoe et al. [111], Fujii et al. [129], porém não suficiente em tamanhos e variações dos próximos vizinhos em torno do par CT. Outro problema é que as propriedades de simetria do par CT não nos permite reduzir o número de parâmetros estudados, como no caso dos pares CC e TT, portanto sendo necessárias muito mais sequências do que as utilizadas neste trabalho. Aqui, o menor conjunto de dados foi para C- $\text{Ag}^+$ -C, com 22 sequências, que é o limite inferior do método. Outro aspecto de interesse é a concentração de cátions; aqui, utilizamos temperaturas de desnaturação medidas com um equivalente de metal por MM e, claro, zero para os CTRL. Seria desejável repetir esse processo para um conjunto de concentrações que nos proporcionaria uma compreensão do processo de estabilização induzidos pelos cátions. As estabilidades das diferentes estruturas de DNA com metal, apresentadas na figura 27, nos levaram a outra pergunta: qual tipo de hélice é mais importante para o transporte de carga? Uma hélice rígida ou uma hélice flexível? Apesar de nossos resultados estarem condizentes com a literatura, necessitamos de mais sequências com temperaturas de desnaturação contendo os dois metais na estrutura para confirmar nossas observações e tentar responder a essa pergunta.

Outra coisa que podemos fazer, é utilizar uma modificação do modelo PB, conhecida como PB Holstein [131], que trata diretamente sobre transporte de carga em DNA. Esperamos que nossos resultados estimulem a realização de mais experimentos de UV *melting* com metais entre os pares de base.

## 4.5 Conclusões e Perspectivas

Neste capítulo, analisamos as temperaturas de desnaturação com a presença de metais entre os pares de base para T-Hg<sup>2+</sup>-T e C-Ag<sup>+</sup>-C. Para T-Hg<sup>2+</sup>-T, encontramos um grande valor de potencial de Morse  $D$  da ordem de um par CG, enquanto para C-Ag<sup>+</sup>-C encontramos um valor muito menor, aproximadamente  $2/3$  de um par AT. No caso de C-Ag<sup>+</sup>-C, o valor de Morse pequeno favorece a configuração N3-Ag<sup>+</sup>-N3 proposta por Dairaku et al. [132]. Embora o empilhamento CC-CC tenha sido muito fraco na presença de Ag<sup>+</sup>, o potencial de Morse para CC foi suficientemente forte para conferir alguma estabilidade em comparação com a ausência de Ag<sup>+</sup>. O empilhamento TA-TT encontrado foi também muito pequeno, porém o potencial de Morse T-Hg<sup>2+</sup>-T foi grande o suficiente para estabilizar o duplexo. Conseguimos estimar o parâmetro de empilhamento CC-TT para uma hélice híbrida. O valor encontrado para este empilhamento é bem maior do que nos casos CC-CC para C-Ag<sup>+</sup>-C e TT-TT para T-Hg<sup>2+</sup>-T. Os perfis de abertura das três situações mostram que a hélice de DNA T-Hg<sup>2+</sup>-T é mais estável que a hélice C-Ag<sup>+</sup>-C. A situação híbrida, C-Ag<sup>+</sup>-C | T-Hg<sup>2+</sup>-T, encontra-se entre esses dois extremos. É interessante notar que, dependendo do arranjo dos pares CC e TT na hélice híbrida do DNA, encontramos casos onde a estabilidade é maior do que a de uma hélice de DNA composta puramente por pares de base AT.

Nossos resultados estão de acordo com o que se conhece sobre DNA mediado por metal em T-Hg<sup>2+</sup>-T, C-Ag<sup>+</sup>-C e C-Ag<sup>+</sup>-C | T-Hg<sup>2+</sup>-T. Portanto, acreditamos que seja possível expandir o método para outros casos de DNA mediado por metal, por exemplo C-Hg<sup>2+</sup>-T, quando um conjunto de temperaturas de desnaturação estiver disponível. Além disso, com base nos nossos resultados, podemos ampliar o estudo para estes metais com bases artificiais, tais como: imidazol, fluoracil, tiotimina, pirazoli e outras mais presentes na literatura [37, 97, 98, 100, 103, 133–136]. Uma questão adicional que podemos nos aprofundar é sobre o transporte de carga no DNA e como a adição dos metais auxilia nesse processo.

## 5 Nucleotídeos em ambiente macromolecular

Neste capítulo, vamos apresentar a revisão bibliográfica e os resultados do segundo projeto, que aborda a estabilidade dos nucleotídeos em um ambiente macromolecular. Na próxima subseção, apresentaremos uma revisão bibliográfica sobre como o ambiente em que os nucleotídeos estão inseridos afeta sua estabilidade. Nessa parte incluímos estudos experimentais e teóricos. Após isso, mostraremos nossos resultados, conclusões e perspectivas futuras. No apêndice A, tabelas 23, 24 e 25, estão as sequências encontradas na literatura e utilizadas neste projeto.

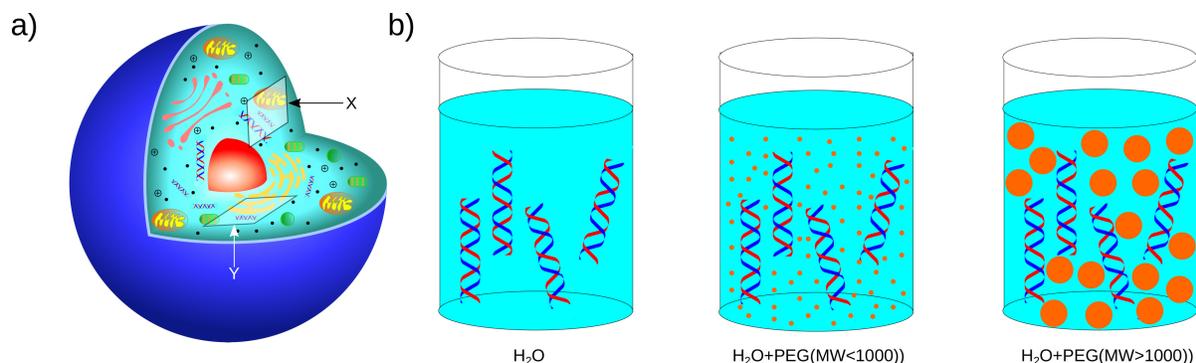
### 5.1 Revisão bibliográfica

### 5.2 Nucleotídeos em ambiente macromolecular

Aproximadamente 20-40% do volume celular é composto por biomoléculas [41, 137–139]. Além das biomoléculas, cátions são abundantes dentro das células, desempenhando papéis importantes na regulação das funções biológicas dos ácidos nucleicos [35, 49, 140]. A maioria dos experimentos sobre estabilidade de oligonucleotídeos é realizada em soluções aquosas bastante diluídas [78] (figura 28b à esquerda), o que é completamente diferente do meio celular. Tanto as macromoléculas quanto as micromoléculas presentes no ambiente celular influenciam a estabilidade dos nucleotídeos, e compreender este fenômeno é fundamental para entender os ácidos nucleicos e suas funções nos seres vivos [33].

O citoplasma da célula é composto principalmente por água (figura 28a). Entretanto, uma parte significativa de seu volume é ocupada por moléculas orgânicas e inorgânicas, solúveis ou insolúveis, tais como: aminoácidos, proteínas, lipídios, sais inorgânicos ( $\text{Na}^+$  e  $\text{K}^+$ , por exemplo) e nucleotídeos (DNA e RNA). Ademais, a célula é composta por várias organelas que ocupam um volume expressivo em seu interior [31]. Este conjunto de micromoléculas e macromoléculas é chamado de *crowding*.

O *Crowding* altera o equilíbrio químico de formação dos duplexos de duas maneiras principais: atividade de água e volume de exclusão. A atividade de água é alterada através da interação de micromoléculas orgânicas com as moléculas de  $\text{H}_2\text{O}$  próximas aos nucleotídeo. Essas moléculas de  $\text{H}_2\text{O}$  são cruciais para a estabilidade da dupla hélice. Entretanto, a simples presença dessas micromoléculas na célula perturba as moléculas de  $\text{H}_2\text{O}$ , deslocando o equilíbrio químico no sentido de formação de fitas simples, ou seja, a estabilidade do duplexo se reduz [33, 48, 141].



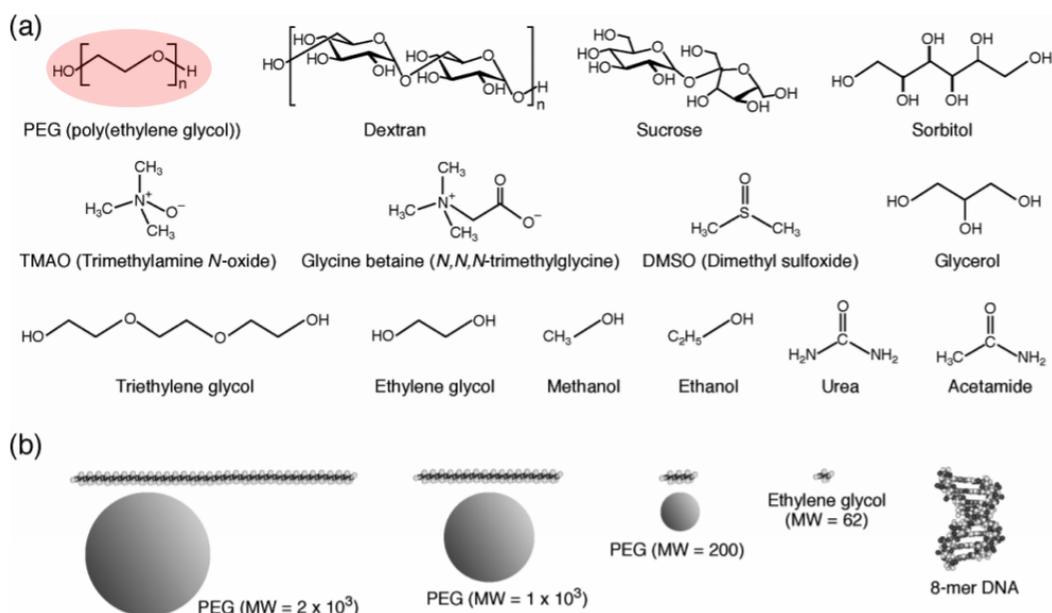
**Figura 28**

Diferenças entre experimentos *in vivo* e *in vitro*. a) célula eucarionte, as regiões *X* e *Y* são discutidas no texto, e b) experimentos *in vitro*.

Por outro lado, o volume de exclusão é o espaço ocupado por moléculas orgânicas na célula, reduzindo assim a quantidade de  $H_2O$  disponível para realização das funções biológicas. Tanto as micromoléculas quanto as macromoléculas apresentam esse efeito. No entanto, o volume de exclusão é mais evidente nas moléculas maiores. O volume indisponível confina o duplexo em um espaço menor, dificultando a separação da dupla hélice. Como resultado, o equilíbrio é deslocado no sentido de formação de fitas duplas, ou seja, a estabilidade do duplexo aumenta [33, 48, 141].

Uma das alternativas para se estudar tais fenômenos seria extrair porções da célula e analisar a termodinâmica do sistema. Contudo, as macro- e micro-moléculas presentes no meio celular estão distribuídas de forma heterogênea devido aos processos físico-químicos ocorridos no interior das células. Portanto, a porção *X* da célula coletada e analisada hoje será diferente da porção *X* coletada e analisada amanhã. Na figura 28a temos destacadas duas regiões *X* e *Y*. O conteúdo da célula se encontra em constante movimento, ou seja, aquela porção *X* demarcada pode se tornar a porção *Y* um tempo depois. Dessa forma, fica difícil validar o experimento, e isso adiciona um grau de complexidade extra às análises [47]. Para contornar esses problemas, os experimentos *in vitro* são realizados em  $H_2O$ , onde uma porcentagem da solução é composta por um solvente orgânico que simula o ambiente molecular (*crowding*). Além disso, a escolha do soluto escolhido deve respeitar alguns critérios específicos para garantir a eficiência do experimento. Esses critérios são:

1. Inerte (o soluto não deve interagir diretamente com o nucleotídeo);
2. Solúvel em  $H_2O$  (pelo menos algum percentual);
3. Não deve absorver em bandas próximas de 260 nm (faixa de absorção dos nucleotídeos);
4. No caso de grandes solutos, polímeros de vários tamanhos devem estar disponíveis;



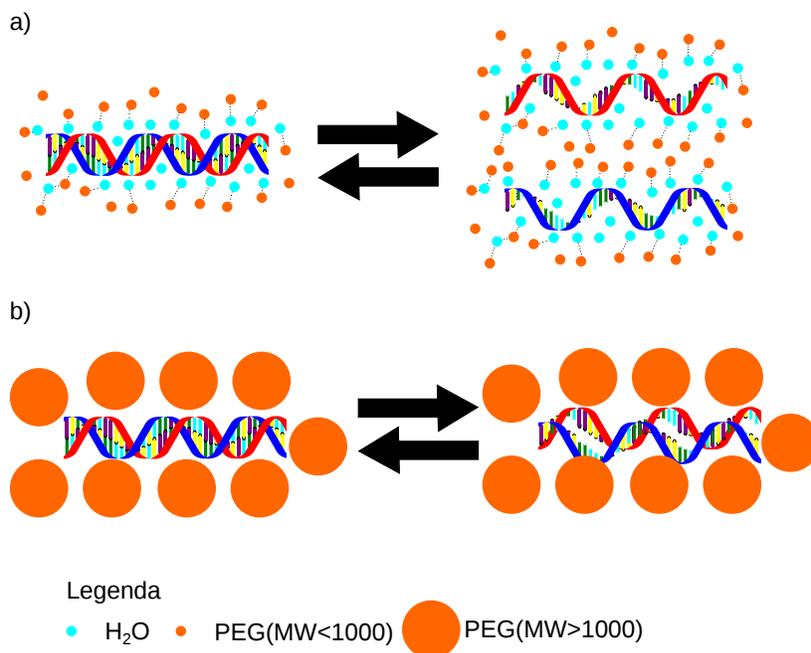
**Figura 29**

a) Estrutura química de alguns solutos utilizados para simular o ambiente molecular; b) comparação dos tamanhos dos PEGs de diferentes pesos moleculares, etilenoglicol e um pequeno duplexo de DNA. Figura extraída de Nakano et al. [23]

- No caso de pequenos solutos, solutos com uma gama de propriedades químicas deve ser usado [138, 139].

Solutos de grande tamanho comumente usados são PoliEtilenoGlicol (PEG), Dextrano e Ficoll [23, 142, 143]. O PEG é frequentemente escolhido por ser inerte e estar disponível em pesos moleculares diferentes. Proteínas como albumina, hemoglobina e lisozima também são utilizadas como *crowding*. Por outro lado, os solutos de pequeno tamanho comumente utilizados são etilenoglicol (EG), álcoois, glicóis, aminoácidos e betaína. Algumas dessas estruturas estão mostradas na figura 29.

A estrutura marcada em vermelho na figura 29a é o soluto mais utilizado nos experimentos [33, 35, 48, 49, 140, 141, 143–149]. A fórmula estrutural do PEG é  $\text{H}-(\text{O}-\text{CH}_2-\text{CH}_2)_n-\text{OH}$ , onde  $n$  representa a unidade estrutural que se repete ao longo da cadeia. Todas as estruturas com  $n \geq 4$  são coletivamente chamadas de PEG. O peso molecular da estrutura é determinado pela fórmula  $44n+18$ , geralmente mostrado a direita do nome PEG, por exemplo: PEG200, PEG400, PEG8000, entre outros. Essa informação fornece o tamanho da cadeia do PEG utilizado no experimento (figura 29b). PEGs com massa molecular, do inglês *Molecular Weight* (MW), entre 200 e 600 g/mol são líquidos à temperatura ambiente e possuem densidade entre 1.12 e 1.13 g/cm<sup>3</sup>. Enquanto, PEGs com MW maiores que 600 g/mol têm a aparência de cera sólida, com densidade de 1.2 g/cm<sup>3</sup> [150]. A medida que o tamanho da cadeia aumenta, a solubilidade do PEG em água diminui. Por exemplo, o PEG4000

**Figura 30**

Efeito de PEG na estabilidade do nucleotídeo. Mostramos o efeito de pequenos solutos e grandes solutos, esferas laranja, nos estados de fita dupla e fita simples na temperatura de desnaturação. a) PEGs com massas moleculares menores do que 1000 e b) PEGs com massas moleculares maiores do que 1000

é 55% solúvel em H<sub>2</sub>O [150].

A figura 30 ilustra o efeito do PEG na estabilidade do nucleotídeo. O efeito de volume de exclusão é causado tanto por PEGs(MW<1000) quanto por PEGs(MW>1000). Este efeito pode ser visualizado na figura 28b, representado pelas esferas em laranja presentes na solução. Como mencionado anteriormente, micromoléculas possuem efeito de volume de exclusão baixo, PEGs(MW<1000), enquanto as macromoléculas possuem o efeito contrário PEGs (MW>1000) [48, 141]. PEGs(MW<1000) afetam a atividade de água próximas aos nucleotídeos (figura 30a). As esferas em laranja na figura 30a representam os PEGs(MW<1000), enquanto as esferas azuis representam as moléculas de H<sub>2</sub>O próximas ao nucleotídeo. A parte polar do PEG interage via ligações de hidrogênio com as moléculas de H<sub>2</sub>O. Esse pequeno deslocamento das moléculas de H<sub>2</sub>O causado pelo PEG é responsável por desestabilizar o nucleotídeo e, conseqüentemente, reduzir a temperatura de desnaturação [143–147]. A figura 30b representa o efeito do volume de exclusão na estabilidade do nucleotídeo causado por PEG (MW>1000). Assim como na figura 30a, as esferas em cor laranja representam os PEGs, porém a dimensão da esfera é maior. Devido ao considerável volume ocupado pelo PEG na solução, apenas uma pequena parte fica disponível para o nucleotídeo. Isto dificulta a separação das fitas, e conseqüentemente, aumenta a temperatura de desnaturação [33, 35, 49, 147].

Há poucas informações disponíveis sobre o impacto do meio molecular no empilhamento e nas ligações de hidrogênio dos nucleotídeos (DNA, RNA e DNA-RNA). Os poucos estudos presentes na literatura, relacionados à termodinâmica do sistema utilizam o modelo de próximos vizinhos [143–146], que não fornece informações sobre as interações moleculares presentes nos

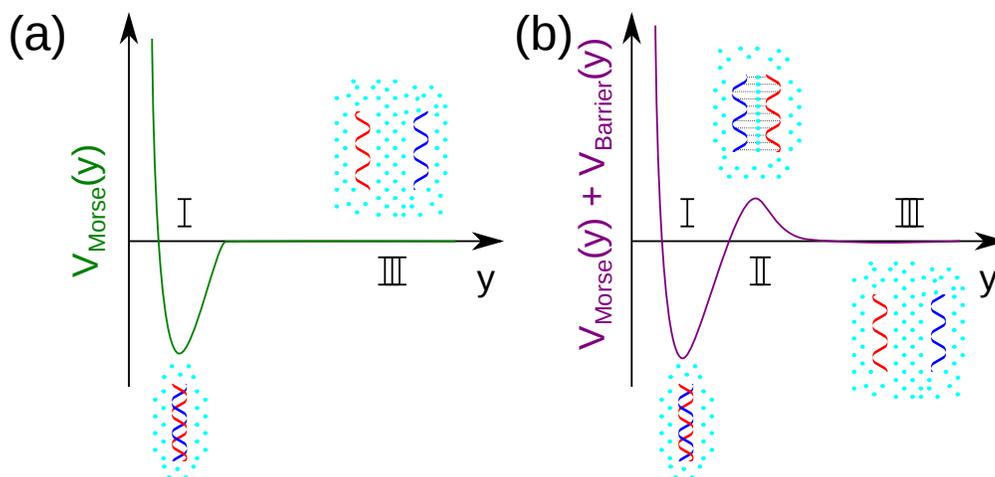
nucleotídeos. Portanto, vamos investigar a estabilidade térmica dos nucleotídeos utilizando o modelo estatístico de Peyrard-Bishop [43].

Na próxima seção, falaremos sobre outros modelos mesoscópicos baseados no modelo PB. Os modelos são conhecidos na literatura como modelos PB com barreira de potencial, cujo cada modelo recebeu um nome específico de acordo com o autor. A presença da barreira de energia após a separação das bases significa que o solvente penetra no duplexo e as bases das fitas simples passam a interagir indiretamente através do solvente. Neste cenário, temos um estado intermediário base-solvente-base antes das fitas se separarem completamente, um aspecto que não é descrito pelo modelo PB. No entanto, nenhum dos modelos de barreira foi testado com dados experimentais. Logo, nosso objetivo é avaliar a utilidade dos modelos PB com barreira para prever as temperaturas de desnaturação.

### 5.3 Modelos PB com barreira de potencial

Além do cátion, que desempenha um papel fundamental na estabilidade da estrutura do DNA, as moléculas de água também desempenham um papel crucial na manutenção da dupla hélice [41]. Essas interações influenciam tanto o empilhamento das bases quanto as ligações de hidrogênio entre as bases em fitas complementares. O modelo PB, figura 31a, assume que o DNA passa de fita dupla para fita simples, devido ao aumento de temperatura, ocorre de maneira direta. No entanto, uma análise mais detalhada revela a existência de um estado intermediário durante essa transição. Quando o duplexo começa a se desenrolar e abrir, moléculas de água começam a ocupar o espaço entre as fitas. Se a quantidade de moléculas de  $H_2O$  entre as duas fitas for baixa, o nucleotídeo da fita principal interage indiretamente e de maneira mais fraca com o nucleotídeo da fita complementar. Essa interação é intermediada principalmente pelas ligações de hidrogênio entre as moléculas de água e entre as moléculas de água e os nucleotídeos. À medida que a distância de separação entre as fitas aumenta, o valor da interação entre as moléculas também aumenta até atingir um pico máximo. A partir desse ponto, a interação começa a reduzir até que as fitas se encontrem finalmente separadas, figura 31b. Estudos de dinâmica molecular realizados por Giudice et al. [151] determinaram que as energias livres das fitas simples são bem maiores do que as do duplexo. Isso é um indicativo de que o DNA precisa superar uma barreira de potencial para passar do estado de fita dupla para fita simples. Recentemente, medidas de AFM em DNA detectaram uma pequena força entre os pares AT e CG que perdurou por um longo alcance após a quebra de sua ligação [152].

Por essa razão, Drukker et al. [153] introduziram um termo que simula a interação do nucleotídeo com o solvente para estudar a dinâmica molecular do sistema. As interações com o solvente estabilizam a fita simples de DNA por meio de ligações de hidrogênio que o solvente realiza



**Figura 31**

Comparação entre potenciais de Morse (a) sem e (b) com barreira de potencial. Os elementos I, II e III são estados do sistema. I - fita dupla, II - intermediário e III - fita simples.

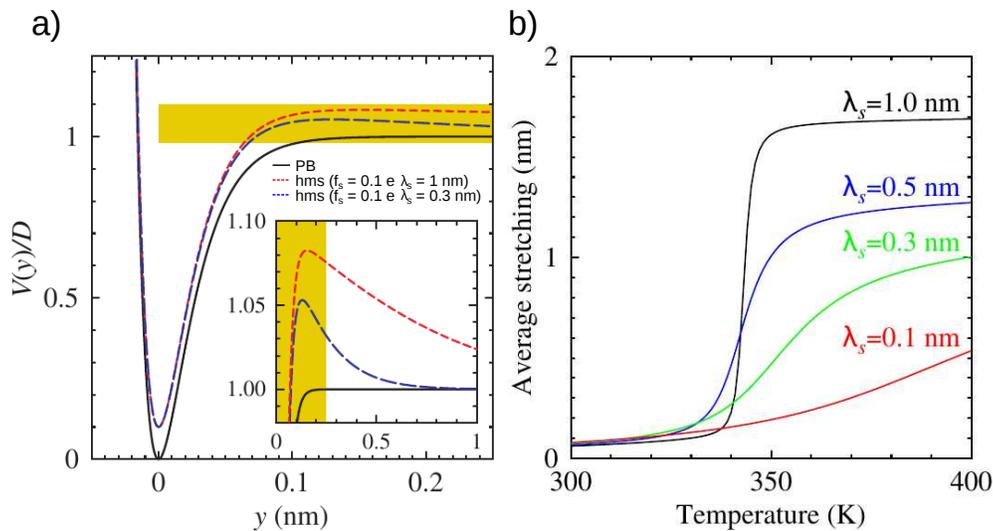
com as bases. No entanto, durante o processo de formação dos duplexos de DNA, essas ligações precisam ser desfeitas. Com o intuito de incorporar esse efeito no modelo PB, várias modificações no modelo PB convencional para descrever a interação dos nucleotídeos com o solvente foram propostas [24–27]. A seguir, apresentaremos as múltiplas variações do modelo PB documentadas na literatura e implementadas no âmbito deste projeto.

**Modelo PB com efeito de solvente (HMS)** Weber [24] adaptou o termo do solvente de Drukker et al. [153] para ser utilizado no modelo mesoscópico PB, discutido na seção 3.1. Com isso, o potencial de Morse passa a ser descrito pelo potencial Morse-solvente, e o Hamiltoniano PB, equação (9), passa a ser chamado de Hamiltoniano Morse-Solvente (HMS).

$$V(y_n) = D(e^{-\frac{y_n}{\lambda_s}} - 1)^2 - f_s D[\tanh(y_n + y_e)/\lambda_s + 1], \quad (34)$$

onde  $f_s$  é a intensidade que o solvente exerce sobre o par de base e  $\lambda_s$  é o alcance da ligação solvente-(par de base). A adição deste termo no potencial de Morse do modelo PB gera uma energia extra a ser vencida na barreira de potencial (figura 32a). Essa energia seria a ligação duplexo-solvente ( $y_n < \lambda_s$ ). No momento em que as ligações de hidrogênio são quebradas as fitas simples passam a se ligar apenas ao solvente ( $y_n > \lambda_s$ ), representado pelos picos em azul e vermelho na figura 32a. O alcance da ligação entre (fita codante)-solvente-(fita complementar) é dado pelo termo  $\lambda_s$ . Na figura 32a, observamos que quanto maior  $\lambda_s$  mais longa será a interação. O potencial de Morse (linha preta da figura 32) foi obtido ao definir o termo  $f_s$  da equação (34) como zero.

Outro fato interessante é o efeito do  $\lambda_s$  na abertura do duplexo (figura 32b). Para  $\lambda_s = 1$  a



**Figura 32**

a) Comparação entre o potencial de Morse PB (curva em preto) e o potencial de Morse-solvente - HMS (curvas em vermelho e azul) b) estiramento médio  $\langle y \rangle$  em função da temperatura para o Hamiltoniano Morse-solvente — equação (11) substituindo o potencial de Morse  $V(y_n)$  pelo potencial HMS da equação (34). Figuras retiradas de Weber [24].

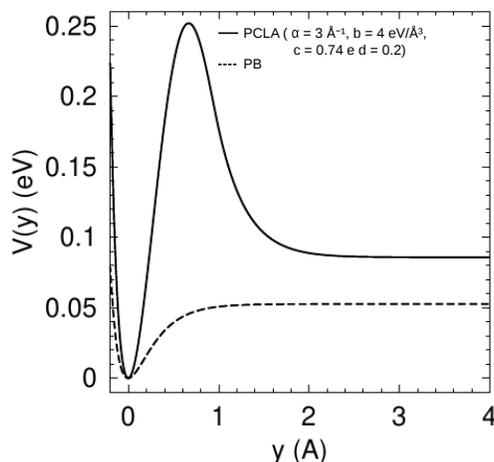
Hamiltoniano	parâmetro	unidade
$D(e^{-y_i/\lambda}-1)^2$	$D$	eV
	$\lambda$	Å
$-f_s D [\tanh(y_i + y_e)/\lambda_s + s]$	$f_s$	adimensional
	$\lambda_s$	Å
	$y_e$	Å
	$s$	adimensional
	$k$	eV/Å <sup>2</sup>
$\frac{k}{2}(y_i^2 - 2y_i y_{i-1} \cos\theta + y_{i-1}^2)$	$\theta$	rad

**Tabela 5**

Parâmetros para o modelo HMS.

transição de fita dupla para fita simples é abrupta, enquanto para valores menores do que 1, essa transição ocorre de maneira mais suave. Portanto,  $\lambda_s$  pode ser utilizado como uma ferramenta teórica para modificar o perfil da curva de desnaturação.

Weber [24] substituiu o valor +1, apresentado na equação (34) pelo parâmetro  $s$ . Este termo possui a função de reduzir ou elevar a profundidade do poço de potencial  $D$ . Entre os artigos que tratam deste tipo de potencial encontramos dois valores de  $s$  para o modelo HMS utilizados para estudos teóricos. Os valores de  $s$  são  $s=+1$  [24] e  $s=-1$  [153, 154]. Exploraremos estes dois casos no capítulo de resultados 5. Na tabela 5 listamos os parâmetros do potencial HMS e suas unidades.



**Figura 33**

Comparação entre o potencial de Morse PB (linha tracejada em preto) e o potencial de Morse com barreira entrópica — PCLA (linha em preto). Os parâmetros dos potenciais foram escolhidos para reproduzir a mesma temperatura de desnaturação de uma sequência poly(A). Figura retirada de Peyrard et al. [25].

**Modelo PB com barreira entrópica (PCLA)** Peyrard, Cuesta-Lopez e Angelov (PCLA) [25] modificaram o modelo PB ao acrescentar um termo de barreira entrópica ao potencial de Morse, equação (35). Esse termo possui o mesmo significado físico do termo Morse-solvente introduzido por Weber [24], o qual descreve a interação do nucleotídeo com o solvente.

$$V(y_n) = D(e^{-\frac{y_n}{\lambda}} - 1)^2 + \Theta(y_n) \frac{by_n^3}{\cosh^2[c(\alpha y_n - d \ln 2)]}, \quad (35)$$

onde  $\Theta(y)$  é uma função degrau que assegura que a equação (35) só participa efetivamente do potencial de Morse para  $y > 0$ , quando as bases começam a se separar. O parâmetro  $b$  determina a altura da barreira de potencial,  $c$  a largura e  $d$  a posição da barreira.

Na figura 33, observam-se as diferenças entre os dois potenciais. A diferença primordial entre os dois potenciais é a barreira de potencial gerada pela adição da equação (35) ao potencial de Morse. O potencial de Morse (linha tracejada figura 33) foi gerado fazendo o termo  $b$  da equação (35) igual à zero. Para reproduzir os resultados experimentais de uma sequência poly(A) para ambos os modelos, a escolha de parâmetros do modelo PCLA resultou em uma gigante barreira de potencial (linha em preto da figura 33). No entanto, temos que ter em mente que um par de base no estado de fita simples é puxado para o estado de fita dupla também pelo potencial de empilhamento entre os seus próximos vizinhos. Portanto, o potencial efetivo de fechamento que a fita simples experimental é muito menor.

Apesar dos autores definirem o significado físico e as unidades de  $b$  e  $c$  como eV e  $\text{Å}^{-1}$ , respectivamente [25], notamos que, a partir da equação (35), não seria possível existir tais unidades.

Hamiltoniano	parâmetro	unidade
$D(e^{-y_i/\lambda}-1)^2$	$D$	eV
	$\lambda$	Å
	$b$	eV/Å <sup>3</sup>
$by_i^3/\cosh^2[c(\alpha y_i - d \ln 2)]$	$c$	adimensional
	$d$	adimensional
	$\alpha$	Å <sup>-1</sup>
$\frac{k}{2}(y_i^2 - 2y_i y_{i-1} \cos \theta + y_{i-1}^2)$	$k$	eV/Å <sup>2</sup>
	$\theta$	rad

**Tabela 6**

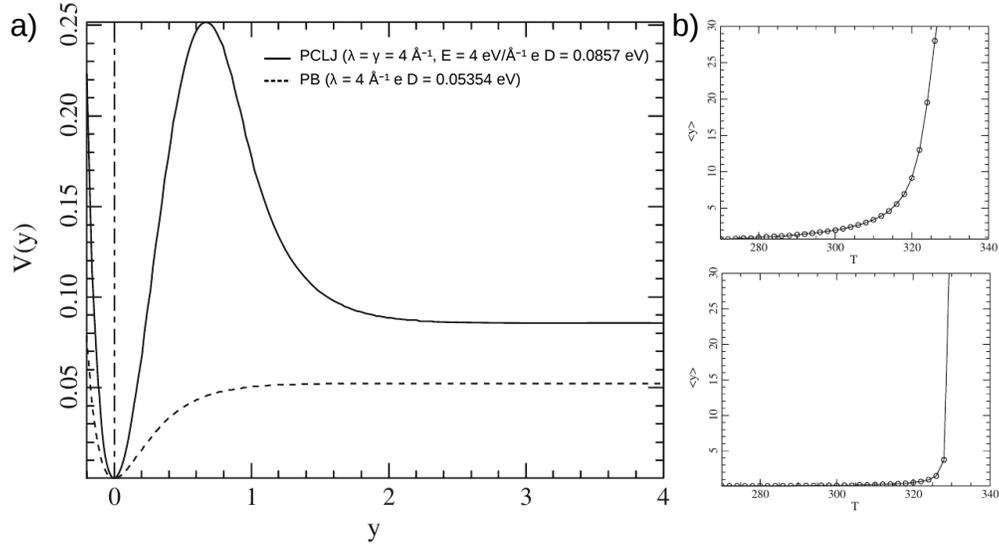
Parâmetros para o modelo PCLA com unidades redefinidas.

Na tabela 6, listamos os parâmetros do potencial PCLA e suas respectivas unidades.

**Modelo PB com potencial de hump (PCLJ)** Peyrard, Cuesta-Lopez e James (PCLJ) [26] modificaram o potencial de Morse pelo potencial de 'hump'. O nome 'hump' (corcova) vem de um apelido dado a literatura ao modelo devido ao formato da barreira de potencial. Simulações por dinâmica molecular não conseguem estudar o tempo de vida do estado de fita simples por causa do número de graus de liberdade associados aos pares de base e das bases com solvente. Entretanto, as simulações podem ser feitas para observar o caminho da energia livre associado à abertura da fita. Isso é realizado adicionando uma restrição geométrica para uma dada abertura. Logo, as flutuações da estrutura do DNA são registradas para todas as posições intermediárias e sua distribuição de probabilidades fornece as energias livres como funções das aberturas [151]. Os resultados são altamente sensíveis a dinâmica do solvente e do sal, no entanto, eles mostram que o estado aberto deve possuir um pequeno valor de energia. Isso é consistente com a dinâmica do sistema em que o par de base deve permanecer aberto por um período mais longo, em vez de vibrar como um *breather*<sup>2</sup>.

Nos modelos mesoscópicos, que descrevem apenas alguns graus de liberdade do sistema, os potenciais são na verdade energias livre que incluem todos os graus de liberdade que não constam no modelo. Portanto, é natural substituir o potencial de Morse por uma função que leva em conta os efeitos discutidos anteriormente. Na equação (36), apresentamos o potencial de 'hump' que

<sup>2</sup>DNA *breathing* (ou *fraying*): flutuações conformacionais locais espontâneas no DNA dupla hélice. Essas flutuações levam a separação dos pares de base em temperaturas abaixo da temperatura de desnaturação expondo essas bases ao solvente.



**Figura 34**

a) Comparação entre o potencial de Morse PB (linha tracejada em preto) e o potencial de Morse com potencial 'hump' — PCLJ (linha em preto). Os parâmetros dos potenciais foram escolhidos para reproduzir a mesma temperatura de desnaturação. b) Deslocamento médio dos pares de base no DNA descritos pelo modelo PB, figura de cima, e pelo modelo PB modificado com o potencial de 'hump', figura de baixo. Figuras retiradas de Peyrard et al. [26].

renomeamos para PCLJ.

$$V_h(y_n) = \begin{cases} A(e^{\frac{-y_n}{\lambda_h}} - 1)^2 & y < 0 \\ ay^2 + by^3 + cy^4 & 0 \leq y \leq 1 \\ D_h + Ee^{-y/\gamma}(y + \gamma) & y > 1, \end{cases} \quad (36)$$

A primeira função da equação (36) é um potencial de Morse, a segunda é um polinômio de grau quatro para assegurar uma transição suave entre a primeira e terceira equações e a terceira equação descreve um decaimento exponencial que assegura que a função chegue ao estado de energia raso,  $D_h$ , quando a fita se encontra totalmente aberta.  $D_h$ ,  $E$ ,  $\lambda_h$  e  $\gamma$  são os parâmetros a serem inseridos no sistema, enquanto os demais parâmetros são calculados para assegurar a continuidade do potencial e suas primeira e segunda derivadas em  $y = 0$  e  $y = 1$  Å, equação (37). Aplicando as condições, temos,

$$\begin{aligned} A &= a\lambda_h^2 \\ a &= 6D_h + \frac{1}{2}Ee^{-1/\gamma}(\gamma^{-2} + 5\gamma^{-1} + 12\gamma + 12) \\ b &= -8D_h - Ee^{-1/\gamma}(\gamma^{-2} + 4\gamma^{-1} + 8\gamma + 8) \\ c &= 3D_h + \frac{1}{2}Ee^{-1/\gamma}(\gamma^{-2} + 3\gamma^{-1} + 6\gamma + 6) \end{aligned} \quad (37)$$

Hamiltoniano	parâmetro	unidade
$A(e^{-y_i/\lambda_h - 1})^2$	$A$	eV - calculado pela equação (37)
	$\lambda_h$	Å
$ay_i^2 + by_i^3 + cy_i^4$	$a$	eV/Å <sup>2</sup> - calculado pela equação (37)
	$b$	eV/Å <sup>3</sup> - calculado pela equação (37)
	$c$	eV/Å <sup>4</sup> - calculado pela equação (37)
$D_h + Ee^{-y_i/\gamma}(y_i + \gamma)$	$D_h$	eV
	$E$	eV/Å
	$\gamma$	Å
$\frac{k}{2}(y_i^2 - 2y_i y_{i-1} \cos\theta + y_{i-1}^2)$	$k$	eV/Å <sup>2</sup>
	$\theta$	rad

**Tabela 7**

Parâmetros para o modelo PCLJ.

A figura 34a compara o potencial de Morse com o potencial de 'hump'. O potencial de 'hump' possui uma barreira de potencial muito grande. Isso acontece por causa das condições que os autores colocaram na temperatura de desnaturação, mas isso é consistente com a física do DNA, porque as bases são hidrofóbicas. Ou seja, o interior do duplexo de DNA não interage com solvente no estado de fita dupla. A partir do momento em que a fita começa a se romper, há um custo energético da interação da fita simples com o solvente e o potencial de 'hump' leva isso em conta, descrito pelo terceiro termo da equação (36).

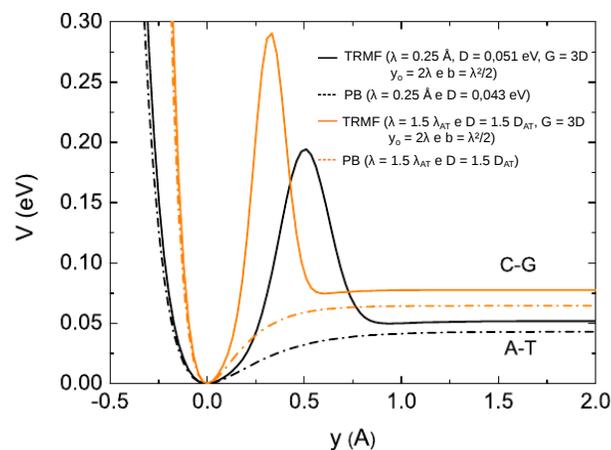
As temperaturas de desnaturação para os dois modelos são mostradas na figura 34b. A desnaturação é mais fina com o potencial de 'hump' (figura 34a abaixo) do que com o potencial de Morse (figura 34a acima). Isso é condizente com homopolímeros de DNA. Na tabela 7, listamos os parâmetros do potencial PCLJ e suas unidades.

**Modelo PB com potencial gaussiano (TRMF)** Tapia-Rojo, Mazo, and Falo (TRMF) [27] introduziram um potencial gaussiano ao potencial de Morse,

$$V(y_n) = D(e^{-\frac{y_n}{\lambda}} - 1)^2 + Ge^{-(y_n - y_o)^2/b_G}, \quad (38)$$

onde,  $G$  é a altura da barreira,  $y_o$  é a posição da barreira e  $b_G$  descreve a largura da barreira. A figura 35 compara o modelo com o potencial gaussiano com o potencial de Morse. O valor da altura da barreira de CG é maior do que AT. Isso é devido a escolha dos valores dos parâmetros e, também, do fato de que o par CG possui mais sítios de ligação de hidrogênio do que o par AT. Como nos dois modelos anteriores PCLA e PCLJ a barreira de potencial é bem grande. Porém, devemos nos lembrar que o solvente influencia o empilhamento entre as bases. Portanto, o potencial efetivo deve ser bem menor. Na tabela 8 listamos os parâmetros do potencial TRMF e suas unidades

Os modelos mencionados acima foram utilizados apenas em cálculos teóricos [24–27,



**Figura 35**

Comparação entre o potencial de Morse PB (linha tracejada) e o potencial de Morse com potencial gaussiano — TRMF (linha sólida). Os parâmetros dos potenciais foram escolhidos para reproduzir a mesma temperatura de desnaturação. Figura retirada de Tapia-Rojo et al. [27].

Hamiltoniano	parâmetro	unidade
$D(e^{-y_i/\lambda}-1)^2$	$D$	eV
	$\lambda$	Å
	$G$	eV
$Ge^{-(y_i-y_o)^2/b_G}$	$y_o$	Å
	$b_G$	Å
	$k$	eV/Å <sup>2</sup>
$\frac{k}{2}(y_i^2 - 2y_i y_{i-1} \cos\theta + y_{i-1}^2)$	$\theta$	rad

**Tabela 8**

Parâmetros para o modelo TRMF.

153–158]. Será que algum dos modelos, ou todos eles, conseguem determinar temperaturas de desnaturação melhores que o modelo PB? Em caso afirmativo, significa que teremos um modelo superior para prever temperaturas de desnaturação, capaz de descrever um estado do sistema não explicado pelo modelo PB.

Nas seções seguintes, apresentaremos toda a metodologia empregada no estudo de nucleotídeos em ambientes que simulam o meio molecular. Como se deu a coleta de dados, os procedimentos de minimização do modelo PB e dos modelos PB com barreira. Em seguida, discutiremos os resultados obtidos pelo modelo PB e avaliaremos se os modelos com barreira de potencial são melhores do que o modelo original na predição de temperaturas. Por último, apresentaremos nossas conclusões e perspectivas futuras.

## 5.4 Coleta e preparação dos dados

**Notação** Usamos a notação usual para os pares canônicos do DNA e RNA apresentada no capítulo 2. As ligações de hidrogênio serão representados pelo par participante na ligação, por exemplo o par CG. Os empilhamentos entre as bases serão representados por um (par de base 1)-(par de base 2). Como exemplo, utilizaremos o par CG e o par AT, correspondente ao dímero 5'-CA-3'/3'-GT-5' será representado por CG-AT. Devido a condições de simetria o dímero CG-AT é equivalente ao dímero TA-GC. Utilizaremos este fato para reduzir o número de parâmetros analisados. Fazemos isso quando o número de sequências utilizados nas análises é relativamente pequeno.

**Dados de temperatura de desnaturação** Extraímos sequências de DNA e RNA em PEG da literatura [143–147], e primeiramente, analisamos os tipos de primeiros vizinhos presentes nessas sequências. Com isso, identificamos os parâmetros de *stacking*,  $k$ , para otimizarmos, além claro, do potencial de Morse,  $D$ , entre as bases nitrogenadas. A tabela 9 apresenta a quantidade de sequências extraídas da literatura e os parâmetros a serem estudados. Além das sequências, obtivemos dos artigos a concentração de duplexos ( $C_t$ ), a temperatura de desnaturação ( $T_m$ ) e a concentração de sal ( $[Na^+]$ ). As sequências de DNA e RNA extraídas da literatura se encontram sob as mesmas condições experimentais. Desse modo, não foi necessário ajustar a concentração de sal ou de duplexos nessas sequências.

Encontramos outras proporções e tamanhos de PEG na literatura para DNA e RNA [33–35, 48, 49, 140–142, 159–162], porém o número de sequências é bem pequeno. A maioria das sequências encontradas na literatura constitui um conjunto menor do que 10 sequências. Conseqüentemente, a frequência em que esses parâmetros aparecem na base de dados é muito baixa, podendo até mesmo ser zero. Dessa forma, no momento, não é viável trabalhar com DNA ou RNA em outras proporções

Dados	$N$	Morse ( $D$ )	empilhamento ( $k$ )				
DNA-P200/40	40	AT(164)	AT-AT(20)	AT-CG(44)	AT-GC(40)	AT-TA(30)	CG-AT(45)
		CG(254)	CG-CG(45)	CG-GC(62)	GC-AT(49)	GC-CG(58)	TA-AT(25)
RNA-P200/20	35	AU(119)	AU-AU(14)	AU-CG(20)	AU-GC(25)	AU-UA(30)	CG-AU(20)
		CG(127)	CG-CG(18)	CG-GC(32)	GC-AU(25)	GC-CG(32)	UA-AU(30)
RNA-P200/40	45	AU(178)	AU-AU(31)	AU-CG(36)	AU-GC(45)	AU-UA(33)	CG-AU(46)
		CG(207)	CG-CG(36)	CG-GC(40)	GC-AU(43)	GC-CG(46)	UA-AU(29)

**Tabela 9**

Resumo dos pares de base e próximos vizinhos associados aos parâmetros analisados. Na coluna  $N$ , apresentamos o número total de sequências extraídas da literatura de cada base de dados. O número  $n$  entre parênteses, nas colunas Morse e empilhamento, informa a quantidade de vezes que cada parâmetro aparece em nossa base de dados.

ou tamanhos de PEG. Portanto, concentraremos nossa análise no PEG200 nas concentrações de 20%, apenas RNA [146, 147], e 40%, DNA [143, 144] e RNA [145], do volume em água.

Utilizamos a nomenclatura DNA-P200/40, RNA-P200/20 e RNA-P200/40 para separar os conjuntos de dados com PEG200. DNA e RNA se referem ao tipo de nucleotídeo utilizado, P200 é uma abreviação de PEG200, e 40 ou 20 denotam a concentração de PEG200 utilizada para medir as temperaturas de desnaturação. Ao compararmos as condições experimentais entre os conjuntos de dados DNA-P200/40, RNA-P200/20 e RNA-P200/40, verificamos que eles estão quase sob às mesmas condições experimentais. Os conjuntos DNA-P200/40 e RNA-P200/40 possuem as mesmas concentrações de sal ( $[Na^+]$ ), de duplexo ( $C_t$ ), pH e concentração de PEG. No entanto, a  $[Na^+]$  e concentração de PEG das sequências RNA-P200/20 é diferente das dos conjuntos DNA-P200/40 e RNA-P200/40. A concentração de  $[Na^+]$  é  $10\times$  maior, enquanto a concentração de PEG é 20% do volume de água.

Não dispomos das mesmas sequências de DNA e RNA na ausência de PEG200 para serem nossa referência. O modelo de próximos vizinhos, discutido na seção 3.2 do capítulo 3, possui parâmetros de entalpia e entropia bem estabelecidos em  $H_2O$  [30, 163]. A partir destes parâmetros, podemos prever quais às temperaturas das sequências dos conjuntos DNA-P200/40, RNA-P200/20 e RNA-P200/40 na ausência de PEG200. O cálculo se baseia em somar as contribuições de entalpia e entropia dos próximos vizinhos que compõem cada sequência presente na base de dados. Os próximos vizinhos, ou dímeros, são os mesmos apresentados na tabela 9, onde o  $k$  é substituído por  $\Delta H$  ou  $\Delta S$ . Ao somar todas as contribuições de energia, obtemos valores de entalpia e entropia totais únicas para cada sequência. Utilizando a equação (8), calculamos às temperaturas de desnaturação na concentração  $C_t$  desejada, que neste caso é de 100 mM [143–147]. Esse mesmo tipo de análise já foi utilizada por outros autores na literatura [143–145]. Nomeamos o conjunto de dados de temperaturas de desnaturação calculadas em  $H_2O$  como DNA-W/40, RNA-W/20, RNA-W/40. Diferenciamos

o conjunto controle (DNA-W/40, RNA-W/20, RNA-W/40), temperaturas calculadas em H<sub>2</sub>O, do conjunto de temperaturas medidas em PEG200 (DNA-P200/40, RNA-P200/20 e RNA-P200/40) apenas substituindo P200 por W, onde W = *water* que é água em inglês. As temperaturas calculadas utilizando o modelo de próximos vizinhos estão disponíveis nas tabelas 23, 24 e 25. Ao todo, analisamos, com os modelos mesoscópicos, seis conjuntos de dados chamados de DNA-W/40, RNA-W/20, RNA-W/40, DNA-P200/40, RNA-P200/20 e RNA-P200/40.

Dividimos os resultados dos modelos em duas partes: Modelo Peyrard-Bishop (PB) e Modelos Peyrard-Bishop (PB) com barreira de potencial. Na primeira parte, apresentaremos os resultados do modelo PB padrão e na segunda parte, apresentaremos os resultados dos modelos PB com barreira de potencial. O motivo principal da divisão é que os modelos PB com barreira de potencial nunca foram analisados com dados experimentais; há apenas discussões teóricas [24–27, 153–158]. Esta é a primeira vez que este tipo de análise foi realizada. Logo, não existem parâmetros de referência iniciais para as minimizações dos modelos PB com barreira. Então, precisamos primeiro identificar esses parâmetros iniciais. Essa abordagem não é necessária para o modelo PB padrão, que já está bem estabelecido para prever temperaturas de desnaturação, e nosso grupo já possui parâmetros iniciais determinados para várias concentrações de sal diferentes [28–30, 163, 164].

## 5.5 Modelo Peyrard-Bishop (PB)

### 5.5.1 Procedimento de otimização

Os procedimentos de otimização para o modelo PB foram similares aos descritos na seção 3.1.3. Como valores iniciais, utilizamos os das referências Weber et al. [28], Ferreira et al. [30] mostrados na tabela 10. Os conjuntos DNA-W/40, DNA-P200/40, RNA-W/20, RNA-P200/20, RNA-W/40 e RNA-P200/40 foram otimizados separadamente. Foram realizadas quatro rodadas de minimizações. Cada rodada de minimização foi realizada 1000 vezes, pois às minimizações estão sujeitas a um grande número de mínimos locais. Para evitar este problema, o cálculo foi repetido várias vezes. O número 1000 é suficientemente grande para termos um bom valor médio e dispersão para cada parâmetro minimizado do modelo. As três primeiras otimizações foram feitas para confirmar se realmente nos encontramos na região de mínimo da função  $\chi^2$ . Isso é observado quando os valores de  $\chi^2$  entre as minimizações variam pouco. Após confirmarmos que nos encontramos na região do mínimo, uma quarta e última rodada de minimização foi realizada para estimar a influência do erro experimental nos novos parâmetros, seguindo uma abordagem semelhante à descrita na seção 3.1.3. Para essa estimativa, consideramos uma incerteza de 0.7 °C, obtida das médias de cada incerteza para os valores de temperatura de desnaturação das sequências extraídas

Parâmetro	DNA-W/40	Parâmetro	RNA-W/20	RNA-W/40
	DNA-P200/40		RNA-P200/20	RNA-P200/40
$D_{AT}$	31.96 meV	$D_{AU}$	38.70 meV	36.77 meV
$D_{CG}$	73.26 meV	$D_{CG}$	67.14 meV	75.61 meV
$k_{AT-AT}$	2.22 eV/nm <sup>2</sup>	$k_{AU-AU}$	1.91 eV/nm <sup>2</sup>	2.08 eV/nm <sup>2</sup>
$k_{AT-CG}$	2.58 eV/nm <sup>2</sup>	$k_{AU-CG}$	3.04 eV/nm <sup>2</sup>	3.39 eV/nm <sup>2</sup>
$k_{AT-GC}$	2.31 eV/nm <sup>2</sup>	$k_{AU-GC}$	2.51 eV/nm <sup>2</sup>	2.43 eV/nm <sup>2</sup>
$k_{AT-TA}$	1.60 eV/nm <sup>2</sup>	$k_{AU-UA}$	2.21 eV/nm <sup>2</sup>	0.87 eV/nm <sup>2</sup>
$k_{CG-AT}$	3.52 eV/nm <sup>2</sup>	$k_{CG-AU}$	2.44 eV/nm <sup>2</sup>	2.84 eV/nm <sup>2</sup>
$k_{CG-CG}$	2.06 eV/nm <sup>2</sup>	$k_{CG-CG}$	2.77 eV/nm <sup>2</sup>	2.51 eV/nm <sup>2</sup>
$k_{CG-GC}$	2.75 eV/nm <sup>2</sup>	$k_{CG-GC}$	1.33 eV/nm <sup>2</sup>	1.24 eV/nm <sup>2</sup>
$k_{GC-AT}$	2.80 eV/nm <sup>2</sup>	$k_{GC-AU}$	3.27 eV/nm <sup>2</sup>	4.15 eV/nm <sup>2</sup>
$k_{GC-CG}$	3.37 eV/nm <sup>2</sup>	$k_{GC-CG}$	3.12 eV/nm <sup>2</sup>	3.03 eV/nm <sup>2</sup>
$k_{TA-AT}$	2.55 eV/nm <sup>2</sup>	$k_{UA-AU}$	2.99 eV/nm <sup>2</sup>	5.78 eV/nm <sup>2</sup>

**Tabela 10**

Parâmetros  $k$  e  $D$  iniciais utilizados para as minimizações do modelo PB.

Conjunto de dados	regressão por tamanho da sequência			
	antes da primeira minimização		depois da última minimização	
	$\langle \Delta T \rangle$ (°C)	$\langle \chi^2 \rangle$ (°C <sup>2</sup> )	$\langle \Delta T \rangle$ (°C)	$\langle \chi^2 \rangle$ (°C <sup>2</sup> )
DNA-W/40	2.92	8.55	1.69	4.15
DNA-P200/40	2.32	10.26	0.77	1.35
RNA-W/20	1.05	1.66	0.41	0.26
RNA-P200/20	1.31	2.34	0.92	1.21
RNA-W/40	1.02	1.51	0.30	0.15
RNA-P200/40	1.65	4.46	0.79	1.02

**Tabela 11**

Comparação entre os  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$  antes da primeira e após a última minimização na presença e ausência de PEG200. Para ficar mais fácil de compararmos os resultados, normalizamos o valor de  $\chi^2$  dividindo-o pelo número de sequências  $N$  mostrados na tabela 9.

dos seus respectivos artigos [143–147].

## 5.5.2 Resultados

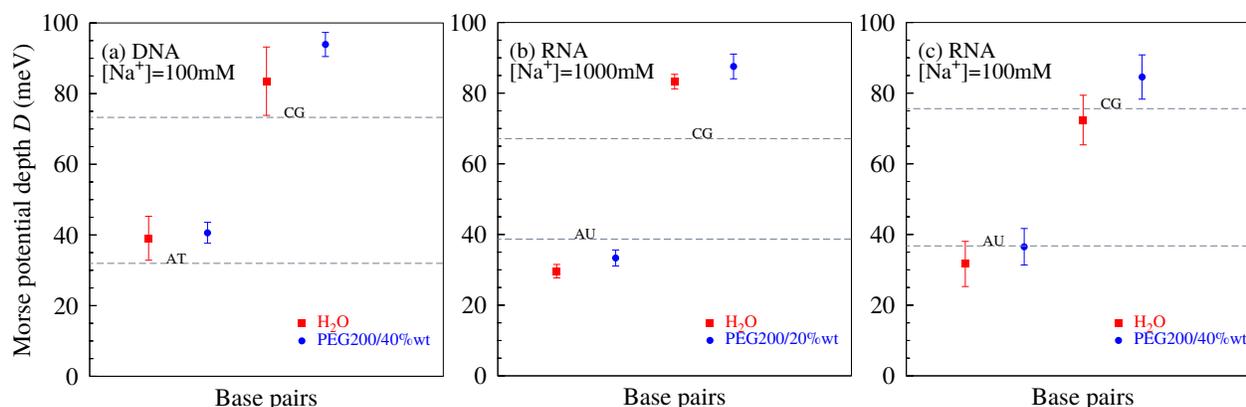
Foram necessárias três rodadas de minimizações, além de uma última rodada para avaliar a influência do erro experimental para as 6 bases de dados apresentadas na tabela 9. Na tabela 11, apresentamos os parâmetros de qualidade  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$  antes da primeira rodada de minimizações e após a quarta e última rodada. As diferenças de temperaturas médias,  $\langle \Delta T \rangle$ , se encontram abaixo de 2 °C e de  $\langle \chi^2 \rangle$  abaixo de 4.15 °C<sup>2</sup>, indicando que as temperaturas calculadas estão bem próximas das temperaturas experimentais. O conjunto de dados RNA-P200/40 foi o que mais se aproximou das temperaturas experimentais.

Nas figuras 36, apresentamos os valores de profundidades de poço  $D$  para os seis conjuntos.

Somente na figura 36b temos uma diferença bem grande entre os valores de Morse  $D$  de sequências em  $H_2O$ , pontos em vermelho, de estudos com outras sequências sob as mesmas condições experimentais [28–30], linhas tracejadas. Apesar disso, os resultados encontrados foram bastante interessantes. Em todos os casos, observamos um aumento do valor médio de  $D$  quando adicionamos PEG200 às soluções. Recentemente, por meio de um estudo conduzido pelo nosso grupo, Muniz et al. [165] constataram que a variação da concentração de  $[Na^+]$  em soluções aquosas não resultou em nenhuma alteração significativa no valor médio de  $D$ . As moléculas de água agem no par de base reduzindo o valor da ligação de hidrogênio entre os nucleotídeos. Supondo que o par de base se encontre no vácuo, sem a influência de outras moléculas, o par de base CG possui uma energia maior do que o par AT, ou AU para RNA. Quando formamos um duplexo de DNA ou RNA com diferentes combinações de nucleotídeos, como AT e CG no DNA, a hélice de DNA torna-se mais distorcida devido às forças desiguais entre os pares CG e AT. Ao adicionarmos água ao sistema, observamos que as moléculas de água atuam de forma a desestabilizar a ligação de hidrogênio entre os nucleotídeos [70]. Se um dos pares de base for mais hidratado, isso pode levar a duas possibilidades distintas:

1. CG ser *mais* hidratado do que AT, ou AU no caso do RNA;
2. CG ser *menos* hidratado do que AT, ou AU no caso do RNA.

No caso 1. teremos uma redução maior entre a ligação de hidrogênio do par CG, enquanto que no caso 2. observaremos o contrário. Pelo caso 1. esperamos que as moléculas de água atuem de forma a deixar as ligações de hidrogênio dos pares AT e CG com diferenças menores. A consequência direta disso é que como as energias de interação entre os pares de base AT, ou AU, e CG estão mais próximas a hélice apresentará menos distorções tendendo a uma conformação mais regular. No caso 2. observaremos um efeito contrário. Os resultados apresentados nas figuras 36 estão de acordo com o caso 1. O PEG200 interfere nas interações das moléculas de  $H_2O$  com os nucleotídeos, deslocando o sistema para o estado de fita simples [33, 48, 49, 140, 141]. Menos moléculas de água atuando como contrapeso nas ligações de hidrogênio significam que as energias das ligações entre os nucleotídeos se tornam mais intensas. Nossos resultados mostram que o par CG é o mais afetado na presença de PEG200/40%wt, figuras 36a e 36c. Portanto, CG é mais hidratado tanto em DNA quanto em RNA. No RNA, a diferença entre a profundidade do poço  $D$  do par CG em  $H_2O$  do caso em PEG200/40%wt é um pouco maior do que o DNA. Ao compararmos AU e AT observamos que PEG200/40%wt afeta mais o par AU, ou seja, AU é mais hidratado que o par AT. Pelas figuras 36a e 36c, podemos inferir uma ordenação de quais pares de bases são os mais hidratados e afirmar qual tipo de nucleotídeo, DNA ou RNA, é mais hidratado. De acordo com nossos resultados, o RNA é mais hidratado do que o DNA. Nossos resultados estão em linha com a literatura que afirma que o

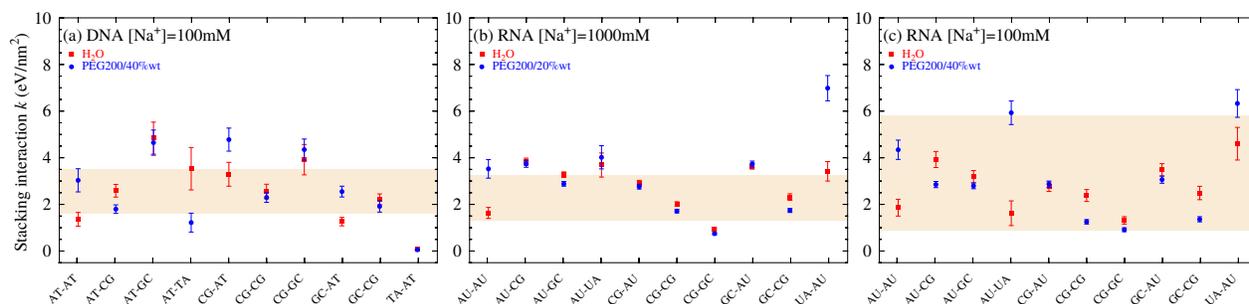


**Figura 36**

Profundidades médias de Morse  $D$  calculadas para DNA e RNA em H<sub>2</sub>O como pontos vermelhos e em PEG200 como pontos azuis. Para comparação mostramos os parâmetros  $D_{AT}$ , ou  $D_{AU}$ , (linha inferior) e  $D_{CG}$  (linha superior) de outras publicações do nosso grupo em H<sub>2</sub>O sob as mesmas concentrações de [Na<sup>+</sup>] [28–30] como linhas tracejadas em cinza. (a) DNA-W/40 e DNA-P200/40 (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40.

RNA é realmente mais hidratado do que o DNA [4, 34] e o par CG é mais hidratado do que o par AT, ou AU no caso de RNA [3, 4, 65]. Outro fato interessante é que as diferenças de profundidades  $D_{AU}$  e  $D_{CG}$  entre H<sub>2</sub>O e PEG200/40%wt, ou PEG200/20%wt, são bem próximas. Para o DNA, é nítido que  $D_{CG}$  apresenta uma variação bem maior do que  $D_{AT}$ . Isso ocorre devido às diferenças de energia entre H<sub>2</sub>O e PEG200/40%wt que são bem distintas entre AT e CG no DNA. O DNA, nesse caso, pode apresentar uma conformação diferente do que apresenta em soluções contendo apenas H<sub>2</sub>O. Ou seja, o DNA estaria se transformando de uma hélice do tipo B para uma hélice do tipo A. Outro detalhe importante é que os valores de AT são praticamente iguais tanto em H<sub>2</sub>O quanto em PEG200. Logo, sequências puramente AT sempre apresentam hélices do tipo B, conforme mencionado em Hunter [62]. Por causa da diferença entre as concentrações de [Na<sup>+</sup>] entre os conjuntos de RNA, figuras 36b e 36c, não é possível atribuir que as diferenças entre os dois gráficos seja puramente do PEG200.

Nas figuras 37, apresentamos os parâmetros de empilhamento  $k$  dos 10 dímeros possíveis no sistema. Poucos pontos em vermelho saíram da região sombreada que representa o valor máximo e mínimo dos parâmetros de empilhamento  $k$  encontrados em estudos anteriores de DNA e RNA na H<sub>2</sub>O sob condições experimentais similares [28–30]. Apenas o dímero TA-AT, figura 37a, se encontra muito distante da região sombreada. TA-AT aparece em 14 das 40 sequências utilizadas no estudo de DNA em ambientes que simulam o meio molecular, tabela 23. Destas sequências três nos chamaram a atenção: **CGGTACCG**, **CCGTACGG** e **CGCGTACGCGTACGCG**. Observe que há muitos pares CG nas três sequências. Devido ao seu ordenamento na estrutura, dão origem à todas as combinações de dímeros CG possíveis. A última sequência apresenta uma temperatura de desnaturação muito



**Figura 37**

Interações de empilhamento médias  $k$  calculadas para DNA e RNA em  $H_2O$  como pontos vermelhos e em PEG200 como pontos azuis. A região sombreada representa o intervalo de parâmetros de empilhamento de outras publicações do nosso grupo em  $H_2O$  sob as mesmas concentrações de  $[Na^+]$  [28–30]. (a) DNA-W/40 e DNA-P200/40 (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40

maior do que as duas primeiras, tanto em PEG200 quanto em  $H_2O$ , principalmente por causa de seu tamanho. Como existem vários tipos de dímeros formados pelo par CG nestas sequências e apenas um ponto, ou dois na última sequência com o dímero TA-AT, os cálculos de minimizações consideraram que  $k_{TA-AT}$  é aproximadamente zero. Isso acontece, pois o modelo se baseia que o par CG é o principal responsável pela transição de fita dupla para simples, CG possui mais ligações de hidrogênio que AT.

O efeito de PEG200 nos três casos, pontos em azul nas figuras 37, varia bastante. O deslocamento das moléculas de água enrijece ou amolece alguns dos empilhamentos  $k$ . Mas há casos em que há pouco ou nenhum efeito por conta da adição de PEG200. Nas figuras 37a e 37c os empilhamentos que mais sofreram variação com adição de PEG200 foram os dímeros AT-AT, AU-AU análogo no RNA, e AT-TA, AU-UA análogo no RNA. Enquanto, AT-AT, AU-AU e AU-UA ficaram mais rígidos, AT-TA se tornou mais flexível. Hélices de DNA ou RNA que possuem uma alta frequência de AT-AT, AU-AU e AU-UA se tornam mais difíceis de se desnaturar dificultando a atuação de enzimas. Na figura 37b observamos que AU-UA e UA-AU são os dímeros que mais variaram com a adição de PEG200. Apenas 20% de PEG200 foram suficientes para aumentar o valor do empilhamento em relação  $H_2O$ . A concentração de  $[Na^+]$  10 vezes maior do que no caso da figura 37c parece ter um papel importante em manter os demais empilhamentos próximos aos valores em solução com apenas  $H_2O$ . Entre os dímeros, devem existir muitas moléculas de  $Na^+$  solvatadas pela  $H_2O$ . A ligação iônica entre  $Na^+$  e  $H_2O$  dificulta a atuação da pequena concentração de PEG200, apenas 20% do volume da solução, nas moléculas de água. Nossos resultados mostram que há poucas moléculas de  $Na^+$  nos dímeros AU-UA e UA-AU, mais afetados pela adição de PEG200.

O modelo PB explica muito bem a dinâmica de nucleotídeos em soluções que simulam o meio molecular. A análise dos resultados fornece informações valiosas sobre as interações inter-

moleculares presentes nos nucleotídeos. No DNA, observamos uma diferença significativa na mudança de energia entre os pares AT e CG após a adição de PEG200, o que não ocorre no RNA. Essa diferença de energia entre os pares AT e CG é o responsável pelo aparecimento de distorções na dupla hélice, consequentemente a hélice passa de uma conformação total tipo B para a conformação tipo A. Nossos resultados concordam com o que existe na literatura sobre hidratação [3, 4] e experimentos de UV *melting* de nucleotídeos em soluções contendo PEG200 [35, 49, 140, 143–147]. Seria interessante fazer este estudo para as mesmas sequências em outras concentrações de PEG200, entre 0%-40%, sob as mesmas condições experimentais. Isso nos permitiria determinar a atividade da água nas sequências analisadas. Além disso, teríamos um mecanismo para prever temperaturas de desnaturação de outras sequências em diferentes concentrações de PEG200. A partir das temperaturas previstas em várias concentrações de PEG200, poderíamos determinar a atividade de água da sequência sem precisar realizar o experimento.

### 5.5.3 Conclusões e Perspectivas

Nossos resultados foram bastante interessantes. Em todos os casos analisados, observamos um aumento de profundidade de poço médio,  $D$ . Esse aumento de energia se deve à influência do PEG200 que afeta diretamente a interação das moléculas de água com os nucleotídeos. Pelos nossos resultados, observamos que CG(RNA) é o mais hidratado, seguido por CG(DNA), AU(RNA) e AT(DNA). O fato de CG ser mais hidratado torna as diferenças de energia entre AT e CG, ou AU e CG, menores. A atuação do PEG200 é determinante para inferir se ocorrem mudanças de conformação devido à hidratação dos nucleotídeos. Uma mudança de energia na presença de PEG200 em relação ao caso com H<sub>2</sub>O (ausência de PEG200), especialmente maior no par CG do que no par AT e AU, pode indicar essa mudança. Para o RNA, identificamos que as mudanças de energia são pequenas, enquanto no DNA, não. Isso sugere que o DNA, na presença de PEG200, pode assumir outro tipo de conformação, passando de uma hélice do tipo B para uma hélice do tipo A. Outro detalhe importante é que os valores de AT são praticamente idênticos tanto em H<sub>2</sub>O quanto em PEG200. O modelo PB explica muito bem a dinâmica do sistema na ausência e na presença de PEG200. Seria interessante fazer este estudo das mesmas sequências para outras concentrações de PEG200, entre 0%-40%, sob as mesmas condições experimentais. Dessa forma, além de determinarmos as atividades de água para cada uma das sequências, poderíamos prever temperaturas de desnaturação em diferentes concentrações de PEG200 para sequências que não estão presentes em nossa base de dados. Com isso, poderíamos determinar a atividade de água de uma sequência sem precisar realizar o experimento.

## 5.6 Modelos Peyrard-Bishop (PB) com barreira de potencial

A proposta do modelo Peyrard-Bishop foi estudar a dinâmica de DNA durante o processo de transição de fita dupla para fita simples, através do aumento de temperatura do sistema, como discutido no capítulo 3, seção 3.1 e no primeiro artigo [43] publicado em 1989. No entanto, com o passar dos anos, novas metodologias foram empregadas no tratamento termodinâmico da dinâmica de nucleotídeos para prever temperaturas de desnaturação e foi percebido que isto pode ser aplicado a outros sistemas envolvendo nucleotídeos, como RNA e híbridos DNA/RNA [28, 29, 164]. No primeiro artigo de cada um dos modelos de barreira, os autores tratam novamente apenas a dinâmica de DNA [24–27]. A partir da mesma proposta do modelo original, assumiremos que os modelos PB com barreira foram criados para explicar as interações do DNA com o solvente. Então, nos concentraremos nessa primeira parte nos conjuntos de dados DNA-W/40 e DNA-P200/40. Ao encontrar os valores iniciais para se trabalhar com os modelos PB com barreira de potencial no DNA, faremos um processo de minimização semelhante ao modelo PB para DNA e RNA, como descrito na seção 5.5.1. Serão realizadas 3 minimizações para verificar se realmente estamos no mínimo da função  $\chi^2$ , seguidas por uma última minimização para estimar a influência do erro experimental. Desse modo, os modelos PB barreira passaram pelo mesmo tratamento que o modelo PB convencional, discutido na seção 5.5, e os resultados obtidos serão comparáveis. Para reproduzir os efeitos do PEG na estabilidade dos nucleotídeos e devido à falta de referências para os parâmetros dos potenciais barreira, utilizaremos como parâmetros iniciais de empilhamento,  $k$ , e Morse,  $D$ , os valores ajustados pela quarta rodada de minimizações do modelo PB, seção 5.5.2. Esperamos que os termos de barreira melhorem as previsões do modelo PB e forneçam informações de como as moléculas de água interagem com os pares de base durante o processo de desnaturação do DNA.

As otimizações dos modelos PB com barreira são bem diferentes das do modelo PB padrão, seção 5.5.1. O principal motivo é que não dispomos de referências dos parâmetros para os modelos PB com barreira de potencial. Logo, foi necessária uma análise inicial para avaliar o comportamento dos parâmetros nesses modelos antes de seguir o procedimento descrito na seção 5.5.1.

### 5.6.1 Análise inicial

Algumas afirmações já podem ser feitas a respeito da altura da barreira, que representa fisicamente a força de interação entre o solvente, neste caso a  $H_2O$ , e o nucleotídeo. Para existir a possibilidade de formação de dupla hélice, quando as ligações de hidrogênio das fitas simples com a fita dupla forem rompidas, a força que o solvente deve exercer sobre a base, quando esta se torna fita simples, deve ser menor ou igual à ligação base-base. Isso ocorre porque, se a ligação solvente-base for mais estável do que a ligação base-base, o sistema nunca mudaria para um duplexo. Deste modo,

como valores iniciais para o termo de altura da barreira, utilizaremos a faixa entre 0 e  $D$ , onde  $D$  é o valor da profundidade do poço do potencial de Morse que representa a ligação entre as bases. Para a distância de equilíbrio entre o solvente e base utilizaremos a faixa entre 0.05 e 0.1 nm. Essa distância foi determinada por análises experimentais, a distância de 0.05 nm é a região onde se concentra mais de 40% das moléculas de água ao redor do DNA [65]. As demais distâncias são utilizadas no estudo de dinâmica molecular do DNA [69, 70, 166]. Essas distâncias nos permitirão analisar o efeito das águas próximas ao DNA (distância de até 0.3 nm), à meio alcance do DNA (distância até 0.5 nm) e longe do DNA (distância até 1 nm). As águas próximas ao DNA são as principais responsáveis pela estabilização dos nucleotídeos [70]. Como parâmetro de alcance da barreira, utilizamos os valores de 22.5 nm para CG e 17.5 nm para AT de Luo et al. [152]. Eles encontraram esses valores em estudos de microscopia de força atômica em DNA. Durante o processo de separação da fita de DNA, as moléculas de  $H_2O$  se intercalam entre os pares de base. Essa interação, (par de base)- $H_2O$ -(par de base), é de longo alcance e se concentra na faixa de 20–25 nm para o par CG e 15–20 nm para o par AT, até que os pares de base estejam completamente separados [152]. Assim, usaremos os valores de 22.5 nm para CG e 17.5 nm para AT, que é a média desta faixa.

**Modelos HMS, PCLA e TRMF** Os modelo HMS (tabela 5), PCLA (tabela 6) e TRMF (tabela 8), apresentados nas páginas 79, 80 e 84, respectivamente, possuem uma parte do Hamiltoniano bastante similar,  $V_{Morse} + V_{barreira}$ . Entretanto, cada termo de barreira possui suas particularidades. Como os modelos pretendem discutir o mesmo efeito, ou seja, a interação base-solvente, manteremos os parâmetros de posição de equilíbrio, alcance da barreira e altura da barreira com os mesmos valores. Para o modelo HMS, avaliaremos duas funções de  $V_{barreira}$  uma com  $s$  positivo e outra com  $s$  negativo [24, 153]. Diferenciaremos esses dois casos por  $HMS^-$ , quando  $s = -1$ , e  $HMS^+$ , quando  $s = +1$ .

**Modelo PCLJ** Diferente dos outros modelos, o modelo PCLJ (tabela 7), apresentado na página 82, tem um problema de que o valor do poço de potencial de Morse,  $A$ , muda de acordo com os valores de  $\gamma$ ,  $\lambda_h$ ,  $E$  e  $D_h$  utilizados. Estudos anteriores do modelo PB realizados pelo nosso grupo, não alteramos o parâmetro  $\lambda$  que representa o alcance da ligação de hidrogênio entre os pares de base (potencial de Morse, equação (11)), pois ele não traz diferença significativa para os nossos resultados [28]. Assumiremos que o parâmetro  $\lambda_h$  do modelo PCLJ deve possuir comportamento semelhante ao modelo PB durante as minimizações. Assim, para termos certeza de que cobriremos a faixa em que o potencial de Morse  $A$  é similar ao Morse  $D$  dos demais modelos, alteramos os valores de  $\lambda_h$  (AT) e  $\lambda_h$  (CG) do modelo PB para os valores mostrados na tabela 12. Uma diferença deste modelo em relação aos demais é que após a separação base-solvente-base para base-solvente, há uma energia pequena remanescente representada pelo parâmetro  $D_h$ , isto faz todo sentido, pois

após a separação dos nucleotídeos e do estado intermediário base-solvente-base para o estado de fita simples, existe uma energia de ligação entre a base e o solvente. Embora não sabemos o valor desta interação, sabemos que ela deve ser pequena. Assim, testaremos os valores de  $E$  para três valores de  $D_h$ : 0, 1 e 10 meV. Posteriormente, avaliaremos se é necessário variar este parâmetro quando formos realizar às minimizações.

Nessa etapa, não sabemos qual valor de altura da barreira a ser utilizado para iniciar o processo de minimização. Assumimos que a altura da barreira deve ser menor ou igual ao potencial de Morse  $D$ . No modelo HMS, o termo  $f_s$  representa uma porcentagem do valor da profundidade  $D$  do potencial de Morse, conforme a equação (34). Utilizaremos a faixa entre 0 e 1 desse parâmetro, a fim de observar o efeito prático da adição do potencial solvente ao modelo PB. Faremos isso utilizando um processo chamado de varredura. O processo de varredura consiste em calcular os parâmetros de qualidade,  $\chi^2$  ou  $\langle T \rangle$ , para cada valor de  $f_s$  em um intervalo e passo determinados. Nos concentraremos apenas nos valores de  $\chi^2$ . De modo a facilitar as análises, mostraremos nos gráficos os valores de  $\langle \chi^2 \rangle = \chi^2/N$ , onde  $N$  é o número de sequências utilizadas, neste caso 40, conforme a tabela 9. Como mencionado anteriormente, o intervalo avaliado foi entre 0 e 1, já o passo foi de 0.1, ao todo teremos 11 pontos. Os pares de base AT e CG interagem diferente com o solvente, por causa da sua estrutura química. Desse modo, cada valor de  $\langle \chi^2 \rangle$  será determinado por um valor de  $f_s$  (AT) para AT e um  $f_s$  (CG) para CG. De forma a normalizar todos os modelos em unidades de energia, multiplicaremos  $f_s$  por  $D$  e usaremos esses valores de energia para observar o mapa de  $\langle \chi^2 \rangle$  de cada modelo. Assim, realizaremos uma varredura dos parâmetros relacionados à altura da barreira na faixa mostrada na tabela 12,  $0 < \text{altura da barreira} < D$ , para todos os modelos e manteremos os demais parâmetros constantes. Ao final, teremos uma matriz  $11 \times 11$  de  $\langle \chi^2 \rangle$  do espectro da região analisada. Construiremos um mapa de calor de  $\langle \chi^2 \rangle$  desta região e, em seguida, avaliaremos se o modelo produz valores de  $\langle \chi^2 \rangle$  melhores do que os do modelo PB. Utilizaremos o melhor valor  $\langle \chi^2 \rangle$ , ou seja, o menor valor de cada modelo para iniciar o processo de minimização. Caso o valor de um dos parâmetros seja zero, usaremos o valor do melhor  $\langle \chi^2 \rangle$ , onde os parâmetros de altura da barreira sejam diferentes de zero. Um resumo dos parâmetros utilizados para a busca dos parâmetros iniciais para iniciar o processo de minimização encontra-se na tabela 12

**Otimizações de teste** Realizamos três testes para determinar qual é o melhor processo de minimização para se trabalhar com os modelos:

1. Variar apenas os parâmetros relevantes relacionados ao solvente mantendo os parâmetros de empilhamento,  $k$ , e Morse,  $D$ , fixos;
2. Variar os parâmetros relevantes relacionados ao solvente e os de Morse,  $D$ , mantendo os parâmetros de empilhamento,  $k$ , fixos;

Parâmetro	tipo	DNA-W/40 DNA-P200/40
modelo HMS		
$f_s (AT) = f_s (CG)$	percentual da altura da barreira	$0 < f_s < 1$
$\lambda_s (AT)$	alcance da barreira	17.5 nm
$\lambda_s (CG)$	alcance da barreira	22.5 nm
$y_e (AT) = y_e (CG)$	posição de equilíbrio (solvente-base)	entre 0.05 e 1 nm
$s_{AT} = s_{CG}$	funções utilizadas na literatura [24, 153]	+1 ou -1
modelo PCLA		
$b_{AT}$	altura da barreira	$0 < b < D_{AT}$
$b_{CG}$	altura da barreira	$0 < b < D_{CG}$
$c_{AT}$	inverso do alcance da barreira	0.0057
$c_{CG}$	inverso do alcance da barreira	0.0044
$d_{AT} = d_{CG}$	posição de equilíbrio (solvente-base)	entre 0.05 e 1
$\alpha_{AT}$	inverso do alcance do potencial de Morse	$0.28 \text{ nm}^{-1}$
$\alpha_{CG}$	inverso do alcance do potencial de Morse	$0.94 \text{ nm}^{-1}$
modelo PCLJ		
$D_h (AT) = D_h (CG)$	energia final da fita simples	0 ou 1 ou 10 meV
$E_{(AT)}$	altura da barreira	$0 < E < D_{AT}$
$E_{(CG)}$	altura da barreira	$0 < E < D_{CG}$
$\lambda_h (AT)$	alcance do potencial de Morse	$5.87 \times 10^{-3} \text{ nm}$
$\lambda_h (CG)$	alcance do potencial de Morse	$7.18 \times 10^{-3} \text{ nm}$
$\gamma_{(AT)}$	alcance da barreira	17.5 nm
$\gamma_{(CG)}$	alcance da barreira	22.5 nm
modelo TRMF		
$G_{AT}$	altura da barreira	$0 < G < D_{AT}$
$G_{CG}$	altura da barreira	$0 < G < D_{CG}$
$y_o (AT) = y_o (CG)$	posição de equilíbrio (solvente-base)	entre 0.05 e 1 nm
$b_G (AT)$	alcance da barreira	17.5 nm
$b_G (CG)$	alcance da barreira	22.5 nm

**Tabela 12**

Parâmetros utilizados para iniciar o processo de minimização dos modelos com barreira. Os parâmetros de empilhamento,  $k$  e  $D$ , iniciais foram os ajustados pela terceira minimização do modelo PB, mostrados nas figuras 36a e 37a.

3. Variar os parâmetros relevantes relacionados ao solvente e os de empilhamento,  $k$ , mantendo os parâmetros de Morse,  $D$ , fixos.

Os parâmetros relevantes dos modelos com barreira são relacionados às energias. Isso inclui as energias de barreira para os modelos HMS, PCLA, PCLJ e TRMF, bem como a energia da interação entre a fita simples e às moléculas de H<sub>2</sub>O do modelo PCLJ. Os termos referentes ao alcance e distâncias de equilíbrio da interação base-solvente manteremos constantes, pois esses valores vieram de referências experimentais e cálculos teóricos de dinâmica molecular [65, 70, 152] para compararmos nossos resultados. Nos modelos PCLA e PCLJ, há um parâmetro referente ao alcance do potencial de Morse que manteremos constantes. No modelo PCLJ, sairemos do regime de profundidade de poço  $A$  diferente do modelo PB, o que dificulta uma comparação direta entre os modelos. Já no modelo PCLA, teremos valores diferentes de alcance do potencial de morse para os termos  $V_{Morse}$  e  $V_{barreira}$ . Assumiremos que  $\alpha$  do modelo PCLA tem pouca relevância nas minimizações, análogo à outras análises do modelo PB realizadas pelo nosso grupo [28].

Nesta etapa, para poupar tempo, faremos apenas uma otimização com 100 rodadas, visto que isso é suficiente para observar o comportamento dos parâmetros. Com isso, faremos uma análise preliminar para verificar se os resultados fazem sentido com a literatura. Em caso positivo, completaremos as 1000 rodadas de minimizações para refinar nossos resultados.

Os modelos de barreira explicam um estado intermediário que não existe no modelo PB, como discutido na seção 5.3. A existência deste estado intermediário faz todo o sentido físico. Porém, o modelo PB consegue descrever muito bem o que ocorre no sistema, como foi mostrado em estudos anteriores [28–30]. Surge então a questão: Os modelos com barreira conseguem prever temperaturas de desnaturação melhores do que o modelo PB e, adicionalmente, fornecer mais informações sobre a dinâmica do sistema? Antes de respondermos a essa pergunta, precisamos identificar os parâmetros iniciais dos diferentes modelos com barreira. Pela tabela 12, nos resta apenas definir qual valor de energia de barreira inicial utilizar nas minimizações. Então, primeiro procuramos as energias de barreiras iniciais para os diferentes modelos PB com barreira, HMS (página 79), PCLA (página 80), PCLJ (página 82) e TRMF (página 84). Em seguida, realizamos diversos procedimentos de minimização para encontrar qual a melhor forma de se trabalhar com cada um dos modelos PB com barreira. Nessa etapa, aplicaremos alguns critérios para avaliar os modelos:

- Os valores de predição de temperaturas sejam melhores do que do modelo PB padrão;
- Os resultados devem ter sentido físico com a dinâmica do DNA.

Ainda nesta parte, faremos uma comparação dos resultados dos modelos PB com barreira com os

resultados encontrados pelo modelo PB, conforme discutido na seção 5.5.

### 5.6.2 Otimizações finais

Os modelos aprovados nas análises iniciais, conforme a seção 5.6.1, serão submetidos ao mesmo procedimento de otimização empregado para o modelo PB, descrito na seção 5.5.1. Completaremos as 100 rodadas de minimizações da etapa anterior para 1000 e faremos mais 2 rodadas de otimizações com 1000 rodadas para confirmarmos se nos encontramos no mínimo da função  $\chi^2$ , como descrito na seção 3.1.3. Dessa forma, tanto a análise pelo modelo PB padrão quanto as dos modelos PB com barreira passarão pelo mesmo número de procedimentos de otimizações. Por fim, avaliaremos a influência do erro experimental em nossos resultados, da mesma maneira que fizemos anteriormente com o modelo PB. Para isso, utilizaremos a mesma incerteza de 0.7 °C.

Nesta etapa, voltaremos a estudar as sequências de RNA e compararemos os resultados dos modelos de barreira com os resultados do modelo PB padrão. Logo, teremos que identificar os parâmetros iniciais dos modelos de barreira aprovados na análise inicial para o RNA. O alcance e a posição de equilíbrio são idênticos ao DNA, como especificado na tabela 12, pois os grupos funcionais que formam as ligações de hidrogênio entre os pares de base são os mesmos nas bases nitrogenadas, conforme mostrado na figura 2c. Resta apenas identificar as energias de barreira. Para isso, utilizaremos o mesmo procedimento de varredura que realizamos para o DNA. Na próxima seção, apresentaremos as análises iniciais dos modelos com barreira. Os modelos aprovados serão submetidos a mais rodadas de otimizações, como descrito nesta seção.

### 5.6.3 Análise Inicial - Parâmetros iniciais dos modelos com barreira

Como discutido nas seções anteriores, determinamos quase todos os parâmetros iniciais para os modelos com barreira, mostrados na tabela 12. Os únicos parâmetros que nos restam encontrar são as energias de barreira para cada modelo. A princípio, os valores de energia podem ser quaisquer números. Entretanto, impomos a restrição de que esse parâmetro deve se limitar ao intervalo entre 0 e o valor de profundidade média do poço  $D$ , como encontrado pelo modelo PB (figura 36a). Essa restrição é baseada no seguinte argumento: a ligação base(fita principal)-solvente-base(fita complementar), estado intermediário entre fita dupla e fita simples, deve ser menor ou igual à ligação base(fita principal)-base(fita complementar) no estado de fita dupla. Após as ligações de hidrogênio entre as fitas simples forem rompidas, surge uma ligação base(fita principal)-solvente-base(fita complementar), onde o solvente realiza uma interação indireta entre os pares de base. Acreditamos que se a energia de barreira for maior do que a energia de ligação entre as fitas, o sistema nunca mudaria para o estado de fita dupla a temperaturas abaixo da temperatura de desnaturação. Isso

ocorre porque, neste caso, a ligação solvente-base seria mais estável do que a ligação base-base. Mesmo com esta restrição, ainda há um grande número de possibilidades para as energias de barreira iniciais. Para contornar esse problema, optamos por manter os demais parâmetros constantes e variar apenas as energias de barreira AT e CG no intervalo de 0 a  $D$ . Dessa forma, em vez de otimizar todos os parâmetros, calculamos a função de partição, determinamos o índice de *melting* e, por fim, computamos as temperaturas de desnaturação, conforme detalhado nos capítulos 3.1, 3.1.1 e 3.1.2. Assim, os parâmetros de qualidade nos forneceram valores não otimizados para cada valor de energias de barreira AT e CG. Em outras palavras, determinamos quão distante as temperaturas calculadas estão das experimentais para cada ponto (AT, CG) no domínio. Entre os parâmetros de qualidade que utilizamos, conforme as equações da seção 3.1.3, escolhemos o  $\chi^2$ , equação (30). Para cada valor de energia de barreira, temos um valor associado de  $\chi^2$ . Alguns dos valores de  $\chi^2$  são muito grandes, então, nos gráficos representamos para cada ponto (AT, CG) o valor de  $\langle \chi^2 \rangle = \chi^2/N$ , onde  $N$  é o número total de sequências na base de dados (neste caso, 40, como indicado na tabela 9). Ao final, construímos uma matriz de pontos de energias de barreira AT  $\times$  CG. A princípio, escolheríamos o ponto de menor valor de  $\langle \chi^2 \rangle$  fora do eixo 0 de um dos parâmetros, pois não faria sentido que a barreira de energia existisse apenas para um par de bases. Além disso, esse é o ponto mais próximo do mínimo da função  $\langle \chi^2 \rangle$ , que representa a menor diferença entre as temperaturas calculadas e experimentais. Entretanto, usamos essa abordagem apenas para os modelos HMS e TRMF. Para os modelos PCLA e PCLJ, utilizamos outros critérios de seleção de energias de barreira, discutidos nos próximos parágrafos.

**Modelo PCLA** Para o conjunto DNA-P200/40, escolhemos o ponto de menor valor de  $\langle \chi^2 \rangle$  fora do eixo 0 de um dos parâmetros, seguindo o mesmo critério aplicado aos modelos HMS e TRMF. Entretanto, para as sequências DNA-W/40, esse ponto tem valor de energia de barreira AT igual à profundidade de poço  $D_{AT}$  e uma energia de barreira CG pequena. Uma energia de barreira maior no par AT do que no par CG não faria sentido, pois CG é mais hidratado do que AT [4, 167]. Notamos que, para o valor de energia de barreira CG fixo em  $b_{CG} = 8.35 \text{ meV/\AA}^3$ , tabela 13, os valores de  $\langle \chi^2 \rangle$  são muito próximos, o que significa que eles fornecem os mesmos valores de predição de temperaturas. Assim, escolhemos para as sequências DNA-W/40 o mesmo ponto que foi escolhido para as sequências DNA-P200/40. No entanto, os valores de energia de barreira são diferentes entre os dois conjuntos, conforme mostrado na tabela 13, devido aos valores diferentes de profundidade do poço  $D$  das sequências em  $\text{H}_2\text{O}$ , DNA-W/40, e em PEG200, DNA-P200/40, apresentados na (figura 36a). Como resultado, o intervalo entre 0 e  $D$  avaliado é diferente entre os dois conjuntos.

**Modelo PCLJ** No caso do modelo PCLJ, vale a pena recordar que para cada valor de  $\gamma$ ,  $\lambda_h$ ,  $E$  e  $D_h$ , obtemos valores de profundidades de poço  $A$  diferentes, como definido pelas equações (37).

Após a separação do par de bases, as moléculas de água interagem indiretamente com os nucleotídeos AT e CG. Existe apenas uma diferença entre os grupos funcionais que se localizam entre os nucleotídeos, o grupo -CH na adenina e o grupo -NH<sub>2</sub> na guanina (posição C(4)), conforme ilustrado na figura 2c (página 29). Devido ao grupo -NH<sub>2</sub> ser mais eletronegativo e formar mais ligações de hidrogênio com a H<sub>2</sub>O em comparação ao grupo -CH, é esperado que o par CG possua uma energia de barreira ligeiramente maior do que o par AT. Nesse contexto, escolhemos os valores de energia de barreira  $E_{AT} \approx 11 \text{ meV/\AA}$  e  $E_{CG} \approx 12 \text{ meV/\AA}$  para as sequências DNA-W/40. Em seguida, variamos os parâmetros  $\lambda_h$  (AT) e  $\lambda_h$  (CG) até encontrarmos valores de  $A_{AT}$  e  $A_{CG}$  próximos ao  $D_{AT}$  e  $D_{CG}$  do modelo PB, conforme mostrado na figura 36. Os valores resultantes para  $\lambda_h$  (AT) e  $\lambda_h$  (CG) foram  $5.87 \times 10^{-3} \text{ nm}$  e  $7.18 \times 10^{-3} \text{ nm}$ , respectivamente. Após encontrarmos os valores de  $\lambda_h$ , variamos os parâmetros de  $E_{AT}$  e  $E_{CG}$  para as sequências em PEG200 que produzem valores de profundidade de poço  $A_{AT}$  e  $A_{CG}$  próximos ao  $D_{AT}$  e  $D_{CG}$  do modelo PB, conforme ilustrado na figura 36. Os valores encontrados foram  $E_{AT} \approx 11.2 \text{ meV/\AA}$  e  $E_{CG} \approx 13.4 \text{ meV/\AA}$ . Com isso, utilizamos os mapas de calor de  $\langle \chi^2 \rangle$  apenas para avaliar o quão distante do mínimo da função  $\chi^2$  nos encontramos, visto que alterar os valores de  $E_{AT}$  ou  $E_{CG}$  nos levaria para fora do regime em que as profundidade de poço  $A_{AT} \approx D_{AT}$  e  $A_{CG} \approx D_{CG}$ . Isso, por sua vez, dificultaria uma comparação direta entre os modelos.  $D_h$  representa a energia da fita simples após a completa separação dos nucleotídeos. Avaliamos o mapa de calor  $E_{AT} \times E_{CG}$  para três valores de  $D_h$ : 0, 1 e 10 meV. Embora não notamos nenhuma melhora nos valores de  $\langle \chi^2 \rangle$  ao variar  $D_h$ , consideramos que uma pequena energia residual após a separação da fita dupla faz todo sentido físico. Assim, escolhemos o valor de  $D_h = 1 \text{ meV}$  como valor inicial tanto para AT quanto para CG, avaliados de forma distinta para cada par de base.

Na tabela 13, apresentamos um resumo os valores de energia de barreira iniciais para cada modelo, bem como os valores de  $\langle \chi^2 \rangle$  iniciais dos cálculos. Os valores de energia encontrados permanecem os mesmos independente da distância de equilíbrio entre a base e o solvente utilizada (modelos HMS, PCLA e TRMF), porém observamos ligeiras variações nos valores  $\langle \chi^2 \rangle$ . Acreditamos que os modelos podem ser capazes de diferenciar moléculas de águas próximas daquelas mais distantes do DNA, fornecendo energias de barreira que seguem uma dependência com a distância entre a base e o solvente. Investigaremos essa possibilidade na próxima seção. Para obter uma descrição detalhada de como encontramos os valores iniciais de energia de barreira para cada modelo, consulte o apêndice B.

#### 5.6.4 Análise Inicial - Otimizações de teste

Nesta seção, avaliaremos se os modelos com barreira são úteis para prever temperaturas de desnaturação. Lembrando que realizamos três testes: 1. Variar apenas os termos de energia da

Parâmetro	DNA-W/40	DNA-P200/40
modelo HMS		
HMS <sup>-</sup>		
$f_s$ (AT)	0.2	0.1
$f_s$ (CG)	0.1	0.1
$s_{AT} = s_{CG}$	-1	-1
$\langle \chi^2 \rangle$	4.43 °C <sup>2</sup>	1.43 °C <sup>2</sup>
HMS <sup>+</sup>		
$f_s$ (AT)	0.3	0.4
$f_s$ (CG)	0.1	0.1
$s_{AT} = s_{CG}$	+1	+1
$\langle \chi^2 \rangle$	4.50 °C <sup>2</sup>	1.25 °C <sup>2</sup>
modelo PCLA		
$b_{AT}$	7.81 meV/Å <sup>3</sup>	8.12 meV/Å <sup>3</sup>
$b_{CG}$	8.35 meV/Å <sup>3</sup>	9.39 meV/Å <sup>3</sup>
$\langle \chi^2 \rangle$	9.74 °C <sup>2</sup>	4.74 °C <sup>2</sup>
modelo PCLJ		
$D_h$ (AT) = $D_h$ (CG)	1 meV	1 meV
$E_{(AT)}$	11 meV/Å	11.2 meV/Å
$E_{(CG)}$	12 meV/Å	13.4 meV/Å
$\langle \chi^2 \rangle$	entre 17.85 °C <sup>2</sup> e 102.3 °C <sup>2</sup>	entre 25.97 °C <sup>2</sup> e 43.11 °C <sup>2</sup>
modelo TRMF		
$G_{AT}$	3.91 meV	4.06 meV
$G_{CG}$	8.35 meV	9.39 meV
$\langle \chi^2 \rangle$	4.69 °C <sup>2</sup>	2.03 °C <sup>2</sup>

**Tabela 13**

Parâmetros de energia de barreira utilizados para iniciar o processo de minimização dos modelos com barreira.

barreira, 2. Variar os termos de energia da barreira e os termos do potencial de Morse  $D$ , e 3. Variar os termos de energia da barreira e os termos de empilhamento  $k$ . Realizamos os três testes em quatro diferentes valores de distância entre os nucleotídeos e as moléculas de  $H_2O$ . Para avaliar os modelos, utilizaremos os seguintes critérios:

- I. O valor de  $\chi^2$  deve ser menor do que o modelo PB;
- II. O valor de energia de barreira CG deve ser levemente maior do que AT no caso com  $H_2O$ ;
- III. O valor de profundidade de poço do potencial de Morse deve ser maior nas sequências em PEG200 do que em  $H_2O$ .

O critério I. nos informa se a modificação com o potencial de barreira resulta em previsões de temperaturas melhores do que o modelo original. Já os demais critérios nos informam sobre a físico-química do sistema. O critério II. destaca que, após a separação dos nucleotídeos, as moléculas de água se inserem entre os pares de base, estabelecendo uma ligação base- $H_2O$ -base. Observando a figura 2c (posição C(4)), percebemos que existe apenas um grupo funcional diferente entre as pares de bases: o grupo  $-NH_2$  na guanina e o grupo  $-CH$  na adenina. O grupo  $-NH_2$ , além de ser mais eletronegativo, faz mais ligações de hidrogênio com a  $H_2O$  do que o grupo  $-CH$ . Logo, o par CG deve apresentar uma energia de barreira um pouco maior do que o par AT. Quanto ao critério III., avaliaremos o efeito do PEG200. O PEG200 reduz a interação das moléculas de  $H_2O$  com o nucleotídeo [35, 140, 143, 144]. Assim, as moléculas de água, que antes atuavam de forma a reduzir a ligação de hidrogênio entre os pares de base, o fazem de forma menos intensa. Consequentemente, esperamos um aumento no potencial de Morse, tal como aconteceu com os resultados do modelo PB, discutido na seção 5.5.2.

**Modelo HMS** Das 100 rodadas de minimizações do modelo  $HMS^-$ , conforme a tabela 5 com  $s = -1$ , e do  $HMS^+$ , conforme a tabela 5 com  $s = +1$ , todas as minimizações foram concluídas com sucesso. Nas tabelas 14 e 15, apresentamos os resultados para os três testes. Realizamos as minimizações para quatro valores diferentes de  $y_e$ . O motivo dessa variação foi para investigar se os modelos com barreira conseguem diferenciar moléculas de água próximas do DNA de moléculas de água distantes.

Ao contrário dos demais modelos com barreira, nos quais variamos a energia da barreira, no modelo HMS, variamos o valor de  $f_s$  durante as minimizações. Após completar os cálculos, determinamos a altura da barreira ao multiplicar o valor de  $f_s$  pelo valor de  $D$ . Ambos os valores estão na tabelas 14 e 15. O erro da energia de barreira, valores entre parênteses, foi encontrado utilizando o método de propagação de derivadas na função  $f_s D$ . Com base nos critérios I., II. e III.

avaliaremos a possibilidade de utilizar o modelo HMS para prever temperaturas de desnaturação. Primeiramente, avaliaremos o modelo HMS<sup>-</sup> e posteriormente o modelo HMS<sup>+</sup>.

Pelos resultados da tabela 14, observamos que os testes 2. e 3. conseguiram predições de temperatura em H<sub>2</sub>O (DNA-W/40) melhores que as do modelo PB, enquanto em PEG200 (DNA-P200/40) todos os testes apresentaram temperaturas preditas melhores do que o modelo padrão. Somente no teste 1. em H<sub>2</sub>O, os resultados das predições foram levemente piores do que o modelo PB. Assim, todos os testes do modelo HMS<sup>-</sup> foram aprovados no critério I.

No critério II., as energias de barreira AT e CG devem ser próximas, ademais a energia de CG em H<sub>2</sub>O (DNA-W/40) deve ser levemente maior do que AT. Apesar dos valores médios de  $f_s$  ( $D_{AT}$ ) serem levemente maiores do que  $f_s$  ( $D_{CG}$ ) nos testes 1. e 3., ao incluirmos a barra de erro nas análises indica que os valores de energia de barreira CG podem ser maiores do que AT. No teste 2., os valores médios de energia de barreira AT são bem maiores do que as de CG. Ao incluirmos a barra de erro, os valores de energia de barreira CG podem ser maiores que AT. Entretanto, com base em nossa experiência, é provável que as próximas minimizações produzam valores de energia de barreira médias próximos aos valores encontrados na tabela 14, com menores dispersões. Logo, o teste 2. não conseguirá valores de energia de barreiras de potencial CG levemente maiores do que AT em H<sub>2</sub>O (DNA-W/40). Apenas os testes 1. e 3. foram aprovados no critério 2.. Observe que os valores de  $f_s$  no teste 1. (tabela 14) são aproximadamente zero, ou seja, ao variar apenas o termo de barreira, o modelo HMS assume que o modelo PB é o melhor. Portanto, não é interessante para nós explorar o modelo com o teste 1..

Os valores de Morse CG e AT dos testes 1. e 3. já respeitam automaticamente o critério III., já que não alteramos esses parâmetros nesses testes. No entanto, no teste 2., os valores de profundidades de poço médias  $D_{CG}$  são maiores do que as de  $D_{AT}$ , porém  $D_{AT}$ -DNA-P200/40 é menor do que  $D_{AT}$ -DNA-W/40. Além disso, observamos que as profundidades médias  $D_{CG}$  e  $D_{AT}$  das sequências em H<sub>2</sub>O e PEG200/40%wt são muito próximas. Nesse caso, a adição de PEG200/40%wt não possui efeito significativo nas ligações de hidrogênio. As diferenças de temperatura  $T_{calc}$  e  $T_{exp}$  mostradas na tabela 23 são bem grandes. Acreditamos que essa diferença de temperatura não possa ser atribuída apenas ao termo de empilhamento. Com base no critério III., apenas os testes 1. e 3. foram aprovados.

Diferente do modelo HMS<sup>-</sup>, como mostrado na tabela 14, as predições de temperatura do modelo HMS<sup>+</sup> (tabela 15 - coluna  $\chi^2$ ) são bem melhores. Pelo critério I., apenas os testes 2. e 3. apresentaram valores de  $\chi^2$  melhores do que o modelo PB. O teste 2. em H<sub>2</sub>O (DNA-W/40) teve resultados levemente piores. Já o teste 1. em PEG200 (DNA-P200/40,) foi o que apresentou valores de predições de temperatura melhores. Dos resultados de  $\Delta\chi^2$  apresentados na tabela 15, apenas os testes 2. e 3. foram aprovados no critério I.

As alturas das barreira AT e CG,  $f_s D$ , são bem próximas nos testes 1. e 2. em H<sub>2</sub>O (DNA-

DNA-W/40											
Min.	$y_e$ (nm)	$\chi^2$ ( $^{\circ}\text{C}^2$ )	$\langle \Delta T \rangle$ ( $^{\circ}\text{C}$ )	$\Delta\chi^2$ ( $^{\circ}\text{C}^2$ )	$f_s$ (AT)	$f_s$ (CG)	$f_s$ (AT) $D_{AT}$ (meV)	$f_s$ (CG) $D_{CG}$ (meV)	$D_{AT}$ (meV)	$D_{CG}$ (meV)	
1.	0.05	166.55	1.69	0.61	0.013(0.035)	0.005(0.019)	0.51(1.36)	0.38(1.56)	39.07(6.21)	83.51(9.68)	
	0.3	166.61	1.70	0.66	0.015(0.041)	0.006(0.022)	0.59(1.60)	0.47(1.86)	39.07(6.21)	83.51(9.68)	
	0.5	166.36	1.69	0.41	0.009(0.021)	0.002(0.011)	0.36(0.83)	0.14(0.89)	39.07(6.21)	83.51(9.68)	
	1	166.32	1.69	0.38	0.008(0.021)	0.002(0.010)	0.30(0.80)	0.15(0.87)	39.07(6.21)	83.51(9.68)	
	0.05	154.25	1.57	-11.69	0.24(0.15)	0.07(0.03)	11.17(6.95)	6.94(2.97)	46.87(6.29)	94.01(5.50)	
2.	0.3	153.75	1.56	-12.20	0.23(0.14)	0.07(0.03)	10.84(6.67)	6.44(2.97)	47.32(6.41)	94.22(4.94)	
	0.5	155.83	1.60	-10.11	0.22(0.10)	0.07(0.03)	10.06(5.05)	6.53(3.19)	46.26(6.83)	92.83(5.89)	
	1	153.16	1.53	-12.79	0.22(0.11)	0.07(0.03)	10.58(5.53)	6.33(2.48)	48.17(5.19)	94.65(3.12)	
	0.05	162.34	1.67	-3.61	0.20(0.07)	0.08(0.03)	7.71(3.07)	6.71(2.24)	39.07(6.21)	83.51(9.68)	
	0.3	162.69	1.68	-3.25	0.19(0.08)	0.07(0.03)	7.32(3.45)	6.00(2.73)	39.07(6.21)	83.51(9.68)	
3.	0.5	163.51	1.69	-2.44	0.18(0.09)	0.08(0.03)	6.90(3.69)	6.43(2.54)	39.07(6.21)	83.51(9.68)	
	1	164.11	1.69	-1.83	0.20(0.08)	0.08(0.02)	7.75(3.40)	6.59(2.17)	39.07(6.21)	83.51(9.68)	
DNA-P200/40											
1.	0.05	47.18	0.76	-6.85	0.01(0.04)	0.04(0.04)	0.58(1.43)	4.01(3.47)	40.61(2.96)	93.90(3.41)	
	0.3	45.16	0.74	-8.87	0.01(0.03)	0.03(0.03)	0.37(1.12)	3.12(3.29)	40.61(2.96)	93.90(3.41)	
	0.5	46.87	0.75	-7.16	0.02(0.04)	0.04(0.04)	0.64(1.46)	3.98(3.63)	40.61(2.96)	93.90(3.41)	
	1	45.59	0.74	-8.43	0.01(0.03)	0.04(0.03)	0.31(1.06)	3.32(3.28)	40.61(2.96)	93.90(3.41)	
	0.05	44.53	0.73	-9.49	0.11(0.05)	0.08(0.02)	4.94(2.23)	8.13(4.76)	44.24(3.49)	98.79(3.94)	
2.	0.3	44.11	0.73	-9.91	0.11(0.04)	0.08(0.02)	4.70(1.66)	8.13(3.82)	44.58(3.53)	98.69(0.87)	
	0.5	44.59	0.74	-9.44	0.11(0.03)	0.08(0.02)	4.74(1.34)	8.29(3.20)	44.41(3.13)	98.72(0.68)	
	1	44.64	0.74	-9.38	0.10(0.02)	0.09(0.02)	4.53(1.08)	8.44(2.77)	44.70(2.96)	98.75(0.86)	
	0.05	32.65	0.68	-21.37	0.12(0.04)	0.06(0.02)	4.96(1.57)	5.57(1.73)	40.61(2.96)	93.90(3.41)	
	0.3	43.87	0.68	-10.16	0.12(0.03)	0.06(0.02)	4.83(1.44)	5.82(2.15)	40.61(2.96)	93.90(3.41)	
3.	0.5	47.72	0.75	-6.31	0.13(0.04)	0.06(0.02)	5.35(1.69)	5.45(1.74)	40.61(2.96)	93.90(3.41)	
	1	42.77	0.66	-11.26	0.13(0.04)	0.06(0.02)	5.21(1.62)	5.55(2.00)	40.61(2.96)	93.90(3.41)	

**Tabela 14**

Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo HMS<sup>-</sup>. Mostramos os resultados para as sequências em H<sub>2</sub>O e PEG200 para quatro valores diferentes de  $y_e$ .  $\Delta\chi^2 = \chi^2_{(HMS^-)} - \chi^2_{(PB)}$ , onde  $\chi^2_{(PB)}$  se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos  $f_s$ , 2. Variar os termos  $f_s$  e os termos do potencial de Morse  $D$  e 3. Variar os termos  $f_s$  e os termos de empilhamento  $k$ .

W/40), conforme mostra a tabela 15. Porém, os valores de energia AT são levemente maiores do que CG. Ao considerar as barras de erro, os valores de CG podem ser maiores do que AT. No teste 3., as diferenças de energia de barreira  $f_s$  (AT)  $D_{AT}$  e  $f_s$  (CG)  $D_{CG}$  são bem maiores. Como mencionado anteriormente para o modelo HMS<sup>-</sup>, os valores de energia média finais, após mais rodadas de minimizações, tendem a oscilar próximos aos valores do teste 3. na tabela 15, com dispersões menores. Portanto, apenas os testes 1. e 2. foram aprovados pelo critério II..

No critério III. observamos se valores de profundidade de poço  $D$  em H<sub>2</sub>O (DNA-W/40) são maiores do que em PEG200 (DNA-P200/40). Os testes 1. e 3. já atendem automaticamente ao critério III., pois mantemos  $D$  constante. A tabela 15 mostra que os valores de  $D_{CG}$  são maiores do que  $D_{AT}$ . Ao comparar  $D$  em H<sub>2</sub>O com  $D$  em PEG200, observamos que  $D_{AT}$  das sequências DNA-W/40 são maiores do que  $D_{AT}$  das sequências DNA-P200/40. Considerando a barra de erro, há a possibilidade de  $D_{AT}$  em PEG200 ser maior do que H<sub>2</sub>O, mas as minimizações seguntes tendem a produzir valores médios de parâmetros próximos aos resultados anteriores, com menor variação. Assim, acreditamos que  $D_{AT}$  em H<sub>2</sub>O será maior do que  $D_{AT}$  em PEG200. Com base nisso, apenas os testes 1. e 3. passaram no critério III..

Em resumo, o teste 3. do modelo HMS<sup>-</sup> foi aprovado em todos os nossos critérios. Por outro lado, para o modelo HMS<sup>+</sup>, nenhum dos testes conseguiu ser aprovado nos três critérios utilizados para proseguir com as minimizações. Notamos que as energias de barreira não seguem

DNA-W/40										
Min.	$y_e$ (nm)	$\chi^2$ ( $^{\circ}\text{C}^2$ )	$\langle \Delta T \rangle$ ( $^{\circ}\text{C}$ )	$\Delta\chi^2$ ( $^{\circ}\text{C}^2$ )	$f_s$ (AT)	$f_s$ (CG)	$f_s$ (AT) $D_{AT}$ (meV)	$f_s$ (CG) $D_{CG}$ (meV)	$D_{AT}$ (meV)	$D_{CG}$ (meV)
1.	0.05	174.10	1.75	8.16	0.07(0.10)	0.02(0.04)	2.58(4.00)	1.93(3.63)	39.07(6.21)	83.51(9.68)
	0.3	175.83	1.76	9.89	0.10(0.12)	0.04(0.05)	4.76(3.75)	3.32(3.04)	39.07(6.21)	83.51(9.68)
	0.5	174.88	1.75	8.94	0.09(0.11)	0.03(0.05)	3.32(4.39)	2.56(4.07)	39.07(6.21)	83.51(9.68)
	1	175.85	1.75	9.91	0.09(0.12)	0.04(0.05)	3.63(4.61)	3.03(4.26)	39.07(6.21)	83.51(9.68)
	0.05	163.29	1.66	-2.65	0.22(0.08)	0.091(0.008)	11.09(4.29)	8.07(2.50)	51.28(8.11)	91.49(8.24)
2.	0.3	166.19	1.69	0.25	0.22(0.09)	0.09(0.04)	11.13(4.88)	8.28(3.74)	49.56(8.51)	90.56(9.97)
	0.5	165.11	1.68	-0.84	0.20(0.09)	0.09(0.03)	9.89(4.68)	7.77(3.24)	49.25(7.82)	90.76(8.58)
	1	164.93	1.68	-1.01	0.21(0.09)	0.09(0.03)	10.57(4.76)	8.39(2.82)	49.55(8.07)	91.05(9.04)
	0.05	143.35	1.59	-22.60	0.33(0.11)	0.06(0.04)	12.80(4.76)	5.25(3.32)	39.07(6.21)	83.51(9.68)
	0.3	143.90	1.59	-22.05	0.31(0.10)	0.07(0.04)	12.05(4.30)	5.81(3.36)	39.07(6.21)	83.51(9.68)
3.	0.5	143.12	1.59	-22.82	0.30(0.09)	0.07(0.04)	11.61(4.01)	6.19(3.42)	39.07(6.21)	83.51(9.68)
	1	140.58	1.57	-25.37	0.32(0.11)	0.07(0.04)	12.47(4.80)	5.83(3.32)	39.07(6.21)	83.51(9.68)
DNA-P200/40										
1.	0.05	34.34	0.71	-19.69	0.104(0.003)	0.001(0.001)	4.21(0.34)	0.12(0.13)	40.61(2.96)	93.90(3.41)
	0.3	34.34	0.71	-19.69	0.102(0.002)	0.001(0.001)	4.14(0.32)	0.10(0.09)	40.61(2.96)	93.90(3.41)
	0.5	34.32	0.71	-19.70	0.101(0.002)	0.001(0.001)	4.10(0.31)	0.10(0.09)	40.61(2.96)	93.90(3.41)
	1	34.37	0.71	-19.66	0.099(0.003)	0.001(0.001)	4.00(0.31)	0.10(0.10)	40.61(2.96)	93.90(3.41)
	0.05	45.95	0.74	-8.07	0.32(0.06)	0.11(0.03)	14.44(3.16)	10.55(2.75)	45.78(3.69)	99.36(1.56)
2.	0.3	46.19	0.74	-7.83	0.32(0.05)	0.11(0.02)	14.40(2.65)	10.58(2.06)	45.64(3.86)	99.34(1.60)
	0.5	46.14	0.74	-7.88	0.32(0.06)	0.11(0.02)	14.54(2.91)	10.46(2.41)	46.05(3.87)	99.35(1.38)
	1	46.00	0.74	-8.02	0.31(0.06)	0.11(0.02)	13.98(3.16)	10.64(1.85)	45.41(4.11)	99.40(1.08)
	0.05	45.54	0.68	-8.49	0.27(0.08)	0.08(0.02)	10.78(3.27)	7.69(2.18)	40.61(2.96)	93.90(3.41)
	0.3	50.62	0.74	-3.41	0.27(0.08)	0.08(0.03)	11.10(3.48)	7.36(2.75)	40.61(2.96)	93.90(3.41)
3.	0.5	51.05	0.77	-2.98	0.26(0.08)	0.08(0.03)	10.67(3.24)	7.56(2.60)	40.61(2.96)	93.90(3.41)
	1	50.76	0.77	-3.26	0.27(0.08)	0.08(0.03)	10.84(3.42)	7.33(2.70)	40.61(2.96)	93.90(3.41)

**Tabela 15**

Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo HMS<sup>+</sup>. Mostramos os resultados para as sequências em H<sub>2</sub>O e PEG200 para quatro valores diferentes de  $y_e$ .  $\Delta\chi^2 = \chi^2_{(HMS^+)} - \chi^2_{(PB)}$ , onde  $\chi^2_{(PB)}$  se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos  $f_s$ , 2. Variar os termos  $f_s$  e os termos do potencial de Morse  $D$  e 3. Variar os termos  $f_s$  e os termos de empilhamento  $k$ .

um padrão ao variar as distâncias  $y_e$  no modelo HMS<sup>-</sup> (tabela 14). Em outras palavras, o modelo HMS não consegue distinguir entre moléculas de água a diferentes distâncias do nucleotídeo. Então, prosseguimos com as minimizações apenas para o valor  $y_e = 3 \text{ \AA}$ . Escolhemos esse valor, pois é onde se concentra a primeira casca de hidratação do DNA [3]. A primeira casca de hidratação é a região mais importante na estabilidade dos nucleotídeos e responsável por vários processos de interação entre o DNA e enzimas [3, 65, 69]. Na próxima seção, avaliaremos a possibilidade de utilizar o modelo PCLA para prever temperaturas de desnaturação.

**Modelo PCLA** A tabela 16 apresenta os resultados dos testes para o modelo PCLA. Os valores representados por traços são dos cálculos que não convergiram. Quando temos o valor do parâmetro mas não o erro, isso significa que apenas uma das 100 rodadas de minimizações conseguiu finalizar os cálculos. Cálculo finalizado significa que o *simplex*, neste caso um triângulo, não encontra mudanças significativas no valor de  $\chi^2$  e encerra o processo de minimização. No modelo PCLA, mesmo nos testes 2. e 3., uma grande parte das 100 rodadas de minimização não foi concluída. Nos testes 2., aproximadamente 60 rodadas convergiram, enquanto nos testes 3. apenas 15 rodadas convergiram. Com essa análise inicial, fica evidente que o modelo PCLA provavelmente não é capaz de prever temperaturas de desnaturação.

Ao continuar a análise dos resultados, vemos que nenhum dos valores de  $\Delta\chi^2 = \chi^2_{(PCLA)} - \chi^2_{(PB)}$  é negativo, conforme mostrado na tabela 16. Isso indica que o modelo PCLA não conseguiu

DNA-W/40								
Min.	$d$	$\chi^2$ (°C <sup>2</sup> )	$\langle \Delta T \rangle$ (°C)	$\Delta\chi^2$ (°C <sup>2</sup> )	$b_{AT}$ (eV/Å <sup>3</sup> )	$b_{CG}$ (eV/Å <sup>3</sup> )	$D_{AT}$ (meV)	$D_{CG}$ (meV)
1.	0.05	191.01	1.78	25.07	66.54(-)	0.15(-)	39.07(6.21)	83.51(9.68)
	0.3	203.25	1.86	37.30	28.95(-)	0.1(-)	39.07(6.21)	83.51(9.68)
	0.5	-	-	-	-	-	39.07(6.21)	83.51(9.68)
	1	-	-	-	-	-	39.07(6.21)	83.51(9.68)
2.	0.05	166.20	1.62	0.26	7.96(1.97)	4.48(0.97)	56.45(7.51)	56.73(5.27)
	0.3	166.03	1.62	0.08	8.18(1.65)	4.39(1.04)	58.24(8.25)	58.09(5.55)
	0.5	166.81	1.62	0.87	8.42(1.99)	4.55(1.02)	56.39(9.85)	56.89(6.47)
	1	166.44	1.61	0.50	8.68(1.42)	4.37(0.94)	55.60(7.85)	56.82(5.20)
3.	0.05	280.24	2.23	114.29	7.41(2.29)	4.00(1.03)	39.07(6.21)	83.51(9.68)
	0.3	321.02	2.42	155.07	8.91(2.68)	5.64(2.25)	39.07(6.21)	83.51(9.68)
	0.5	300.54	2.35	134.60	8.90(3.73)	5.79(3.07)	39.07(6.21)	83.51(9.68)
	1	312.97	2.38	147.02	8.87(5.69)	4.90(2.45)	39.07(6.21)	83.51(9.68)
DNA-P200/40								
1.	0.05	-	-	-	-	-	40.61(2.96)	93.90(3.41)
	0.3	-	-	-	-	-	40.61(2.96)	93.90(3.41)
	0.5	-	-	-	-	-	40.61(2.96)	93.90(3.41)
	1	177.33	1.67	123.30	2.98(0.19)	6.919(0.001)	40.61(2.96)	93.90(3.41)
2.	0.05	78.72	1.07	24.69	6.22(2.47)	5.72(2.25)	42.71(12.60)	54.68(8.94)
	0.3	75.44	1.07	21.42	5.48(2.46)	4.49(2.57)	39.42(18.32)	54.50(11.51)
	0.5	76.92	1.07	22.90	5.89(2.76)	5.18(2.55)	39.35(18.48)	53.49(11.87)
	1	78.60	1.08	24.57	6.53(2.39)	5.63(2.32)	35.53(13.79)	51.14(8.19)
3.	0.05	115.48	1.45	61.46	8.96(2.94)	6.70(2.38)	40.61(2.96)	93.90(3.41)
	0.3	114.16	1.45	60.14	7.64(2.39)	6.18(1.96)	40.61(2.96)	93.90(3.41)
	0.5	140.68	1.57	86.65	7.73(3.35)	5.58(3.37)	40.61(2.96)	93.90(3.41)
	1	151.45	1.60	97.42	7.30(3.87)	5.50(2.41)	40.61(2.96)	93.90(3.41)

**Tabela 16**

Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo PCLA. Mostramos os resultados para as sequências em H<sub>2</sub>O e PEG200 para quatro valores diferentes de  $d$ .  $\Delta\chi^2 = \chi^2_{(PCLA)} - \chi^2_{(PB)}$ , onde  $\chi^2_{(PB)}$  se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos  $b$ , 2. Variar os termos  $b$  e os termos do potencial de Morse  $D$  e 3. Variar os termos  $b$  e os termos de empilhamento  $k$ .

melhorar as previsões de temperatura encontradas pelo modelo PB. Dessa forma, O modelo PCLA não atendeu ao nosso critério de avaliação número I. As energias de barreira  $b_{AT}$  e  $b_{CG}$  para as sequências DNA-W/40 apresentam valores absolutos de  $b_{AT} > b_{CG}$ . Apenas nos testes 3., se considerarmos a barra de erro,  $b_{CG}$  pode ser maior que  $b_{AT}$ . Os potenciais de Morse do teste 3. do conjunto DNA-P200/40 são maiores do que os do conjunto DNA-W/40. Logo, os testes 3. foram aprovados nos critérios II. e III..

Apesar de respeitar os critérios II. e III., o modelo não apresenta previsões de temperatura melhores que o modelo PB, não atendendo assim ao critério I. Além disso, uma parte considerável das 100 rodadas do teste 3. não conseguiu finalizar os cálculos e a função partição divergiu. Um dos motivos para a divergência da função partição é que o termo  $y^3$  da equação (35) gera uma barreira de potencial enorme. Outro problema é que os autores Peyrard et al. [25] definem significado físico aos parâmetros  $c$  e  $d$  que na verdade não existem, pois, de acordo com a equação (35),  $c$  e  $d$  são apenas números sem unidade de medida. Diante disso, identificamos problemas nesse modelo e não prosseguiremos com mais rodadas de minimização.

Decidimos modificar a equação de barreira do modelo PCLA para ver se corrigimos os

problemas de divergência. A segunda função da equação (35) foi alterada da seguinte forma:

$$V(y_n) = \Theta(y_n) \frac{by_n^p}{\cosh^2[c(\alpha y_n^2 - d \ln 2)]} \quad (39)$$

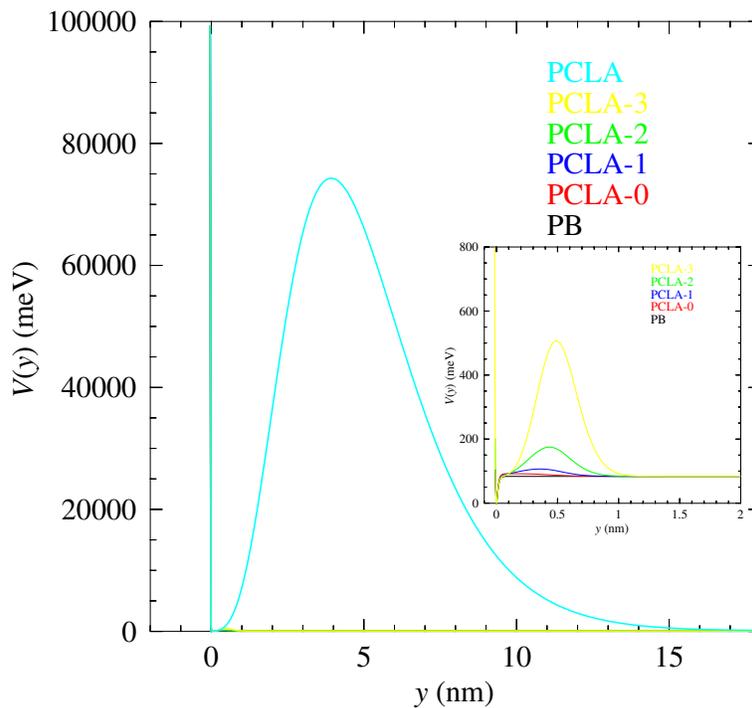
Com o termo  $y^2$  no denominador, conseguimos atribuir unidades de medida para os parâmetros  $c$  e  $d$  iguais as descritas pelo artigo Peyrard et al. [25].  $c$  possui unidade de  $\text{Å}^{-1}$  e  $d$  possui unidade de  $\text{Å}$ . Essa modificação não apenas reduz a altura da barreira em relação ao modelo original (figura 38), mas também atribuí significado físico aos parâmetros  $c$  e  $d$ . O expoente  $p$  no numerador servirá para observarmos os seguintes casos:

- Caso  $y^3$ : Modelo PCLA com  $c$  e  $d$  tendo realidade física (PCLA-3);
- Caso  $y^2$ : Assemelha-se ao potencial harmônico. O estudo realizado por Luo et al. [152] fez cálculos de DFT para observar como as moléculas de  $\text{H}_2\text{O}$  se arranjam na estrutura base- $\text{H}_2\text{O}$ -base. As águas se organizam em forma helicoidal, como uma mola molecular. Caso as previsões de temperatura sejam melhores ou semelhantes ao modelo PB, isso abriria caminho para desenvolver uma barreira de potencial com esta modificação (PCLA-2);
- Caso  $y^1$ : O modelo PCLA com essa modificação, a menos dos diferentes parâmetros, seria à derivada primeira do modelo HMS. Caso as previsões de temperatura sejam tão boas quando as do modelo PB abriria caminho para incluir o segundo termo da expansão em série de Taylor do modelo HMS. Esse segundo termo da expansão em série descreveria a força entre o nucleotídeo e o par de base, já que 90% das interações entre o nucleotídeo e a  $\text{H}_2\text{O}$  são de origem elétrica (PCLA-1);
- Caso  $y^0$ : Avaliar o modelo PCLA modificado com  $b$  tendo unidades de energia (PCLA-0).

Note que a medida que o valor de  $p$  diminui, a barreira de energia também diminui, como mostrado na figura 38.

Plotamos o mapa de calor para os quatro caso, avaliados apenas em  $d = 0.3$  nm, com os mesmos parâmetros do modelo PCLA descritos na tabela 12. Isso foi feito porque não observamos nenhuma dependência entre as energias de barreira e o parâmetro  $d$  no modelo PCLA. Apesar disso, faremos uma verificação futura se existe alguma relação entre as energias de barreira e o parâmetro  $d$  das modificações do modelo PCLA. Os modelos foram chamados de PCLA-3, PCLA-2, PCLA-1 e PCLA-0. O resumo dos parâmetros do modelo PCLA modificado está na tabela 17.

Os resultados encontrados foram surpreendentes, conforme ilustrado figuras 39 e 40. Apesar de nenhum dos valores de  $\langle \chi^2 \rangle$  dos modelos modificados apresentar temperaturas de predição melhor



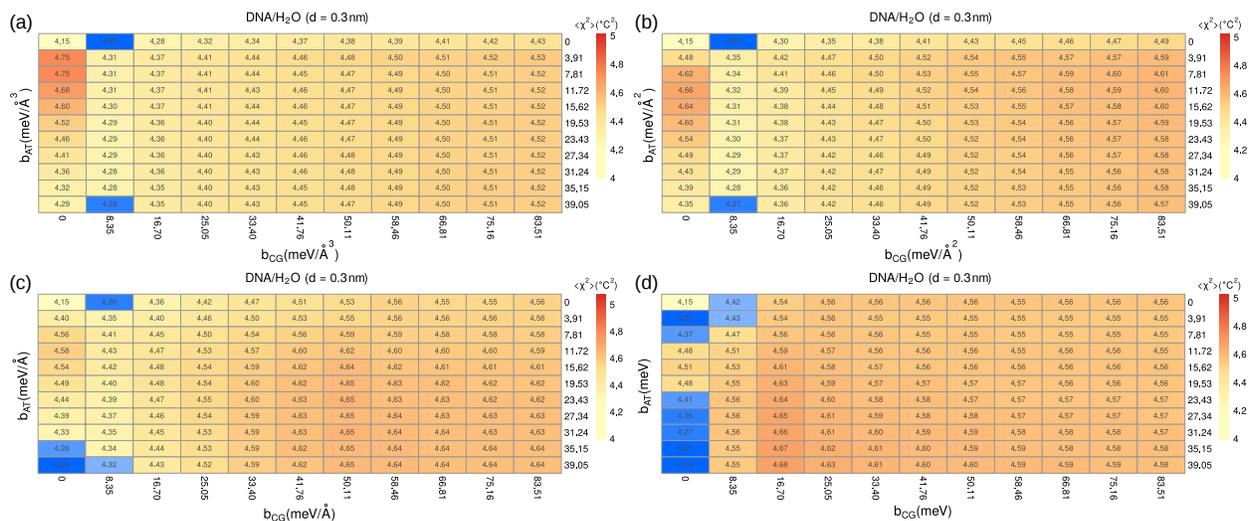
**Figura 38**

Comparação entre os potenciais do modelo PCLA, equação (35), e PCLA modificados, equação (39). PCLA-3 ( $p = 3$ ), PCLA-2 ( $p = 2$ ), PCLA-1 ( $p = 1$ ) e PCLA-0 ( $p = 0$ ). Os parâmetros de todas as curvas são:  $D=83.51$  meV;  $\lambda=0.11$  Å;  $b=8.35$  meV/Å<sup>p</sup>,  $b=8.35$  meV/Å<sup>3</sup> (PCLA);  $c=4.4 \times 10^{-3}$  Å<sup>-3</sup>;  $\alpha=9.39$  Å<sup>-1</sup>

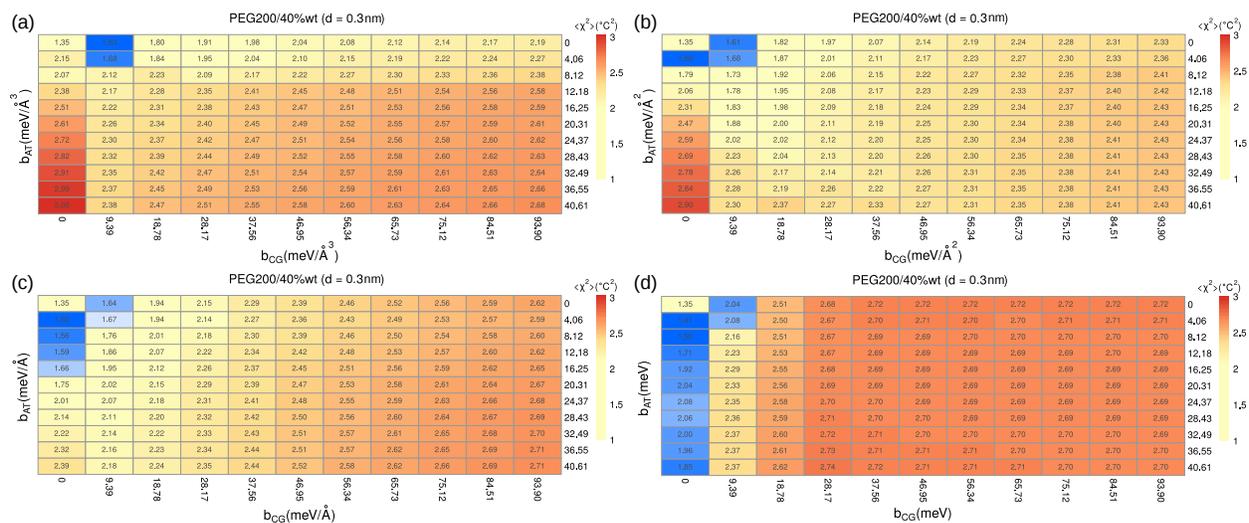
Hamiltoniano	parâmetro	unidade
$D(e^{-y_i/\lambda}-1)^2$	$D$	eV
	$\lambda$	Å
	$b$	eV/Å <sup>p</sup>
$by_i^p / \cosh^2[c(\alpha y_i^2 - d \ln 2)]$	$c$	Å <sup>-1</sup>
	$d$	Å
	$\alpha$	Å <sup>-1</sup>
$\frac{k}{2}(y_i^2 - 2y_i y_{i-1} \cos \theta + y_{i-1}^2)$	$k$	eV/Å <sup>2</sup>
	$\theta$	rad

**Tabela 17**

Parâmetros para o modelo PCLA modificado.



**Figura 39**  
 Modelo PCLA modificado. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $5^{\circ}C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a) PCLA-3, (b) PCLA-2, (c) PCLA-1, (d) PCLA-0.



**Figura 40**  
 Modelo PCLA modificado. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $3^{\circ}C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a) PCLA-3, (b) PCLA-2, (c) PCLA-1, (d) PCLA-0.

do que o PB, as predições de temperatura se reduziram à metade do que foi apresentado pelo modelo PCLA, como visto nas figuras 49 e 50. Os valores de  $\langle \chi^2 \rangle$  fora do eixo zero  $b_{AT}$  e zero  $b_{CG}$  são muito próximos aos do modelo PB. Além disso, o intervalo de  $\langle \chi^2 \rangle$  mostrado pela escala de cores possui valores muito próximos aos do modelo PB. O que nos encoraja a dizer que o modelo PCLA com as modificações propostas conseguirá predizer temperaturas de desnaturação melhores do que o modelo PB. Outro fato que observamos é que as predições de temperatura pouco melhoram à medida que  $p$  diminui, de  $p=3$  (figuras 39a e 40a) até  $p=0$  (figuras 39d e 40d). De forma análoga ao modelo PCLA original, os melhores valores de  $\langle \chi^2 \rangle$  são aqueles que possuem  $b_{AT}$  muito alto no caso com  $H_2O$ . Apenas na figura 39d isso não acontece. Porém, observamos que a diferença entre os valores de  $\langle \chi^2 \rangle$  na coluna  $b_{CG} = 8.35 \text{ meV}/\text{Å}^p$ , onde  $p$  varia de 0 a 3, são bem pequenas. Portanto, podemos calcular as minimizações a partir de qualquer um desses pontos. De modo a manter a mesma referência entre os modelos modificados, escolheremos o ponto (2,2), que é o mesmo ponto do melhor  $\langle \chi^2 \rangle$  das sequências em PEG200, como indicado na figura 40. Infelizmente, não tivemos tempo de proceder com as análises. Deixaremos essa análise como uma perspectiva futura. Na próxima seção, avaliaremos a possibilidade de utilizar o modelo TRMF para predizer temperaturas de desnaturação.

**Modelo TRMF** A tabela 18 apresenta os resultados dos testes para o modelo TRMF. Ao contrário do modelo PCLA, todas as 100 rodadas de todos os testes convergiram. O valor dos parâmetros de energia de barreira e seus erros são a média e o desvio padrão das 100 rodadas. Vamos avaliar se o modelo TRMF respeita os critérios que definimos para prosseguir com as minimizações.

O primeiro critério nos informa que o  $\chi^2$  do modelo TRMF deve ser menor do que o do modelo PB. Pela tabela 18, coluna  $\Delta\chi^2$ , observamos que isso acontece apenas no teste 2. em  $H_2O$  (DNA-W/40) e nos testes 2. e 3. em PEG200 (DNA-P200/40). Alguns dos testes do modelo TRMF conseguiram passar em nosso critério número I.. O critério II. diz que a diferença entre  $G_{AT}$  e  $G_{CG}$  deve ser pequena, e  $G_{CG}$  deve ser levemente maior do que  $G_{AT}$  no caso em  $H_2O$ , sequências DNA-W/40. Apenas no teste 3 isso acontece, como indicado na tabela 18. Se avaliarmos a barra de erro do teste 2, observamos que  $G_{CG}$  pode ser maior do que  $G_{AT}$ , o que não acontece no teste 1. Contudo, nossa experiência com as minimizações nos diz que o valor do parâmetro deve oscilar em torno do valor encontrado nos cálculos anteriores e reduzir o erro, ou seja, desvio padrão entre todas as rodadas. Assim, acreditamos que após novas rodadas de minimizações,  $G_{AT}$  será maior do que  $G_{CG}$  com uma barra de erro bem pequena. Apenas o teste 3 foi aprovado no critério II..

O último critério analisado refere-se às profundidades de poço  $D$ . Além de  $D_{CG} > D_{AT}$ ,  $D$  das sequências em  $H_2O$  (DNA-W/40) deve ser menor do que nas sequências em PEG200 (DNA-P200/40). Todos os testes respeitam o critério III., conforme mostrado na tabela 18.

DNA-W/40								
Min.	$y_o$ (nm)	$\chi^2$ ( $^{\circ}\text{C}^2$ )	$\langle\Delta T\rangle$ ( $^{\circ}\text{C}$ )	$\Delta\chi^2$ ( $^{\circ}\text{C}^2$ )	$G_{\text{AT}}$ (meV)	$G_{\text{CG}}$ (meV)	$D_{\text{AT}}$ (meV)	$D_{\text{CG}}$ (meV)
1.	0.05	166.26	1.69	0.31	0.22(0.34)	0.02(0.03)	39.07(6.21)	83.51(9.68)
	0.3	166.28	1.69	0.33	0.45(0.54)	0.02(0.02)	39.07(6.21)	83.51(9.68)
	0.5	166.50	1.69	0.55	0.85(0.61)	0.02(0.02)	39.07(6.21)	83.51(9.68)
	1	166.19	1.69	0.25	1.17(0.26)	0.02(0.02)	39.07(6.21)	83.51(9.68)
2.	0.05	164.57	1.67	-1.38	8.14(11.85)	1.58(2.27)	48.01(7.89)	92.88(6.55)
	0.3	163.10	1.66	-2.84	8.73(12.51)	1.57(2.12)	48.53(6.76)	93.58(5.80)
	0.5	158.13	1.63	-7.82	5.08(7.24)	1.55(1.92)	46.90(7.63)	91.86(6.07)
	1	163.56	1.64	-2.38	13.93(23.06)	4.53(4.81)	49.96(7.33)	94.98(6.71)
3.	0.05	173.43	1.74	7.49	4.41(4.37)	3.55(3.56)	39.07(6.21)	83.51(9.68)
	0.3	183.49	1.77	17.54	4.29(3.61)	4.94(4.96)	39.07(6.21)	83.51(9.68)
	0.5	189.96	1.80	24.02	6.24(8.58)	5.27(4.77)	39.07(6.21)	83.51(9.68)
	1	186.97	1.77	21.03	7.10(6.66)	7.29(4.93)	39.07(6.21)	83.51(9.68)
DNA-P200/40								
1.	0.05	55.32	0.82	1.30	1.21(0.32)	0.20(0.08)	40.61(2.96)	93.90(3.41)
	0.3	55.44	0.82	1.41	1.07(0.50)	0.24(0.11)	40.61(2.96)	93.90(3.41)
	0.5	55.38	0.82	1.36	0.96(0.71)	0.26(0.15)	40.61(2.96)	93.90(3.41)
	1	54.95	0.81	0.93	2.06(0.54)	0.63(0.17)	40.61(2.96)	93.90(3.41)
2.	0.05	46.36	0.76	-7.67	3.77(1.59)	6.33(2.18)	47.50(5.96)	104.98(5.16)
	0.3	46.21	0.76	-7.81	3.76(1.30)	6.61(2.42)	47.61(3.86)	105.19(3.01)
	0.5	46.50	0.76	-7.53	3.72(1.52)	6.45(2.58)	46.20(3.22)	104.14(3.15)
	1	46.12	0.74	-7.91	4.23(1.82)	6.78(2.78)	43.28(3.17)	101.31(2.22)
3.	0.05	48.55	0.75	-5.48	3.40(1.75)	1.95(2.26)	40.61(2.96)	93.90(3.41)
	0.3	34.57	0.70	-19.45	3.84(1.73)	2.03(2.78)	40.61(2.96)	93.90(3.41)
	0.5	44.34	0.66	-9.69	3.87(1.70)	2.28(2.84)	40.61(2.96)	93.90(3.41)
	1	43.87	0.66	-10.15	4.47(1.65)	3.76(2.69)	40.61(2.96)	93.90(3.41)

**Tabela 18**

Comparação entre os três testes para ver qual o melhor processo de minimização para se trabalhar com o modelo TRMF. Mostramos os resultados para as sequências em H<sub>2</sub>O e PEG200 para quatro valores diferentes de  $y_o$ .  $\Delta\chi^2 = \chi^2_{(TRMF)} - \chi^2_{(PB)}$ , onde  $\chi^2_{(PB)}$  se encontra na seção 5.5.2, tabela 11. 1. Variar apenas os termos  $G$ , 2. Variar os termos  $G$  e os termos do potencial de Morse  $D$  e 3. Variar os termos  $b$  e os termos de empilhamento  $k$ .

Em geral, os testes 3 do modelo TRMF conseguiram prever temperaturas de desnaturação e explicar o que está acontecendo no sistema. No entanto, as previsões de temperatura são piores do que as do modelo PB, com  $\Delta\chi^2 > 0$  (tabela 18). O modelo PB explica muito bem o que acontece no sistema e se aproxima mais das temperaturas experimentais. Portanto, não realizaremos novas rodadas de minimizações com o modelo TRMF. Na próxima seção, avaliaremos a possibilidade de utilizar o modelo PCLJ para prever temperaturas de desnaturação.

**Modelo PCLJ** O modelo PCLJ possui equações acopladas para descrever o potencial  $V_{\text{Morse}} + V_{\text{barreira}}$ , conforme as equações (36) e (37). Mudanças nos termos de barreira geram novos valores de profundidade de poço  $A$ . Logo, não é possível avaliar o teste 1. Os demais testes são possíveis; porém, no teste 3, devido ao acoplamento das equações, variamos os termos de Morse, barreira e empilhamento. Iniciamos as minimizações pelo mesmo valor de  $D_h(\text{AT}) = D_h(\text{CG}) = 1$  meV. No modelo PCLJ,  $D_h$  representa a energia das fitas simples com as demais moléculas presentes na solução. Como esperado, encontramos valores diferentes entre essas duas energias após as rodadas de minimizações, conforme apresentado na tabela 19.

Calculamos  $A_{\text{AT}}$  e  $A_{\text{CG}}$  pelas equações (37) e utilizamos a propagação de erro por derivadas para estimar os erros de  $A_{\text{AT}}$  e  $A_{\text{CG}}$ .  $A$  é equivalente ao  $D$  no modelo PB. Apesar das enormes diferenças entre  $A_{\text{AT}}$  e  $A_{\text{CG}}$ , o modelo consegue descrever que a energia dos pares CG é maior do que a dos pares AT, como mostrado na tabela 19. No entanto, os erros são muito maiores do que nos demais modelos.

Das 100 rodadas de minimizações dos dois testes, poucas terminaram os cálculos da função partição. Nas rodadas com PEG200, DNA-P200/40, o máximo de rodadas finalizadas foram 8 (testes 2. e 3.), e em H<sub>2</sub>O, DNA-W/40, foram 67 rodadas (teste 2.) e 40 rodadas (teste 3.). Os valores e os erros dos parâmetros  $E$  e  $D_h$  mostrados na tabela 19 representam a média e o desvio padrão das rodadas finalizadas. O número de rodadas finalizadas muito inferior às 100 propostas acontece porque iniciamos o processo de minimização com  $\langle\chi\rangle$  muito ruins, isto é, intersecção entre os retângulos verdes das figuras 53 e 54. Quando iniciamos as minimizações por  $\langle\chi\rangle$  melhores, nos quais  $E_{\text{CG}} > E_{\text{AT}}$ , conforme figuras 53 e 54, chegamos às mesmas conclusões apresentadas abaixo, porém com valores  $E$ ,  $D_h$  e  $A$  diferentes (dados não apresentados).

De acordo com nossos critérios, descritos no início da seção 5.6.4, o modelo PCLJ não consegue prever temperaturas de desnaturação melhores do que o modelo PB. Os motivos são os seguintes:

1. As previsões de temperatura do modelo PCLJ são muito ruins,  $\Delta\chi^2 \gg 0$ , mostrada na tabela 19;

DNA-W/40									
Min.	$\chi^2$ ( $^{\circ}\text{C}^2$ )	$\langle\Delta T\rangle$ ( $^{\circ}\text{C}$ )	$\Delta\chi^2$ ( $^{\circ}\text{C}^2$ )	$D_h$ (AT) (meV)	$D_h$ (CG) (meV)	$E_{\text{AT}}$ (meV/Å)	$E_{\text{CG}}$ (meV/Å)	$A_{\text{AT}}$ (meV)	$A_{\text{CG}}$ (meV)
2.	482.28	2.65	316.34	0.81(1.68)	0.96(1.16)	21.33(16.60)	33.37(13.98)	77.30(60.16)	232.56(97.43)
3.	461.45	2.58	295.50	0.80(0.49)	0.78(0.43)	9.28(7.10)	20.65(5.67)	33.64(25.72)	143.93(39.48)
DNA-P200/40									
2.	303.60	2.03	249.57	0.80(0.67)	1.05(0.51)	4.56(1.62)	14.46(3.35)	16.54(5.88)	100.83(23.32)
3.	264.68	1.79	210.66	0.74(0.49)	0.81(0.43)	8.73(7.32)	17.22(10.54)	31.66(26.51)	120.01(73.43)

**Tabela 19**

Comparação entre os dois testes para ver qual o melhor processo de minimização para se trabalhar com o modelo PCLJ.  $\Delta\chi^2 = \chi^2_{(PCLJ)} - \chi^2_{(PB)}$ , onde  $\chi^2_{(PB)}$  se encontra na seção 5.5.2, tabela 11. 2. Variar apenas os termos  $E$  e  $D_h$ , 3. Variar os termos  $E$ ,  $D_h$  e os termos de empilhamento  $k$ .

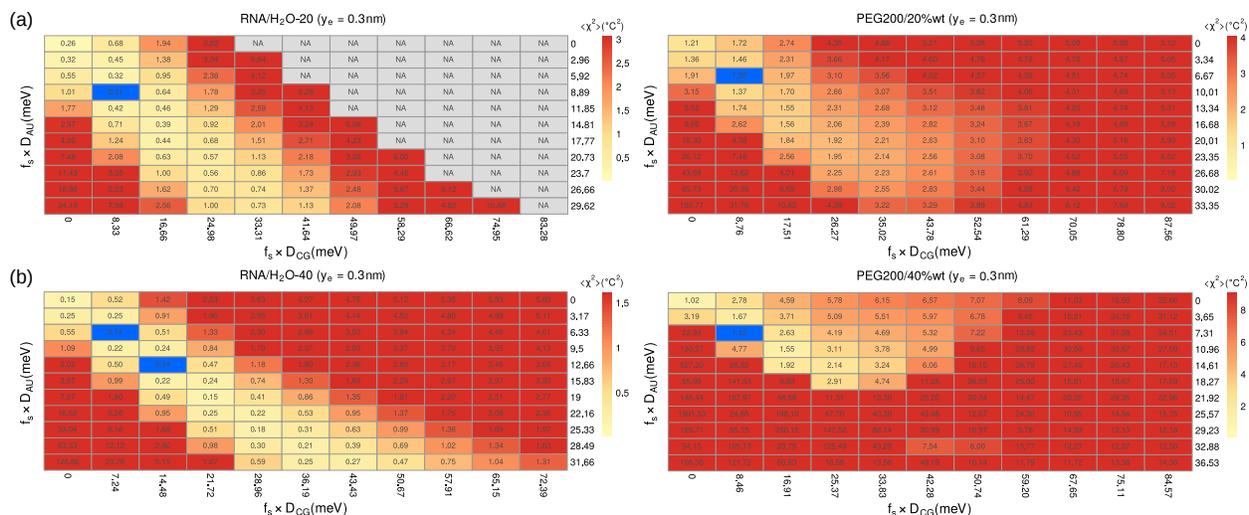
2. Apesar de  $E_{\text{CG}}$  ser maior do que  $E_{\text{AT}}$  em  $\text{H}_2\text{O}$ , DNA-W/40, os valores de  $E_{\text{CG}}$  e  $E_{\text{AT}}$  não são próximos, como indicado na tabela 19;
3. As profundidades de poço do potencial  $A$  são bem maiores em  $\text{H}_2\text{O}$ , DNA-W/40, do que em PEG200, DNA-P200/40. O PEG200 reduz a interação entre as moléculas de água e os nucleotídeos. Como consequência, a interação entre as bases aumenta, o que leva a maiores diferenças na intensidade de ligação entre os pares CG e AT. O efeito prático disso resulta em uma mudança na conformação global da hélice de DNA.

O modelo PCLJ foi reprovado em todos os nossos critérios de avaliação. Portanto, não realizaremos novas rodadas de minimizações.

### 5.6.5 Resultados

Dentre os critérios utilizados para se avaliar os modelos, apenas o modelo  $\text{HMS}^-$  foi aprovado. O único teste que obteve sucesso, conforme os nossos critérios, foi o teste 3., no qual variamos os parâmetros de empilhamento e energia de barreira. Nos próximos parágrafos, discutiremos os resultados encontrados.

**HMS<sup>-</sup>** Nessa etapa, incluímos as sequências de RNA e utilizamos a mesma metodologia empregada para avaliar o modelo  $\text{HMS}^-$  para as sequências de DNA, como descrito na seção B. Plotamos os mapas de calor das sequências de RNA em PEG200, RNA-P200/20 e RNA-P200/40, e de suas respectivas sequências calculadas em  $\text{H}_2\text{O}$ , RNA-W/20 e RNA-W/40, mostrado nas figuras 41. Análogo ao que foi feito para as sequências de DNA, usamos os parâmetros  $D$  e  $k$  minimizados pelo modelo PB como valores iniciais das sequências de RNA, conforme detalhado na seção 5.5.2. Já os parâmetros  $y_e$  e  $\lambda_s$ , foram os mesmos valores mostrados na tabela 12, trocando AT por AU. Para  $y_e$ , nos restringimos ao valor de 0.3 nm (região onde estão a maior parte das moléculas de  $\text{H}_2\text{O}$  responsáveis pela estabilidade dos nucleotídeos [3, 65]), pois não observamos uma dependência com a energia da barreira para as sequências de DNA. O alcance descrito pelo parâmetro  $\lambda_s$  deveria ser



**Figura 41**

Modelo HMS<sup>-</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AU e CG em  $y_e = 0.3$  nm. Colorimos todos os pontos acima do valor máximo mostrado nas escalas de cores de  $\langle \chi^2 \rangle$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). os valores NA no mapa de calor são de cálculos que geraram valores infinitos.

similar aos valores de DNA, pois ambos, DNA e RNA, são moléculas altamente polares que possuem grupos funcionais semelhantes, a menos de algumas diferenças pontuais, como o grupo metil, CH<sub>3</sub>, presente na Timina e ausente na Uracila, como ilustrado na figura 2. Porém, o termo de barreira descreve interações dos nucleotídeos com moléculas de água (ligações de hidrogênio e interações eletrostáticas) entre os pares de base AT, AU e CG, que possuem exatamente os mesmos grupos funcionais. Portanto, assumimos que o mesmo comportamento deva ocorrer com as sequências de RNA.

A partir do mapa de calor da figura 41, determinamos os valores de  $f_{sAU}$  e  $f_{sCG}$  para iniciar o processo de minimização. Só recapitulando: o intervalo de  $f_s$  está entre 0 e 1 com passo de 0.1. Os pontos em azul nas figuras são ótimos valores para iniciarmos as otimizações dos parâmetros, pois apresentam energias de barreira próximas, onde  $f_{sCG} D$  é levemente maior do que  $f_{sAU} D$ . No entanto, na figura 41b, temos dois pontos. Por causa da aproximação até segunda casa decimal, não é possível distinguir os dois pontos em azul. A diferença entre os dois pontos está na quarta casa decimal. Selecionamos o ponto com  $\langle \chi^2 \rangle$  levemente menor que é o ponto  $f_{sCG} = 0.1$  e  $f_{sAU} = 0.2$ .

Após três rodadas de minimizações e uma extra para avaliar a influência do erro experimental nos resultados, encontramos os parâmetros de qualidade das minimizações do modelo HMS<sup>-</sup>, conforme apresentado na tabela 20 para todos os conjuntos de dados (DNA-W/40, DNA-P200/40, RNA-W/20, RNA-P200/20, RNA-W/40 e RNA-P200/40). Observamos que os valores de  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$  reduziram muito pouco para as sequências de RNA. Ao comparar com os valores obtidos pelo

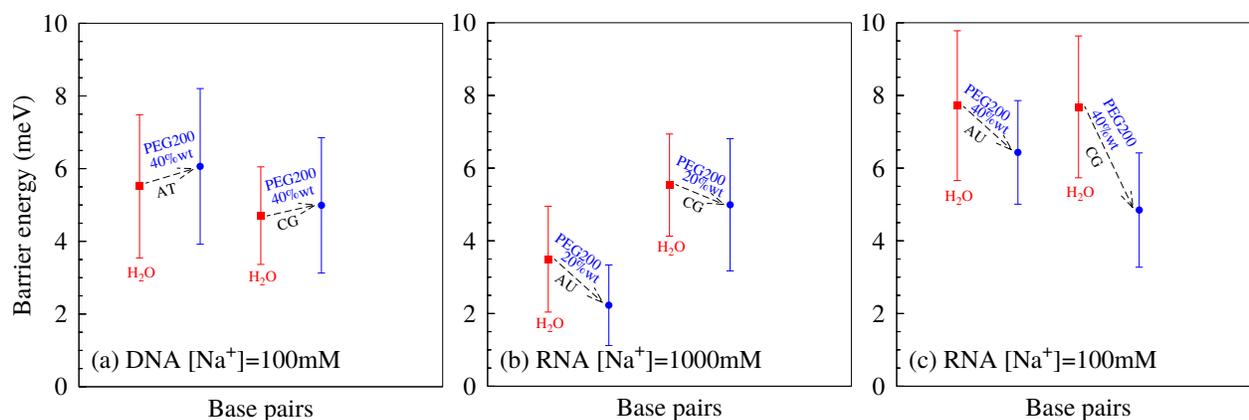
Conjunto de dados	regressão por tamanho da sequência			
	após a primeira minimização (mapa de calor)		depois da última minimização	
	$\langle \Delta T \rangle$ (°C)	$\langle \chi^2 \rangle$ (°C <sup>2</sup> )	$\langle \Delta T \rangle$ (°C)	$\langle \chi^2 \rangle$ (°C <sup>2</sup> )
DNA-W/40	1.75	4.43	1.39	2.87
DNA-P200/40	0.86	1.43	0.68	0.80
RNA-W/20	0.44	0.31	0.41	0.27
RNA-P200/20	0.98	1.32	0.92	1.21
RNA-W/40	0.29	0.14	0.29	0.14
RNA-P200/40	0.88	1.13	0.81	1.04

**Tabela 20**

Comparação entre os  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$  encontrados pelos mapas de calor, figuras 45b e 46b, e após a última minimização na presença e ausência de PEG200. Para ficar mais fácil de compararmos os resultados, normalizamos o valor de  $\chi^2$  dividindo-o pelo número de sequências  $N$  mostrados na tabela 9.

modelo PB, tabela 11, observamos uma pequena ou nenhuma melhora nos valores de  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$ . Diferente do DNA, cujo os parâmetros de qualidade melhoraram bastante em relação ao modelo PB. Apesar disso, os valores de  $\langle \Delta T \rangle$  e  $\langle \chi^2 \rangle$  são iguais ou menores aos encontrados pelo modelo PB. Isso mostra que o modelo  $HMS^-$  é um bom modelo para prever temperaturas de desnaturação.

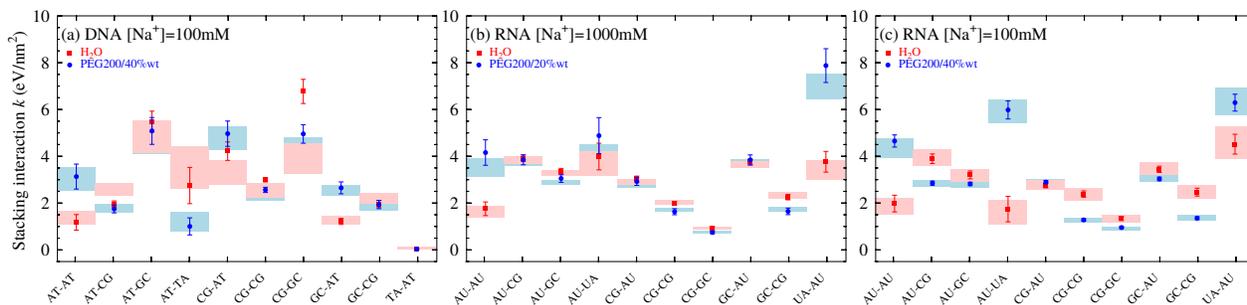
Nas figuras 42, apresentamos as energias de barreira,  $f_s D$ , dos pares de base AT, AU (para RNA) e CG. Nos gráficos, mostramos as mudanças da energia de barreira ao adicionar PEG200 nas soluções, representadas por uma seta. Todos os valores de energia são muito próximos em  $H_2O$ , como esperado. As pequenas diferenças de energia existem por causa da composição química dos nucleotídeos que são um pouco diferentes entre DNA e RNA. Para os nucleotídeos sujeitos à ação apenas de  $H_2O$  e sal, pontos em vermelho, observamos que a concentração de sal possui um papel muito importante. Em baixas concentrações de sal ( $[Na^+]=100$  mM), as barreiras de energia AT e CG, figura 42a, e AU e CG, figura 42c, são praticamente iguais. Apesar dos valores médios de  $f_s D$  serem um pouco diferentes em baixas concentrações de  $Na^+$  no DNA, pontos em vermelho da figura 42a, o intervalo de oscilação do valor dos parâmetros, barra de erro, torna as barreiras de energia AT e CG iguais. O que não acontece para altas concentrações de sal ( $[Na^+]=1000$  mM) CG é nitidamente maior do que AU, figura 42b. Há apenas uma única diferença entre os grupos funcionais presentes nos pares de base AT, AU e CG. A Guanina possui um grupo  $-NH_2$  ligado ao carbono C(4), enquanto a Adenina possui um grupo  $-H$ , conforme ilustrado na figura 2c. O grupo  $-NH_2$  forma mais ligações de hidrogênio com a água do que o grupo  $-H$ . Portanto, esperamos que a barreira de energia de CG seja maior do que AT ou AU para RNA. Essa diferença só foi observada na figura 42b, onde a concentração de sal ( $[Na^+]$ ) é  $10\times$  maior do que nos outros casos. Com mais moléculas de  $Na^+$  presentes na solução, a repulsão eletrostática entre a fita principal e complementar se torna menor, destacando as ligações entre o solvente, no caso  $H_2O$ , e os pares de base. Como resultado, o termo de barreira consegue distinguir as energias de transição de fita dupla para simples entre AU e CG, figura 42b, o que não acontece nos demais casos, figuras 42a e 42c.



**Figura 42**

Energias de barreira  $f_s D$  calculadas para DNA e RNA em H<sub>2</sub>O como pontos vermelhos e em PEG200 como pontos azuis. A seta tracejada indica a mudança na altura da barreira de potencial causada pela adição de PEG200 na solução. (a) DNA-W/40 e DNA-P200/40, (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40.

O efeito de PEG200 nas sequências de DNA é completamente diferente do que ocorre no RNA. Nas sequências de DNA, observamos um pequeno aumento na energia de barreira, enquanto no RNA, observamos uma redução. O duplexo de RNA, além de ser mais hidratado, acumula mais íons de sódio ao seu redor do que o DNA [4, 5, 34, 64]. PEG200 interage com as moléculas de água, diminuindo a interação das moléculas de água com os nucleotídeos [33, 48, 140]. No nosso caso, podemos estender essa interação para outras moléculas presentes no sistema, como os cátions Na<sup>+</sup>. Logo, o deslocamento das moléculas de água afeta diretamente ou indiretamente os nucleotídeos. O efeito direto seria o deslocamento das moléculas de H<sub>2</sub>O, enquanto o efeito indireto seria uma perturbação na solvatação dos íons de sódio, Na<sup>+</sup>. Em ambos os casos, água e sódio estariam próximos aos nucleotídeos. Vamos analisar primeiro as figuras 42a e 42c, que estão nas mesmas condições experimentais. A energia de barreira ser maior para o RNA do que no DNA nas sequências em H<sub>2</sub>O acontece porque tanto os cátions quanto as moléculas de água interagem mais com o RNA. PEG200 possui um efeito ligeiramente positivo tanto em AT e CG, figura 42a, porém, a variação em relação as sequências com H<sub>2</sub>O é a mesma. Ou seja, temos apenas um deslocamento de energia em igual proporção para cada par de base. Nesse caso, PEG200 tem o mesmo efeito em AT e CG no estado intermediário de transição da fita dupla para fita simples. O aumento da barreira de potencial com a adição de PEG200 pode ser por causa que tanto H<sub>2</sub>O quanto PEG200 e alguns íons de sódio têm uma contribuição maior na energia entre as fitas do que do lado de fora. Isso ocorre porque o duplexo de DNA é menos hidratado e acumula menos cátions ao seu redor do que no RNA [4, 5, 34, 64]. Na solução de água e sal, não temos PEG200 participando na energia da barreira, logo a barreira de potencial é menor. No RNA, os efeitos diretos e indireto são mais distinguíveis. Em meios hidrofóbicos, ou seja, soluções com menos volume de água disponível, o RNA possui



**Figura 43**

Interações de empilhamento médias  $k$  calculadas para DNA e RNA em  $H_2O$  como pontos vermelhos e em PEG200 como pontos azuis. A região sombreada representa o intervalo de parâmetros de empilhamento dos resultados do modelo PB apresentados no capítulo 5.5.2. (a) DNA-W/40 e DNA-P200/40, (b) RNA-W/20 e RNA-P200/20 e (c) RNA-W/40 e RNA-P200/40

uma barreira de potencial menor. Isso significa que o RNA se transforma mais facilmente em fita simples em meios com pouca água disponível. Isso é explicado pelo grande deslocamento da energia de barreira CG em relação a AU na presença de PEG200/40%wt, figura 42c. Infelizmente, no presente momento, o modelo não consegue distinguir se o efeito é direto (afetando a interação entre as moléculas de água e os nucleotídeos) ou indireto (afetando a interação entre os íons  $Na^+$  e os nucleotídeos). Tudo o que temos é um efeito acumulado, ou seja, ambos os efeitos direto e indireto definem a energia da barreira. Se as sequências das figuras 42b e 42c estivessem sob as mesmas concentrações de PEG200 ou sob as mesmas concentrações de sódio, teríamos como separar o efeito do sal ou do PEG200. No momento, isso não é possível.

Na figura 42b, observamos um deslocamento da energia de barreira do par AU com a adição de PEG200 igual ao da figura 42c, enquanto CG tem uma pequena diminuição de energia. Ambos os efeitos podem estar associados a alta concentração de sal,  $Na^+$ , que reduz à repulsão eletrostática entre as fitas e consequentemente aumenta a interação do solvente com os nucleotídeos entre os duplexos no estado intermediário entre a fita dupla e as fitas simples. Uma outra alternativa pode estar relacionada à baixa concentração de PEG200, que neste caso desloca menos moléculas de água ao redor dos nucleotídeos e possui um efeito menor na energia da barreira. Reiterando, seria interessante fazer essa análise com as sequências das figuras 42b e 42c sob as mesmas concentrações de PEG200 ou sob as mesmas concentrações de sódio para observamos com clareza os efeitos da água e do sódio.

Na figura 43, apresentamos os empilhamentos  $k$  ajustados pelo modelo  $HMS^-$ . As regiões sombreadas em vermelho e azul representam a faixa de oscilação dos empilhamentos  $k$  ajustados pelo modelo PB, discutidos na seção 5.5.2. O sombreado vermelho são das sequências em  $H_2O$ , enquanto os azuis são das sequências em PEG200. As figuras 43b e figura 43c apresentam pouca variação nos valores de  $k$  ajustados anteriormente pelo modelo PB. Na presença de PEG200,

$k_{\text{AU-AU}}$ ,  $k_{\text{AU-UA}}$  e  $k_{\text{UA-AU}}$  tornaram-se um pouco mais rígidos no modelo  $\text{HMS}^-$ . Em contrapartida, as sequências de DNA são as que apresentam maior diferença de ajuste entre os modelos, como visto na figura 43a.  $k_{\text{AT-CG}}$ ,  $k_{\text{AT-GC}}$ ,  $k_{\text{CG-AT}}$ ,  $k_{\text{CG-CG}}$ ,  $k_{\text{CG-GC}}$ ,  $k_{\text{GC-CG}}$  se comportam de maneira diferente do que foi encontrado pelo modelo PB. Os ajustes dos  $k$  em PEG200 são bastante semelhantes, enquanto os ajustes em  $\text{H}_2\text{O}$  são muito diferentes. Após adição de PEG200, o dímero CG-GC se torna um pouco mais rígido no ajuste do modelo PB, mas no modelo  $\text{HMS}^-$  ele se torna flexível. Alguns empilhamentos, por exemplo  $k_{\text{AT-CG}}$ ,  $k_{\text{CG-AT}}$  e  $k_{\text{GC-CG}}$ , são invariantes com a adição de PEG200, conforme ilustrado na figura 43a.

Ao incluirmos no Hamiltoniano um termo que descreve a interação das fitas com o solvente, mais precisamente como as fitas simples interagem com as demais moléculas presentes na solução, vemos que as interações  $\pi - \pi$  entre as bases no DNA sofrem modificações significativas. O efeito direto disso é visto nos perfis de abertura das sequências. Apesar que os empilhamentos  $k$  sejam praticamente os mesmos nas sequências de RNA, seria interessante observar se a adição do termo de barreira modifica a abertura dos duplexos entre os dois modelos. Para isso, buscamos nas bases de dados DNA-P200/40, RNA-P200/20 e RNA-P200/40 por sequências que fossem idênticas, com uma simples diferença: na posição onde encontramos um T nas sequências de DNA, teríamos um U nas sequências de RNA. Identificamos 13 sequências idênticas nas bases de dados DNA-P200/40 e RNA-P200/40, e 1 sequência igual em todas as bases de dados. A sequência idêntica em todas as bases de dados, possui apenas uma diferença: no DNA, a sequência está na fita complementar, enquanto no RNA ela está na fita principal. Logo, os dímeros são diferentes. Portanto, analisaremos o perfil de abertura apenas das bases de dados DNA-P200/40 e RNA-P200/40.

As figuras 44 apresentam os perfis de abertura de 13 sequências idênticas, com a substituição de T por U para RNA, das bases de dados DNA-P200/40 e RNA-P200/40. Para observar o que acontece no extremo dos perfis de abertura dos modelos PB e  $\text{HMS}^-$ , calculamos mais dois perfis de abertura: uma sequência contendo 8 pares de base CG, ilustrado na figura 44n, e uma sequência contendo 8 pares de base AT, ou AU para RNA, mostrado na figura 44o.

RNA possui temperaturas de desnaturação maiores, em alguns casos próximas, do que DNA, conforme mostrado nas tabelas 23 e 25, em  $\text{H}_2\text{O}$  e PEG200/40%wt. Portanto, a fita de RNA deve ser mais difícil de se abrir. Todos os perfis de abertura extraídos das bases de dados DNA-P200/40 e RNA-P200/40, nas figuras 44a-m, apresentam esse comportamento. Sequências compostas puramente por CG apresentam aberturas praticamente nulas, enquanto sequências AT, ou AU para RNA, exibem perfis de abertura bastante diferentes, mas maiores do que sequências puramente CG (aberturas médias marcadas com o mesmo símbolo e cor mostradas nas figuras 44n e 44o). As sequências de RNA ajustadas pelo modelo  $\text{HMS}^-$  apresentam empilhamentos  $k$  aproximadamente iguais aos ajustados pelo modelo PB, conforme mostrado na figura 43c. Notamos que os perfis de



abertura do modelo PB padrão diferem dos perfis de abertura do modelo  $HMS^-$ . Logo, a adição do termo de solvente modifica um pouco os perfis de abertura em relação ao modelo PB, como pode ser observado nas figuras 44.

Testamos vários modelos com a modificação do termo de barreira e apenas o modelo  $HMS^-$  melhorou as temperaturas de predições em relação ao modelo PB. As modificações inseridas no modelo PCLA conferem significado físico às constantes  $c$  e  $d$ , previamente adimensionais no modelo PCLA original, e reduzem os valores de  $\chi^2$  à metade, como mostrado nas figuras 39 e 49. Esses resultados foram encontrados ao variar o termo  $p$ , entre 0 e 3, do expoente  $y_n$  no numerador da equação (39). Acreditamos que alguma das modificações do modelo PCLA possa prever temperaturas melhores do que o modelo PB padrão. Entretanto, mais testes são necessários para confirmarmos esta hipótese. O teste variando os empilhamentos  $k$  e energia de barreira do modelo TRMF passou nos critérios físico-químicos avaliados, mas as predições de temperatura foram piores. A inserção de uma gaussiana para representar a barreira explicaria o que acontece com o sistema no estado intermediário, ou seja, a interação indireta entre os pares de base via o solvente. No entanto, as temperaturas calculadas estariam distantes dos resultados experimentais em comparação aos resultados obtidos com o modelo PB padrão. Na seção 5.5.2, mostramos que o modelo PB explica perfeitamente o que acontece com o sistema e a diferença entre as temperaturas experimentais e preditas é inferior a 2 °C, dentro da margem de erro experimental. Nosso objetivo nessa parte do projeto foi encontrar um modelo, ou modelos, que se aproximasse(m) mais das temperaturas experimentais e explicasse(m) o estado intermediário. Pelos nossos critérios, isto só foi identificado no modelo  $HMS^-$ . Nossos resultados nos encorajam a utilizar o método encontrado para trabalhar com o modelo  $HMS^-$  em outros sistemas, por exemplo *mismatches* e M-DNA, que foram apresentados no capítulo 4.

### 5.6.6 Conclusões e Perspectivas

Os modelos PB com barreira foram testados pela primeira vez com dados experimentais. Encontramos quatro modificações do modelo PB padrão, às quais denominamos HMS [24, 153], PCLA [25], PCLJ [26] e TRMF [27], propostas para descrever o estado intermediário entre a fita simples e a fita dupla. O modelo HMS foi separado em  $HMS^-$  [153] e  $HMS^+$  [24], pois encontramos essas duas formas na literatura. No entanto, nenhum dos modelos mencionados anteriormente foi validado com dados experimentais. Nesse contexto, surge a seguinte pergunta: Será que algum desses novos modelos funciona para prever temperaturas de desnaturação melhores do que o modelo PB e também explicar a dinâmica do sistema? Não dispomos de referências para os parâmetros iniciais e nem qual o melhor método de otimização a ser utilizado para os modelos. Logo, procuramos pelos parâmetros em referências experimentais e teóricas [28–30, 65, 69, 70, 152, 166]. As energias

de barreira foram obtidas através do método de varredura da função  $\chi^2$  no espaço dos parâmetros dos pares AT, ou AU para RNA, e CG, para cada um dos modelos. O método de otimização foi determinado após avaliarmos dois critérios

- As previsões de temperatura devem ser melhores do que as do modelo PB padrão;
- Os resultados devem ter sentido físico com a dinâmica do sistema.

Com base em nossos resultados, apenas o modelo HMS<sup>-</sup> foi aprovado em todos os critérios de avaliação. Entretanto, o modelo de potencial gaussiano, ou TRMF, é capaz de explicar a física do sistema, mas as temperaturas calculadas são piores do que as do modelo PB padrão. Por sua vez, as previsões do modelo PCLA são muito ruins. No entanto, notamos que os parâmetros  $c$  e  $d$  devem ser adimensionais, diferente do que os autores propuseram no artigo Peyrard et al. [25]. Modificamos o termo de barreira deste modelo para atribuir unidades de distância, ou inverso de distância, aos parâmetros  $c$  e  $d$ . Apenas com estas modificações as previsões de temperatura apresentaram uma melhora significativa. Acreditamos que a modificação no modelo PCLA possa prever temperaturas de desnaturação melhores do que o modelo PB padrão. Contudo, testes posteriores devem ser feitos para confirmar essas observações.

Os valores de energia de barreira média AT, ou AU para RNA, e CG otimizados pelo modelo HMS<sup>-</sup> em H<sub>2</sub>O são bem próximos, pois as grupos funcionais que realizam as ligações de hidrogênio entre os pares de base no duplexo são quase idênticos. Há apenas uma diferença: a Guanina possui o grupo -NH<sub>2</sub> ligado ao Carbono C(4), enquanto a Adenina possui o grupo -H. Em baixas concentrações de sal, Na<sup>+</sup>=100mM, não conseguimos diferenciar claramente as energias de barreira em H<sub>2</sub>O dos pares de base intermediados pelo solvente. No entanto, em altas concentrações de sal, Na<sup>+</sup>=1000mM, a barreira de CG é um pouco maior do que AU, como esperado. O efeito de PEG200 é diferente em DNA e RNA. No DNA, observamos um aumento da barreira de energia, enquanto no RNA observamos uma redução. Essa diferença pode ser explicada pelo fato de o RNA é mais hidratado e acumular mais íons de sódio ao seu redor do que o DNA [4, 5, 34, 64]. A concentração menor de íons e água no DNA quer dizer que temos, além de íons sódio e água, moléculas de PEG200 participando da ligação base-solvente-base. Os perfis de abertura do modelo HMS<sup>-</sup> são diferentes dos perfis de abertura do modelo PB. Essa diferença aparece por causa do termo de barreira, que observamos melhor nas sequências RNA-P200/40 e RNA-W/40, cujo os empilhamentos  $k$  dos modelos são praticamente idênticos. Nossos resultados nos encorajam a utilizar o método encontrado para trabalhar com o modelo HMS<sup>-</sup> em outros sistemas, por exemplo *mismatches* e M-DNA apresentados no capítulo 4.

## 6 Conclusões

Nos propusemos a estudar dois sistemas utilizando modelos mesoscópicos, onde o principal modelo utilizado foi o Peyrard-Bishop (PB). O primeiro projeto foi sobre DNA mediado por metal e o segundo nucleotídeos em solventes que simulam meio molecular. Dividimos o projeto dois em duas partes: na primeira, usamos apenas o modelo PB e na segunda, utilizamos modelos PB com barreira de potencial. No projeto de DNA mediado por metal, identificamos um valor de profundidade médio do poço  $D$  muito alto no par T-Hg<sup>2+</sup>-T e um valor um pouco menor no par C-Ag<sup>+</sup>-C. Isso indica que os metais se ligam no centro do duplexo, e o pequeno valor de Morse C-Ag<sup>+</sup>-C favorece a configuração N3-Ag<sup>+</sup>-N3, proposta por Dairaku et al. [132]. Valores elevados de empilhamento CC-CC e TT-TT na presença dos metais Ag<sup>+</sup> e Hg<sup>2+</sup> poderiam ser explicados por uma interação interplanar. Em outras palavras, observamos interação metal-metal na fita, por exemplo Ag<sup>+</sup>-Ag<sup>+</sup>, no caso CC-CC. No entanto, os baixos valores de empilhamento CC-CC e TT-TT encontrados não são consistentes com este tipo de interação. O empilhamento CC-CC próximo de zero pode ter acontecido devido à baixa frequência desse termo em nossa base de dados.

Na primeira parte do projeto de nucleotídeos em solventes que simulam meio molecular, ao compararmos os valores do potencial de Morse  $D$  em soluções de H<sub>2</sub>O e soluções de PEG200/40%wt, determinamos que CG(RNA) é o mais hidratado, seguido por CG(DNA), AU(RNA) e AT(DNA). O fato de CG ser mais hidratado, torna as diferenças de energia entre AT e CG, ou AU e CG, menores. A atuação do PEG200 é determinante para inferir se ocorre mudanças de conformação por causa da hidratação dos nucleotídeos, pois o PEG200 desloca as moléculas de água próximas ao duplexo [33, 48, 141]. Notamos uma maior diferença de energia nas ligações entre os pares AT e CG no DNA do que nos pares AU e CG no RNA, após adição de PEG200 (soluções com menos água disponível). Diferenças de energia maiores entre os pares AT e CG causam distorções na conformação da dupla hélice. Portanto, na presença de PEG200, o DNA pode estar assumindo um outro tipo de conformação, hélice do tipo B (predominante em soluções com água) para uma hélice de tipo A (predominante em soluções com menos água disponível). Outra observação importante é que os valores de AT são praticamente os idênticos tanto em H<sub>2</sub>O quanto em PEG200. Logo, sequências puramente AT sempre apresentam hélices do tipo B.

Na segunda parte do projeto de nucleotídeos em solventes que simulam meio molecular, exploramos diversos modelos propostos para modelar a barreira de hidratação. Nestes casos, o modelo PB padrão sofreu adaptações no potencial de Morse para descrever um estado intermediário na transição da fita dupla para fita simples, um aspecto não considerado no modelo original. Encon-

tramos quatro modificações do modelo PB padrão que denominamos HMS [24, 153], PCLA [25], PCLJ [26] e TRMF [27], os quais incluem uma barreira de potencial no potencial de Morse do modelo PB padrão. O modelo HMS foi separado em  $\text{HMS}^-$  [153] e  $\text{HMS}^+$  [24], pois encontramos essas duas formas na literatura. Porém, nenhum dos modelos mencionados anteriormente foi validado com dados experimentais. Essa situação nos levou à seguinte pergunta: será que algum desses novos modelos funciona para prever temperaturas de desnaturação melhores do que o modelo PB e também explicar a dinâmica do sistema? Para responder a essa questão, realizamos um conjunto de análises, e apenas o modelo  $\text{HMS}^-$  foi validado em todas as etapas. Os valores de energia de barreira média AT, ou AU para RNA, e CG em  $\text{H}_2\text{O}$  determinados pelo o modelo  $\text{HMS}^-$  são bem próximos. O motivo das energias de barreira dos diferentes pares de base serem próximas é porque existe apenas uma diferença nos grupos funcionais no meio do duplexo. Na Guanina, há um grupo  $-\text{NH}_2$  ligado ao Carbono C(4), enquanto na Adenina, existe um átomo de H. Em baixas concentrações de sal, as energias de barreira são praticamente iguais. No entanto, em altas concentrações de sal, a barreira de CG é um pouco maior do que AU. O PEG200 afeta diferente DNA e RNA. No DNA, observamos um aumento da barreira de energia, enquanto no RNA, notamos uma redução da barreira. Essa diferença pode ser explicada porque o RNA é mais hidratado e acumula mais íons de sódio ao seu redor do que o DNA [4, 5, 34, 64]. A concentração menor de íons e água no DNA quer dizer que temos, além de íons sódio e água, moléculas de PEG200 também participam da ligação base-solvente-base.

Nos três projetos que foram objetos desta tese, submetemos o modelo Peyrard-Bishop a situações completamente novas, e em todos os casos, obtivemos parâmetros que são condizentes com as observações experimentais existentes. Assim, avançamos na validação do modelo como ferramenta para obter informações sobre as interações intra-moleculares de DNA e RNA, inclusive com observações sobre as camadas de hidratação desses oligonucleotídeos.

## Referências

- [1] Arnaldo Zaha, Henrique Bunselmeyer Ferreira, and Luciane MP Passaglia. *Biologia Molecular Básica-5*. Artmed Editora, 2014.
- [2] Dušan Ž Veljković, VB Medaković, Jelena M Andrić, and Snežana D Zarić. C–h/o interactions of nucleic bases with a water molecule: a crystallographic and quantum chemical study. *CrystEngComm*, 16(43):10089–10096, 2014.
- [3] Ahmed M Rozza, Imre Bakó, and Julianna Oláh. Theoretical insights into water network of b-dna duplex with watson-crick and hoogsteen base pairing geometries. *Journal of Molecular Liquids*, 362:119728, 2022.
- [4] Pascal Auffinger and Eric Westhof. Water and ion binding around r (upa) 12and d (tpa) 12oligomers-comparison with rna and dna (cpg) 12 duplexes. *Journal of Molecular Biology*, 305(5):1057–1072, 2001.
- [5] Magdalena Gebala and Daniel Herschlag. Quantitative studies of an rna duplex electrostatics by ion counting. *Biophysical Journal*, 117(6):1116–1124, 2019.
- [6] N Peyret, P A Seneviratne, H T Allawi, and John SantaLucia, Jr. Nearest-neighbour thermodynamics and NMR of DNA sequences with internal A·A, C·C G·G and T·T mismatches. *Biochem.*, 38(12):3468–3477, 1999.
- [7] Kathleen P Howard. Thermodynamics of DNA duplex formation: A biophysical chemistry laboratory experiment. *J. Chem. Educ.*, 77(11):1469, 2000.
- [8] 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.
- [9] Palok Aich, Shaunivan L Labiuk, Les W Tari, Louis JT Delbaere, William J Roesler, Kenneth J Falk, Ronald P Steer, and Jeremy S Lee. M-DNA: a complex between divalent metal ions and DNA which behaves as a molecular wire. *J. Mol. Biol.*, 294(2):477–485, 1999.
- [10] Olivia P Schmidt, Simon Jurt, Silke Johannsen, Ashkan Karimi, Roland KO Sigel, and Nathan W Luedtke. Concerted dynamics of metallo-base pairs in an a/b-form helical transition. *Nature communications*, 10, 2019.

- [11] Hehua Liu, Chen Cai, Phensinee Haruehanroengra, Qingqing Yao, Yiqing Chen, Chun Yang, Qiang Luo, Baixing Wu, Jixi Li, Jinbiao Ma, et al. Flexibility and stabilization of hgii-mediated c: T and t: T base pairs in dna duplex. *Nucleic acids research*, 45(5):2910–2918, 2017.
- [12] Jiro Kondo, Yoshinari Tada, Takenori Dairaku, Yoshikazu Hattori, Hisao Saneyoshi, Akira Ono, and Yoshiyuki Tanaka. A metallo-DNA nanowire with uninterrupted one-dimensional silver array. *Nature chemistry*, 9(10):956–960, 2017.
- [13] Akira Ono, Hiroki Kanazawa, Hikari Ito, Misato Goto, Koudai Nakamura, Hisao Saneyoshi, and Jiro Kondo. A novel dna helical wire containing hgii-mediated t: T and t: G pairs. *Angewandte Chemie*, 131(47):16991–16994, 2019.
- [14] Hiroshi Yamaguchi, Jakub Šebera, Jiro Kondo, Shuji Oda, Tomoyuki Komuro, Takuya Kawamura, Takenori Dairaku, Yoshinori Kondo, Itaru Okamoto, Akira Ono, et al. The structure of metallo-DNA with consecutive thymine–HgII–thymine base pairs explains positive entropy for the metallo base pair formation. *Nucleic acids research*, 42(6):4094–4099, 2014.
- [15] Luciana M. Oliveira, Adam S. Long, Tom Brown, Keith R. Fox, and Gerald Weber. Melting temperature measurement and mesoscopic evaluation of single, double and triple DNA mismatches. *Chem. Sci.*, 11:8273–8287, 2020. doi: 10.1039/d0sc01700k. URL <https://pubs.rsc.org/en/content/articlelanding/2020/SC/D0SC01700K>.
- [16] Takenori Dairaku, Kyoko Furuita, Hajime Sato, Jakub Šebera, Katsuyuki Nakashima, Jiro Kondo, Daichi Yamanaka, Yoshinori Kondo, Itaru Okamoto, Akira Ono, et al. Structure determination of an agi-mediated cytosine–cytosine base pair within dna duplex in solution with  $1\text{h}/15\text{n}/109\text{ag}$  nmr spectroscopy. *Chemistry—A European Journal*, 22(37):13028–13031, 2016.
- [17] Akira Ono, Shiqi Cao, Humika Togashi, Mitsuru Tashiro, Takashi Fujimoto, Tomoya Machinami, Shuji Oda, Yoko Miyake, Itaru Okamoto, and Yoshiyuki Tanaka. Specific interactions between silver (i) ions and cytosine–cytosine pairs in dna duplexes. *Chemical communications*, (39):4825–4827, 2008.
- [18] Hidetaka Torigoe, Yukako Miyakawa, Akira Ono, and Tetsuo Kozasa. Thermodynamic properties of the specific binding between ag<sup>+</sup> ions and c: C mismatched base pairs in duplex dna. *Nucleosides, Nucleotides and Nucleic Acids*, 30(2):149–167, 2011.
- [19] Jiro Kondo, Tom Yamada, Chika Hirose, Itaru Okamoto, Yoshiyuki Tanaka, and Akira Ono. Crystal Structure of Metallo DNA Duplex Containing Consecutive Watson–Crick-like T–HgII–T Base Pairs. *Angewandte Chemie International Edition*, 53(9):2385–2388, 2014.

- [20] Kuklennyik Zsuzsanna and Marzilli Luigi G. Mercury(II) Site-Selective Binding to a DNA hairpin. Relationship of Sequence-Dependent Intra- and Interstrand Cross-Linking to the Hairpin-Duplex Conformational Transition. *Inorganic Chemistry*, 35(19):5654–5662, 1996.
- [21] Yoko Miyake, Humika Togashi, Mitsuru Tashiro, Hiroshi Yamaguchi, Shuji Oda, Megumi Kudo, Yoshiyuki Tanaka, Yoshinori Kondo, Ryuichi Sawa, Takashi Fujimoto, et al. MercuryII-mediated formation of thymine- HgII- thymine base pairs in DNA duplexes. *Journal of the American Chemical Society*, 128(7):2172–2173, 2006.
- [22] Yoshiyuki Tanaka, Shuji Oda, Hiroshi Yamaguchi, Yoshinori Kondo, Chojiro Kojima, and Akira Ono. 15n- 15n j-coupling across hgii: direct observation of hgii-mediated t- t base pairs in a dna duplex. *Journal of the American Chemical Society*, 129(2):244–245, 2007.
- [23] 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, 2014.
- [24] Gerald Weber. Sharp DNA denaturation due to solvent interaction. *Europhys. Lett.*, 73(5): 806–811, 2006. doi: 10.1209/epl/i2005-10466-6.
- [25] M. Peyrard, S. Cuesta-López, and D. Angelov. Experimental and theoretical studies of sequence effects on the fluctuation and melting of short DNA molecules. *J. Phys.: Condens. Matter*, 21:034103, 2009.
- [26] Michel Peyrard, Santiago Cuesta-Lopez, and Guillaume James. Nonlinear analysis of the dynamics of dna breathing. *Journal of biological physics*, 35:73–89, 2009.
- [27] R. Tapia-Rojo, J. J. Mazo, and F. Falo. Thermal and mechanical properties of a DNA model with solvation barrier. *Phys. Rev. E*, 82(3):031916, Sep 2010. doi: 10.1103/PhysRevE.82.031916.
- [28] Gerald Weber, Niall Haslam, Jonathan W. Essex, and Cameron Neylon. Thermal equivalence of DNA duplexes for probe design. *J. Phys.: Condens. Matter*, 21:034106, 2009. doi: 10.1088/0953-8984/21/3/034106.
- [29] 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. URL <http://nar.oxfordjournals.org/content/41/1/e30>.
- [30] 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. URL <https://www.sciencedirect.com/science/article/abs/pii/S0301010418311200>.
- [31] P. Nelson. *Biological Physics: Energy, Information, Life*. W. H. Freeman, 2003. ISBN 9780716743729.
- [32] Lauren A Levine, Matthew Junker, Myranda Stark, and Dustin Greenleaf. A DNA melting exercise for a large laboratory class. *J. Chem. Educ.*, 92:1928–1931, 2015. doi: 10.1021/acs.jchemed.5b00049.
- [33] Shu-ichi Nakano, Hisae Karimata, Tatsuo Ohmichi, Junji Kawakami, and Naoki Sugimoto. The effect of molecular crowding with nucleotide length and cosolute structure on dna duplex stability. *Journal of the American Chemical Society*, 126(44):14330–14331, 2004.
- [34] Eriks Rozners and Janelle Moulder. Hydration of short dna, rna and 2'-ome oligonucleotides determined by osmotic stressing. *Nucleic acids research*, 32(1):248–254, 2004.
- [35] Smritimoy Pramanik, Satoru Nagatoishi, Sarika Saxena, Jhimli Bhattacharyya, and Naoki Sugimoto. Conformational flexibility influences degree of hydration of nucleic acid hybrids. *The Journal of Physical Chemistry B*, 115(47):13862–13872, 2011.
- [36] Gunther L Eichhorn. Metal ions as stabilizers or destabilizers of the deoxyribonucleic acid structure. *Nature*, 194(4827):474–475, 1962.
- [37] Akira Ono, Hidetaka Torigoe, Yoshiyuki Tanaka, and Itaru Okamoto. Binding of metal ions by pyrimidine base pairs in DNA duplexes. *Chemical Society Reviews*, 40(12):5855–5866, 2011.
- [38] Vangelis George Kanellis and Cristobal G dos Remedios. A review of heavy metal cation binding to deoxyribonucleic acids for the creation of chemical sensors. *Biophysical reviews*, 10(5):1401–1414, 2018.
- [39] Gunther L Eichhorn and Yong Ae Shin. Interaction of metal ions with polynucleotides and related compounds. xii. the relative effect of various metal ions on dna helicity. *Journal of the American Chemical Society*, 90(26):7323–7328, 1968.
- [40] Guido H Clever and Mitsuhiko Shionoya. Metal–base pairing in DNA. *Coordination Chemistry Reviews*, 254(19-20):2391–2402, 2010.

- [41] Shuntaro Takahashi and Naoki Sugimoto. Stability prediction of canonical and non-canonical structures of nucleic acids in various molecular environments and cells. *Chemical Society Reviews*, 49(23):8439–8468, 2020.
- [42] Ignacio Tinoco, Olke C Uhlenbeck, and Mark D Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230(5293):362–367, 1971.
- [43] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62(23):2755–2757, 1989.
- [44] Tatsuya Funai, Chizuko Tagawa, Osamu Nakagawa, Shun-ichi Wada, Akira Ono, and Hidehito Urata. Enzymatic formation of consecutive thymine–hg ii–thymine base pairs by dna polymerases. *Chemical Communications*, 2020.
- [45] Tatsuya Funai, Megumi Aotani, Risa Kiri, Junko Nakamura, Yuki Miyazaki, Osamu Nakagawa, Shun-ichi Wada, Hidetaka Torigoe, Akira Ono, and Hidehito Urata. Silver (I)-Ion-Mediated Cytosine-Containing Base Pairs: Metal Ion Specificity for Duplex Stabilization and Susceptibility toward DNA Polymerases. *ChemBioChem*, 2019.
- [46] Hidetaka Torigoe, Akira Ono, and Tetsuo Kozasa. Detection of single nucleotide polymorphisms by the specific interaction between transition metal ions and mismatched base pairs in duplex dna. *Transition Metal Chemistry*, 36(2):131–144, 2011.
- [47] Sumra Shahid, Md Imtaiyaz Hassan, Asimul Islam, and Faizan Ahmad. Size-dependent studies of macromolecular crowding on the thermodynamic stability, structure and functional activity of proteins: in vitro and in silico approaches. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1861(2):178–197, 2017.
- [48] CH Spink and JB Chaires. Selective stabilization of triplex dna by poly (ethylene glycols). *Journal of the American Chemical Society*, 117(51):12887–12888, 1995.
- [49] Hisae Karimata, Shu-ichi Nakano, and Naoki Sugimoto. Effects of polyethylene glycol on dna duplex stability at different nacl concentrations. *Bulletin of the Chemical Society of Japan*, 80(10):1987–1994, 2007.
- [50] G. Blanco and A. Blanco. *Medical Biochemistry*. Elsevier Science, 2017. ISBN 9780128035504.
- [51] Erwin Chargaff. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, 6(6):201–209, 1950.

- [52] Rosalind E Franklin and Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.
- [53] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [54] Hemachander Subramanian and Robert A Gatenby. Evolutionary advantage of anti-parallel strand orientation of duplex dna. *Scientific Reports*, 10(1):9883, 2020.
- [55] Robert Eisberg and Robert Resnick. *Quantum physics of atoms, molecules, solids, nuclei, and particles*. 1985.
- [56] T. W. Graham Solomons. *Organic chemistry*. Tenth edition / Craig B. Fryhle. Hoboken, NJ : Wiley, [2011], 2011.
- [57] LG Wade Jr et al. Chapter 16: Aromatic compounds. *Organic Chemistry*, pages 705–748, 2006.
- [58] Wan-Ru Zhuang, Yi Wang, Peng-Fei Cui, Lei Xing, Jaiwoo Lee, Dongyoon Kim, Hu-Lin Jiang, and Yu-Kyoung Oh. Applications of  $\pi$ - $\pi$  stacking interactions in the design of drug-delivery systems. *Journal of controlled release*, 294:311–326, 2019.
- [59] Christopher A Hunter and Jeremy KM Sanders. The nature of  $\pi$ - $\pi$  interactions. *Journal of the American Chemical Society*, 112(14):5525–5534, 1990.
- [60] Hande Karabıyık, Resul Sevinçek, and Hasan Karabıyık.  $\pi$ -cooperativity effect on the base stacking interactions in dna: is there a novel stabilization factor coupled with base pairing h-bonds? *Physical Chemistry Chemical Physics*, 16(29):15527–15538, 2014.
- [61] Holger Kruse, Pavel Baná, and Sponer Jirı́. Investigations of stacked dna base-pair steps: highly accurate stacking interaction energies, energy decomposition, and many-body stacking effects. *Journal of chemical theory and computation*, 15(1):95–115, 2018.
- [62] Christopher A Hunter. Sequence-dependent DNA structure: the role of base stacking interactions. *J. Mol. Biol.*, 230(3):1025–1054, 1993.
- [63] David Owen Wood. *Biophysical investigation of M-DNA*. PhD thesis, Department of Biochemistry University of Saskatchewan, Saskatoon, 2005.
- [64] Sergio Cruz-León and Nadine Schwierz. Rna captures more cations than dna: Insights from molecular dynamics simulations. *The Journal of Physical Chemistry B*, 126(43):8646–8654, 2022.

- [65] Lada Biedermannová, Jiří Černý, Michal Malý, Michaela Nekardová, and Bohdan Schneider. Knowledge-based prediction of dna hydration using hydrated dinucleotides as building blocks. *Acta Crystallographica Section D: Structural Biology*, 78(8), 2022.
- [66] Christine Jorge, Bryan S Marques, Kathleen G Valentine, and A Joshua Wand. Characterizing protein hydration dynamics using solution nmr spectroscopy. In *Methods in enzymology*, volume 615, pages 77–101. Elsevier, 2019.
- [67] Leonhard Geist, Moriz Mayer, Xiao-Ling Cockcroft, Bernhard Wolkerstorfer, Dirk Kessler, Harald Engelhardt, Darryl B McConnell, and Robert Konrat. Direct nmr probing of hydration shells of protein ligand interfaces and its application to drug design. *Journal of Medicinal Chemistry*, 60(21):8708–8715, 2017.
- [68] Nóra Taricska, Mónika Bokor, Dóra K Menyhárd, Kálmán Tompa, and András Perczel. Hydration shell differentiates folded and disordered states of a trp-cage miniprotein, allowing characterization of structural heterogeneity by wide-line nmr measurements. *Scientific Reports*, 9(1):2947, 2019.
- [69] Kaushik Chakraborty, Sriteja Mantha, and Sanjoy Bandyopadhyay. Molecular dynamics simulation of a single-stranded dna with heterogeneous distribution of nucleobases in aqueous medium. *The Journal of Chemical Physics*, 139(7):08B618\_1, 2013.
- [70] Rongpeng Li and Chi H Mak. A deep dive into dna base pairing interactions under water. *The Journal of Physical Chemistry B*, 124(27):5559–5570, 2020.
- [71] Stephen Neidle and Mark Sanderson. *Principles of nucleic acid structure*. Academic Press, 2021.
- [72] Giulia Rossetti, Pablo D Dans, Irene Gomez-Pinto, Ivan Ivani, Carlos Gonzalez, and Modesto Orozco. The structural impact of DNA mismatches. *Nucleic Acids Res.*, 43(8):4309–4321, 2015.
- [73] Virginie Gervais, Jean AH Cagnet, Marc Le Bret, Lawrence C Sowers, and G Victor Fazakerley. Solution structure of two mismatches A·A and T·T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics. *Eur. J. Biochem.*, 228(2):279–290, 1995.
- [74] Chulhyun Lee, Hae-Kap Cheong, Jee-Hyun Cho, and Chaejoon Cheong. Aa mismatched dnas with a single base difference exhibit a large structural change and a propensity for the parallel-stranded conformation. *Journal of Analytical Science & Technology*, 1:37–48, 2010.

- [75] Sherrie Schreiber-Gosche and Robert A Edwards. Thermodynamics of oligonucleotide duplex melting. *J. Chem. Educ.*, 86(5):644, 2009.
- [76] Poater Jordi, Bickelhaupt F. Matthias Swart Marcel, and Guerra Célia Fonseca. B-DNA structure and stability: the role of hydrogen bonding,  $\pi$ - $\pi$  stacking interactions, twist-angle, and solvation. *Organic & Biomolecular Chemistry*, 12(26):4691–4700, 2014.
- [77] Erik de Oliveira Martins. Desenvolvimento de modelos mesoscópicos assimétricos para oligonucleotídeos e sua aplicação a defeitos tipo single-bulges em rnas. 2018.
- [78] Tianbing Xia, John SantaLucia, Jr., Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoyi 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.
- [79] 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.
- [80] RD Wells, JE Larson, RC Grant, BE Shortle, and CR Cantor. Physicochemical studies on polydeoxyribonucleotides containing defined repeating nucleotide sequences. *Journal of molecular biology*, 54(3):465–497, 1970.
- [81] 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. URL <http://bioinformatics.oxfordjournals.org/content/29/10/1345>.
- [82] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.
- [83] Pâmella Miranda de Moura et al. Desenvolvimento de uma representação mesoscópica para marcadores fluorescentes em dna e predição de intensidade de fluorescência. 2020.
- [84] K. J. Breslauer, R Frank, H Blocker, and L. A. Marky. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83(11):3746–3750, 1986.
- [85] John SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95(4):1460–1465, 1998. URL <http://www.pnas.org/cgi/content/abstract/95/4/1460>.

- [86] Naoki Sugimoto, Shu-ichi Nakano, Misa Katoh, Akiko Matsumura, Hiroyuki Nakamuta, Tatsuo Ohmichi, Mari Yoneyama, and Muneo Sasaki. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochem.*, 34(35):11211–11216, 1995.
- [87] Tetsuo Yamane and Norman Davidson. On the complexing of desoxyribonucleic acid (dna) by mercuric ion<sup>1</sup>. *Journal of the American Chemical Society*, 83(12):2599–2607, 1961.
- [88] Tetsuo Yamane and Norman Davidson. On the complexing of deoxyribonucleic acid by silver (i). *Biochimica et biophysica acta*, 55(5):609–621, 1962.
- [89] N Davidson, J Widholm, US Nandi, R Jensen, BM Olivera, and JC Wang. Preparation and properties of native crab dat. *Proceedings of the National Academy of Sciences of the United States of America*, 53(1):111, 1965.
- [90] Luigi G Marzilli. Metal-ion interactions with nucleic acids and nucleic acid derivatives. *Progress in inorganic chemistry*, 23:255–378, 1977.
- [91] Katz Sidney. The reversible reaction of Hg (II) and double-stranded polynucleotides a step-function theory and its significance. *Biochimica et Biophysica Acta (BBA) - Specialized Section on Nucleic Acids and Related Subjects*, 68:240–253, 1963.
- [92] Buncel E., Joly H. Boone C., and Noris A. R. Kumar R. Metal ion-biomolecule interactions. XII. <sup>1</sup>H and <sup>13</sup>C NMR evidence for the preferred reaction of thymidine over guanosine in exchange and competition reactions with Mercury(II) and Methylmercury(II). *Journal of Inorganic Biochemistry*, 25(1):61–73, 1985.
- [93] Shawn D Wettig, David O Wood, and Jeremy S Lee. Thermodynamic investigation of m-dna: a novel metal ion–dna complex. *Journal of inorganic biochemistry*, 94(1-2):94–99, 2003.
- [94] Jeremy S Lee, Laura JP Latimer, and R Stephen Reid. A cooperative conformational change in duplex dna induced by zn<sup>2+</sup> and other divalent metal ions. *Biochemistry and cell biology*, 71(3-4):162–168, 1993.
- [95] Hidetaka Torigoe, Itaru Okamoto, Takenori Dairaku, Yoshiyuki Tanaka, Akira Ono, and Tetsuo Kozasa. Thermodynamic and structural properties of the specific binding between Ag<sup>+</sup> ion and C: C mismatched base pair in duplex dna to form C–Ag–C metal-mediated base pair. *Biochimie*, 94(11):2431–2440, 2012.
- [96] Hidetaka Torigoe, Yukako Miyakawa, Akira Ono, and Tetsuo Kozasa. Positive cooperativity of the specific binding between Hg<sup>2+</sup> ion and T: T mismatched base pairs in duplex DNA. *Thermochimica acta*, 532:28–35, 2012.

- [97] Osamu Nakagawa, Akane Fujii, Yuki Kishimoto, Yusuke Nakatsuji, Natsumi Nozaki, and Satoshi Obika. 2'-o, 4'-c-methylene-bridged nucleic acids stabilize metal-mediated base pairing in a dna duplex. *ChemBioChem*, 19(22):2372–2379, 2018.
- [98] Tim Richters, Olga Krug, Jutta Kösters, Alexander Hepp, and Jens Müller. A family of “click” nucleosides for metal-mediated base pairing: Unravelling the principles of highly stabilizing metal-mediated base pairs. *Chemistry—A European Journal*, 20(25):7811–7818, 2014.
- [99] Soham Mandal, Marian Hebenbrock, and Jens Müller. A dinuclear mercury (ii)-mediated base pair in dna. *Angewandte Chemie International Edition*, 55(50):15520–15523, 2016.
- [100] Kristof Seubert, Célia Fonseca Guerra, F Matthias Bickelhaupt, and Jens Müller. Chimeric gna/dna metal-mediated base pairs. *Chemical Communications*, 47(39):11041–11043, 2011.
- [101] Mohtashim Hassan Shamsi and Heinz-Bernhard Kraatz. Interactions of metal ions with dna and some applications. *Journal of Inorganic and Organometallic Polymers and Materials*, 23(1):4–23, 2013.
- [102] Tatsuya Funai, Nahomi Adachi, Megumi Aotani, Shun-ichi Wada, and Hidehito Urata. Effects of metal ions on thermal stabilities of DNA duplexes containing homo- and heterochiral mismatched base pairs: comparison of internal and terminal substitutions. *Nucleosides, Nucleotides and Nucleic Acids*, pages 1–12, 2019.
- [103] Hidehito Urata, Eriko Yamaguchi, Yasunari Nakamura, and Shun-ichi Wada. Pyrimidine–pyrimidine base pairs stabilized by silver (I) ions. *Chemical Communications*, 47(3):941–943, 2011.
- [104] Nicole Zimmermann, Eric Meggers, and Peter G Schultz. A novel silver (I)-mediated DNA base pair. *Journal of the American Chemical Society*, 124(46):13684–13685, 2002.
- [105] Isabell Schönrrath, Vladimir B Tsvetkov, Timofei S Zatsëpin, Andrey V Aralov, and Jens Müller. Silver (i)-mediated base pairing in parallel-stranded dna involving the luminescent cytosine analog 1, 3-diaza-2-oxophenoxazine. *JBIC Journal of Biological Inorganic Chemistry*, 24(5):693–702, 2019.
- [106] Hidetaka Torigoe, Yukako Miyakawa, Akira Ono, and Tetsuo Kozasa. Thermodynamic properties of the specific binding between Ag<sup>+</sup> ions and C: C mismatched base pairs in duplex DNA. *Nucleosides, Nucleotides and Nucleic Acids*, 30(2):149–167, 2011.
- [107] Shuvankar Naskar, Marian Hebenbrock, and Jens Müller. Light-induced formation of silver (i)-mediated base pairs in dna: Possibilities and limitations. *Inorganica Chimica Acta*, 512: 119856, 2020.

- [108] Ji Hoon Han, Shingo Hirashima, Soyoung Park, and Hiroshi Sugiyama. Highly sensitive and selective mercury sensor based on mismatched base pairing with dioxo. *Chemical Communications*, 55(69):10245–10248, 2019.
- [109] Guillaume Mata, Olivia P Schmidt, and Nathan W Luedtke. A fluorescent surrogate of thymidine in duplex dna. *Chemical Communications*, 52(25):4718–4721, 2016.
- [110] Olivia P Schmidt, Andrea S Benz, Guillaume Mata, and Nathan W Luedtke. HgII binds to c–t mismatches with high affinity. *Nucleic acids research*, 46(13):6470–6479, 2018.
- [111] Hidetaka Torigoe, Akira Ono, and Tetsuo Kozasa. HgII ion specifically binds with T: T mismatched base pair in duplex DNA. *Chemistry–A European Journal*, 16(44):13218–13225, 2010.
- [112] Xiurong Guo, Sachin A Ingale, Haozhe Yang, Yang He, and Frank Seela. Mercury ii-mediated base pairs in dna: unexpected behavior in metal ion binding and duplex stability induced by 2-deoxyuridine 5-substituents. *Organic & biomolecular chemistry*, 15(4):870–883, 2017.
- [113] Shuvankar Naskar and Jens Müller. Light-induced formation of thymine-containing mercury (II)-mediated base pairs. *Chemistry–A European Journal*, 25(71):16214–16218, 2019.
- [114] Yong Wang, Brandon Ritzo, and Li-Qun Gu. Silver (i) ions modulate the stability of dna duplexes containing cytosine, methylcytosine and hydroxymethylcytosine at different salt concentrations. *RSC advances*, 5(4):2655–2658, 2015.
- [115] Angel Terrón, Blas Moreno-Vachiano, Antonio Bauzá, Angel García-Raso, Juan Jesús Fiol, Miquel Barceló-Oliver, Elies Molins, and Antonio Frontera. X-ray crystal structure of a metallated double-helix generated by infinite and consecutive c\*-agi-c\*(c\*: N1-hexylcytosine) base pairs through argentophilic and hydrogen bond interactions. *Chemistry–A European Journal*, 23(9):2103–2108, 2017.
- [116] Jakub Šebera, Jaroslav Burda, Michal Straka, Akira Ono, Chojiro Kojima, Yoshiyuki Tanaka, and Vladimír Sychrovský. Formation of a thymine-hgii-thymine metal-mediated dna base pair: proposal and theoretical calculation of the reaction pathway. *Chemistry–A European Journal*, 19(30):9884–9894, 2013.
- [117] Dominik A Megger, Célia Fonseca Guerra, F Matthias Bickelhaupt, and Jens Müller. Silver (i)-mediated hoogsteen-type base pairs. *Journal of inorganic biochemistry*, 105(11):1398–1404, 2011.

- [118] Tiziana Marino, Nino Russo, Marirosa Toscano, and Matej Pavelka. Theoretical investigation on dna/rna base pairs mediated by copper, silver, and gold cations. *Dalton Transactions*, 41(6):1816–1823, 2012.
- [119] R Longuinhos, AD Lúcio, H Chacham, and SS Alexandre. Charge-transfer optical absorption mechanism of dna: Ag-nanocluster complexes. *Physical Review E*, 93(5):052413, 2016.
- [120] Leonardo Andres Espinosa Leal and Olga Lopez-Acevedo. On the interaction between gold and silver metal atoms and dna/rna nucleobases—a comprehensive computational study of ground state properties. *Nanotechnology Reviews*, 4(2):173–191, 2015.
- [121] Leonardo Andres Espinosa Leal, Alexander Karpenko, Steven Swasey, Elisabeth G Gwinn, Victor Rojas-Cervellera, Carme Rovira, and Olga Lopez-Acevedo. The role of hydrogen bonds in the stabilization of silver-mediated cytosine tetramers.
- [122] Guo-Jin Cao. Dinuclear metal-mediated homo base pairs with metallophilic interactions: Theoretical studies of  $g_2m_2^{2+}$  ( $m = \text{cu, ag, and au}$ ) ions. *Scientific reports*, 7(1):1–9, 2017.
- [123] Simone S Alexandre, Bernardo J Murta, José M Soler, and Félix Zamora. Stability and electronic structure of M-DNA: Role of metal position. *Phys. Rev. B*, 84(4):045413, 2011.
- [124] Simone S Alexandre, José M Soler, Luis Seijo, and Félix Zamora. Geometry and electronic structure of m-dna ( $m = \text{zn}^{2+}, \text{co}^{2+}, \text{and fe}^{2+}$ ). *Physical Review B*, 73(20):205112, 2006.
- [125] Irena Kratochvílová, Martin Golan, Martin Vala, Miroslava Spérová, Martin Weiter, Ondrej Páv, Jakub Sebera, Ivan Rosenberg, Vladimir Sychrovsky, Yoshiyuki Tanaka, et al. Theoretical and experimental study of charge transfer through DNA: impact of mercury mediated T-Hg-T base pair. *The Journal of Physical Chemistry B*, 118(20):5374–5381, 2014.
- [126] Xi Chen, Alexander Karpenko, and Olga Lopez-Acevedo. Silver-mediated double helix: structural parameters for a robust DNA building block. *ACS omega*, 2(10):7343–7348, 2017.
- [127] Pavel N Kliuev, Petr A Sokolov, and Ruslan R Ramazanov. Qm/mm-md dissociation of  $\text{ag}^+$  and  $\text{h}^+$  mediated cytosine pairs: Monomers and dimers. *Journal of Organometallic Chemistry*, 919:121333, 2020.
- [128] Akira Ono and Yoko Miyake. Highly selective binding of metal ions to thymine-thymine and cytosine-cytosine base pairs in dna duplexes. *Nucleic Acids Res. Suppl*, 3:227–228, 2003.
- [129] Akane Fujii, Osamu Nakagawa, Yuki Kishimoto, Takumi Okuda, Yusuke Nakatsuji, Natsumi Nozaki, Yuuya Kasahara, and Satoshi Obika. 1, 3, 9-triaza-2-oxophenoxazine: An artificial

- nucleobase forming highly stable self-base pairs with three agi ions in a duplex. *Chemistry–A European Journal*, 25(31):7443–7448, 2019.
- [130] Richard Owczarzy, Yong You, Bernardo G. Moreira, Jeffrey A. Manthey, Lingyan Huang, Mark A. Behlke, and Joseph A. Walder. Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochem.*, 43:3537–3554, 2004.
- [131] P Maniadis, G Kalosakas, KØ Rasmussen, and AR Bishop. Polaron normal modes in the peyrard-bishop-holstein model. *Physical Review B*, 68(17):174304, 2003.
- [132] Takenori Dairaku, Kyoko Furuita, Hajime Sato, Jakub Šebera, Katsuyuki Nakashima, Akira Ono, Vladimír Sychrovský, Chojiro Kojima, and Yoshiyuki Tanaka. Hgii/agi-mediated base pairs and their nmr spectroscopic studies. *Inorganica Chimica Acta*, 452:34–42, 2016.
- [133] Biswarup Jash, Philipp Scharf, Nikolas Sandmann, Célia Fonseca Guerra, Dominik A Megger, and Jens Müller. A metal-mediated base pair that discriminates between the canonical pyrimidine nucleobases. *Chemical Science*, 8(2):1337–1343, 2017.
- [134] J Christian Léon, Indranil Sinha, and Jens Müller. 6-pyrazolylpurine as an artificial nucleobase for metal-mediated base pairing in dna duplexes. *International Journal of Molecular Sciences*, 17(4):554, 2016.
- [135] Dominik A Megger, Célia Fonseca Guerra, Jan Hoffmann, Bernhard Brutschy, F Matthias Bickelhaupt, and Jens Müller. Contiguous metal-mediated base pairs comprising two agi ions. *Chemistry–A European Journal*, 17(23):6533–6544, 2011.
- [136] Indranil Sinha, Célia Fonseca Guerra, and Jens Müller. A highly stabilizing silver (i)-mediated base pair in parallel-stranded dna. *Angewandte Chemie International Edition*, 54(12):3603–3606, 2015.
- [137] Steven B Zimmerman and Allen P Minton. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annual review of biophysics and biomolecular structure*, 22(1):27–65, 1993.
- [138] Daisuke Miyoshi and Naoki Sugimoto. Molecular crowding effects on structure and stability of dna. *Biochimie*, 90(7):1040–1051, 2008.
- [139] Hisae Tateishi-Karimata, Shu-ichi Nakano, and Naoki Sugimoto. Quantitative analyses of nucleic acid stability under the molecular crowding condition induced by cosolutes. *Current Protocols in Nucleic Acid Chemistry*, 53(1):7–19, 2013.

- [140] Shu-ichi Nakano, Lei Wu, Hirohito Oka, Hisae Tateishi Karimata, Toshimasa Kirihata, Yuichi Sato, Satoshi Fujii, Hiroshi Sakai, Masayuki Kuwahara, Hiroaki Sawai, et al. Conformation and the sodium ion condensation on dna and rna structures in the presence of a neutral cosolute as a mimic of the intracellular media. *Molecular BioSystems*, 4(6):579–588, 2008.
- [141] Charles H Spink and Jonathan B Chaires. Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex dna. *Biochemistry*, 38(1):496–508, 1999.
- [142] Shu-ichi Nakano, Daisuke Yamaguchi, Hisae Tateishi-Karimata, Daisuke Miyoshi, and Naoki Sugimoto. Hydration changes upon dna folding studied by osmotic stress experiments. *Biophysical journal*, 102(12):2808–2817, 2012.
- [143] Saptarshi Ghosh, Shuntaro Takahashi, Tamaki Endoh, Hisae Tateishi-Karimata, Soumitra Hazra, and Naoki Sugimoto. Validation of the nearest-neighbor model for watson–crick self-complementary dna duplexes in molecular crowding condition. *Nucleic acids research*, 47(7):3284–3294, 2019.
- [144] Saptarshi Ghosh, Shuntaro Takahashi, Tatsuya Ohyama, Tamaki Endoh, Hisae Tateishi-Karimata, and Naoki Sugimoto. Nearest-neighbor parameters for predicting dna duplex stability in diverse molecular crowding conditions. *Proceedings of the National Academy of Sciences*, 2020.
- [145] Saptarshi Ghosh, Shuntaro Takahashi, Dipanwita Banerjee, Tatsuya Ohyama, Tamaki Endoh, Hisae Tateishi-Karimata, and Naoki Sugimoto. Nearest-neighbor parameters for the prediction of rna duplex stability in diverse in vitro and cellular-like crowding conditions. *Nucleic Acids Research*, page gkad020, 2023.
- [146] Miranda S Adams and Brent M Znosko. Thermodynamic characterization and nearest neighbor parameters for rna duplexes under molecular crowding conditions. *Nucleic acids research*, 47(7):3658–3666, 2019.
- [147] Xiaobo Gu, Mai-Thao Nguyen, Abigail Overacre, Samantha Seaton, and Susan J Schroeder. Effects of salt, polyethylene glycol, and locked nucleic acids on the thermodynamic stabilities of consecutive terminal adenosine mismatches in rna duplexes. *The Journal of Physical Chemistry B*, 117(13):3531–3540, 2013.
- [148] Charles H Spink, Nichola Garbett, and Jonathan B Chaires. Enthalpies of dna melting in the presence of osmolytes. *Biophysical chemistry*, 126(1-3):176–185, 2007.

- [149] Shu-ichi Nakano, Yuichi Kitagawa, Daisuke Miyoshi, and Naoki Sugimoto. Hammerhead ribozyme activity and oligonucleotide duplex stability in mixed solutions of water and organic compounds. *FEBS Open Bio*, 4:643–650, 2014.
- [150] A Hartwig and MAK Commission. N-isopropyl-n-phenyl-p-phenylenediamine [mak value documentation, 2013]. *The MAK-Collection for Occupational Health and Safety: Annual Thresholds and Classifications for the Workplace*, 1(2):780–799, 2002.
- [151] Emmanuel Giudice, Péter Várnai, and Richard Lavery. Base pair opening within b-dna: free energy pathways for gc and at pairs from umbrella sampling simulations. *Nucleic acids research*, 31(5):1434–1443, 2003.
- [152] Zhonglong Luo, Hongyan Xiao, Xiaochuan Peng, Yulong Li, Zhongpeng Zhu, Ye Tian, and Lei Jiang. Long-range ordered water correlations between a-t/c-g nucleotides. *Matter*, 3(3):794–804, 2020.
- [153] K. Drukker, G. Wu, and G. C. Schatz. Model simulations of DNA denaturation dynamics. *J. Chem. Phys.*, 114(1):579–590, 2001.
- [154] Marco Zoli. Thermodynamics of twisted dna with solvent interaction. *The Journal of chemical physics*, 135(11):09B606, 2011.
- [155] Marco Zoli. Dna size in confined environments. *Physical Chemistry Chemical Physics*, 21(23):12566–12575, 2019.
- [156] Marta Marty-Roda, Oda Dahlen, Titus S van Erp, and Santiago Cuesta-López. Improving the mesoscopic modeling of DNA denaturation dynamics. *Physical biology*, 15(6):066001, 2018.
- [157] Hernán Cortez Gutiérrez, Elso Drigo Filho, José Roberto Ruggiero, and Milton Cortez Gutierrez. Mobile breathers in a nonlinear model for dna breathing. *Eclética Química*, 42:71–75, 2017.
- [158] N. F. Ribeiro and E. Drigo, Filho. Thermodynamics of a Peyrard–Bishop one-dimensional lattice with on-site “hump” potential. *Brazilian J. Phys.*, 41:195–200, 2011.
- [159] Rivka Goobes, Nava Kahana, Orit Cohen, and Abraham Minsky. Metabolic buffering exerted by macromolecular crowding on dna- dna interactions: Origin and physiological significance. *Biochemistry*, 42(8):2431–2440, 2003.
- [160] Kazuya Koumoto, Hirofumi Ochiai, and Naoki Sugimoto. Hydration is an important factor to regulate thermodynamic stability of a dna duplex under molecular crowding conditions. *Chemistry letters*, 37(8):864–865, 2008.

- [161] Xiao-Bo Gu, Shu-ichi Nakano, and Naoki Sugimoto. Consecutive gc base pairs determine the energy barrier of dna duplex formation under molecularly crowded conditions. *Chemical communications*, (26):2750–2752, 2007.
- [162] Ye Teng, Smritimoy Pramanik, Hisae Tateishi-Karimata, Tatsuya Ohyama, and Naoki Sugimoto. Drastic stability change of XX mismatch in d(CXG) trinucleotide repeat disorders under molecular crowding condition. *Biochem. Bioph. Res. Co.*, 496(2):601–607, 2018.
- [163] Gerald Weber. Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures. *Bioinformatics*, 31(6):871–877, 2015. doi: 10.1093/bioinformatics/btu751. URL <http://bioinformatics.oxfordjournals.org/content/31/6/871>.
- [164] Erik de Oliveira Martins, Vivianne Basílio Barbosa, and Gerald Weber. DNA/RNA hybrid mesoscopic model shows strong stability dependence with deoxypyrimidine content and stacking interactions similar to RNA/RNA. *Chem. Phys. Lett.*, 715C:14–19, 2019. doi: 10.1016/j.cplett.2018.11.015.
- [165] Maria Izabel Muniz, Adrian H Bustos, Sofie Slott, Kira Astakhova, and Gerald Weber. Cation valence dependence of hydrogen bond and stacking potentials in dna mesoscopic models. *Biophysical Chemistry*, 294:106949, 2023.
- [166] Takamasa Arakawa, Narutoshi Kamiya, Haruki Nakamura, and Ikuo Fukuda. Molecular dynamics simulations of double-stranded dna in an explicit solvent model with the zero-dipole summation method. *PLoS One*, 8(10):e76606, 2013.
- [167] Shu-ichi Nakano and Naoki Sugimoto. The structural stability and catalytic activity of dna and rna oligonucleotides in the presence of organic solvents. *Biophysical reviews*, 8(1):11–23, 2016.
- [168] Robert J Fick, Amy Y Liu, Felix Nussbaumer, Christoph Kreutz, Atul Rangadurai, Yu Xu, Roger D Sommer, Honglue Shi, Steve Scheiner, and Allison L Stelling. Probing the hydrogen-bonding environment of individual bases in dna duplexes with isotope-edited infrared spectroscopy. *The Journal of Physical Chemistry B*, 125(28):7613–7627, 2021.

## **APÊNDICE A - Sequências de DNA e RNA utilizadas nos projetos**

Neste apêndice, listamos todas as sequências utilizadas nos capítulos 4 e 5. Todas as tabelas contêm as concentrações totais das fitas  $C_t$ , as concentrações de sal  $[\text{Na}^+]$ , as temperaturas experimentais ( $T_{exp}$ ), as temperaturas ajustadas ( $T_{adj}$ ) ou as temperaturas calculadas ( $T_{calc}$ ) pelo modelo NN da seção 3.2, e as temperaturas calculadas ( $T'$ ) pelos modelos mesoscópicos.

Sequence	Ref.	$C_t$	[Na <sup>+</sup> ]	CC-CTRL			C-Ag <sup>+</sup> -C		
				$T_{exp}$	$T_{adj}$	$T'$	$T_{exp}$	$T_{adj}$	$T'$
GTGACCAC <u>C</u> AGCAGTG	[37]	2	100	43	43	43.42	51	51	52.90
GTGACCA <u>C</u> TGCAGTG	[37]	2	100	43	43	42.88	52	52	49.32
GTGACCT <u>C</u> AGCAGTG	[37]	2	100	45	45	45.19	52	52	56.09
GTGACCC <u>G</u> CGCAGTG	[37]	2	100	52	52	52.31	56	56	56.96
GTGACCC <u>G</u> CCGAGTG	[37]	2	100	54	54	53.46	61	61	60.54
GTGACCC <u>C</u> GCGAGTG	[37]	2	100	51	51	51.18	53	53	53.27
AAAT <u>C</u> TGCG	[45]	6	100	7.3	4.81	5.9	18.4	16.18	12.78
CATTGG <u>A</u> CTGACGCT	[102]	2	100	42.5	42.5	43.53	52.4	52.4	49.50
GACGT <u>C</u> CTACG	[103]	5	1000	31.7	29.26	26.65	36.5	33.55	34.63
AGCAAACAACGC	[97]	2	100	30	30	29.98	42	42	38.08
CACTGCA <u>C</u> TGGTCAC	[17]	1	100	40	44.35	42.88	47	52.31	49.32
CACATT <u>A</u> CTGTTGTA	[104]	1	55	-	-	-	39.1	26.89	28.70
AAAAAAAAAA <u>A</u> CTTTTAAATATTT	[105]	3	150	28.9	33.23	32.65	31.1	38.11	38.04
CTCAGATCCTG <u>C</u> CCTTCAAAAACAA	[106]	1	100	57.9	57.89	56.93	62.7	62.61	61.53
CTCAGATCCTGT <u>C</u> CCTTCAAAAACAA	[106]	1	100	55.8	54.31	54.17	60.7	59.87	60.96
CTCAGATCCTG <u>C</u> CTTCAAAAACAA	[106]	1	100	55.4	57.64	55.03	60.2	61.3	59.51
CTCAGATCCTGT <u>C</u> TTTCAAAAACAA	[106]	1	100	53.1	53.64	52.19	58.3	58.91	58.95
GAGGGACAGAAAG	[107]	1	150	26.7	30.15	32.9	38.2	39.62	42.10
CTCAGATCCTG <u>C</u> CCTTCAAAAACAA	[106]	1	100	49	49.63	52.6	57	57.67	58.31
CTCAGATCCT <u>C</u> CCCTTCAAAAACAA	[106]	1	100	46	49.3	52.08	56	58.06	59.24
AAAAAAAAAA <u>A</u> CAAAAAAAAAAA	[128]	1	100	31	35.25	31.76	39	41.06	38.86
GCGTT <u>C</u> TTTGCT	[129]	2	100	30	30	29.98	42	42	38.08

**Tabela 21**

Sequências de DNA utilizadas no projeto DNA mediado por metal com CC *mismatches* na presença de Ag<sup>+</sup> e suas respectivas sequências controle (CC-CTRL), quando possível, contendo até 3 *mismatches*. Fitos principais estão no sentido 5' para 3' e as posições do *mismatch* CC estão sublinhadas. Algumas temperaturas foram ajustas como descrito na seção 4.  $T_{exp}$  são temperaturas de desnaturação medidas,  $T_{adj}$  as ajustadas e  $T'$  as calculadas. Concentrações de são [Na<sup>+</sup>] estão em mM. Concentrações totais das fitas  $C_t$  exibidas (em  $\mu$ M) são as concentrações originalmente medidas. Todas as temperaturas estão em °C.

Sequence	Ref.	$C_t$	[Na <sup>+</sup> ]	TT-CTRL			T-Hg <sup>2+</sup> -T		
				$T_{exp}$	$T_{adj}$	$T'$	$T_{exp}$	$T_{adj}$	$T'$
GTGACCA <u>T</u> AGCAGTG	[37]	2	100	47	47	47.69	56	56	53.91
GTGACCA <u>T</u> TGCAGTG	[37]	2	100	49	49	47.28	57	57	55.76
GTGACCT <u>T</u> AGCAGTG	[37]	2	100	48	48	48.25	53	53	51.94
GTGACCG <u>T</u> GGCAGTG	[37]	2	100	55	55	55.23	60	60	60.38
GTGACCG <u>T</u> CGCAGTG	[37]	2	100	55	55	55.9	64	64	63.36
GTGACCC <u>T</u> GGCAGTG	[37]	2	100	56	56	54.58	59	59	57.06
AAA <u>T</u> CTGCG	[102]	6	100	19.1	14.51	11.38	46.2	43.13	42.64
GCCCTGCCTG <u>T</u> CTCCCAGATCACTG	[111]	1	100	65.7	70.89	71.47	69.8	71.67	70.66
CTCAGATCCTG <u>T</u> CTTCAAAAACAA	[111]	1	100	59.2	61.16	61.12	64.2	63.69	64.44
CCCTAACCC <u>T</u> AACTCCTAACCC	[10]	5	500	61.5	57.65	58.93	67.5	62.97	63.64
CCCTAACCC <u>T</u> AACTCTAACCC	[10]	5	500	61.5	58.91	58.18	65.1	62.45	62.38
CCCTAACCC <u>T</u> AACTTAACCC	[10]	5	500	62.8	58.71	58.84	65.7	61.93	61.01
AAAAAAAAAA <u>T</u> AAAAAAAAAA	[21]	1	100	34	37	37.05	41	47	47.76
GACG <u>T</u> TCTACG	[103]	5	1000	27.6	24.97	26.88	45.7	46.24	45.43
GTGACCA <u>T</u> TTGCAGTG	[21]	2	100	45	45	45.92	58	58	56.59
CGCG <u>T</u> TGTCC	[14]	5	100	25	21.92	25.55	54	48.25	49.67
CCGCG <u>T</u> TCTCCG	[14]	5	100	-	-	-	61	61.32	63.20
CCGCG <u>T</u> TGTCCG	[14]	5	100	-	-	-	61	59.18	57.56
CCGCG <u>T</u> TTTCCG	[14]	5	100	-	-	-	59	57.71	57.36
CCGCG <u>T</u> TAACCG	[14]	5	100	-	-	-	55	55.79	52.58
CCGC <u>T</u> TTTCCG	[14]	5	100	-	-	-	55	52.33	53.97
CCGC <u>T</u> TGTCCG	[14]	5	100	-	-	-	54	53.86	53.99
CCGC <u>T</u> TTTCCG	[14]	5	100	-	-	-	53	51.59	53.64
CCGC <u>T</u> TATCCG	[14]	5	100	-	-	-	52	48.92	51.99
CCGC <u>T</u> TATCCG	[14]	5	100	-	-	-	50	48.37	51.26
CCGC <u>T</u> TTATCCG	[14]	5	100	-	-	-	46	44.29	48.15
CATTGG <u>A</u> TGACGCT	[44]	2	100	46.5	46.5	47.58	55.4	55.4	55.80
CATTGG <u>A</u> TTGACGCT	[44]	2	100	41.8	41.8	46.57	54.3	54.3	56.62
CATTGG <u>A</u> TTTTGACGCT	[44]	2	100	39.2	39.2	45.8	55.2	55.2	57.39
C <u>T</u> CTAACCCCTAACCCCTAACCC	[109]	5	500	67	58.91	58.18	68.5	62.45	62.38
CG <u>T</u> TTTCATGATACG	[110]	10	200	37.9	32.31	36.2	57.3	51.69	50.37
CGTCCG <u>T</u> ATTACGCACGC	[108]	2	100	58.4	58.4	58.99	66.2	66.2	63.82
GCG <u>T</u> TTTTTTGCT	[129]	2	100	35	35	32.77	46	46	46.66
AGCAA <u>A</u> TAACGC	[97]	2	100	35	35	32.77	47	47	46.66

**Tabela 22**

Sequências de DNA contendo *mismatches* TT utilizados no trabalho DNA mediado por metal na presença de Hg<sup>2+</sup> e suas respectivas sequências controle (TT-CTRL), quando disponíveis. Fitas principais estão no sentido 5' para 3' *mismatches* estão sublinhadas. Algumas temperaturas foram ajustas como descrito na seção 4.  $T_{exp}$  são as temperaturas de desnaturação medidas,  $T_{adj}$  as ajustadas e  $T'$  as calculadas. Concentrações de sal [Na<sup>+</sup>] estão em mM e concentrações totais das fitas  $C_t$  em  $\mu$ M. Todas as temperaturas estão em °C.

Sequence	Ref.	$C_t$	[Na <sup>+</sup> ]	DNA-P200/40			DNA-W/40		
				$T_{exp}$	$T'_{PB}$	$T'_{HMS-}$	$T_{calc}$	$T'_{PB}$	$T'_{HMS-}$
CCGCGG	[143, 144]	100	100	36.0	36.13	35.8	52.31	49.15	50.55
CGGCCG	[143, 144]	100	100	36.9	36.13	35.8	52.31	49.15	50.55
GGACGTCC	[143, 144]	100	100	35.4	35.94	35.67	48.19	49.88	49.87
GACCGGTC	[143, 144]	100	100	35.6	35.89	35.61	48.19	49.86	49.86
CGTCGACG	[143, 144]	100	100	39.0	39.03	38.77	54.37	51.88	53.29
CGACGTGC	[143, 144]	100	100	39.3	39.03	38.77	54.37	51.88	53.29
CAAGCTTG	[143, 144]	100	100	28.9	29.27	29.1	45.11	42.33	43.09
CTTGCAAG	[143, 144]	100	100	29.6	29.27	29.1	45.11	42.34	43.1
CGGTACCG	[143, 144]	100	100	32.3	34.91	34.26	53.16	52.31	52.09
CCGTACGG	[143, 144]	100	100	34.7	34.91	34.26	53.16	52.31	52.09
GATCCGGATC	[143, 144]	100	100	37.3	38.59	38.47	50.99	51.96	52
GGATCGATCC	[143, 144]	100	100	38.2	38.62	38.48	50.99	51.97	52
ATGAGCTCAT	[143, 144]	100	100	34.3	33.30	33.18	50.41	48.70	48.77
ATCAGCTGAT	[143, 144]	100	100	34.0	33.29	33.18	50.41	48.70	48.78
TGCCGCGGCA	[143, 144]	100	100	53.9	54.70	54.2	70.33	69.33	69.85
TGGCGCGCCA	[143, 144]	100	100	54.8	51.09	54.24	70.33	69.34	69.88
CATAGGCCTATG	[143, 144]	100	100	39.8	40.22	40.4	56.70	57.62	57.45
CTATGGCCATAG	[143, 144]	100	100	40.5	40.23	40.4	56.70	57.64	57.47
AGTCATGACT	[143, 144]	100	100	32.3	31.47	31.41	48.66	47.22	47.08
GCGAATTCGC	[143, 144]	100	100	43.1	39.45	41.22	56.51	53.97	54
ATCGCTAGCGAT	[143, 144]	100	100	43.1	44.49	44.48	63.30	60.93	61.16
GACGACGTCTGC	[143, 144]	100	100	48.2	48.89	48.93	64.46	62.50	62.82
GCAAGCCGGCTTGC	[143, 144]	100	100	58.5	58.60	58.87	73.38	73.97	73.26
CGATCGGCCGATCG	[143, 144]	100	100	56.9	56.87	70.02	70.02	72.21	70.45
CATATGGCCATATG	[143, 144]	100	100	40.7	41.50	41.68	58.26	58.84	58.77
CAAGATCGATCTTG	[143, 144]	100	100	44.6	44.53	44.94	57.70	57.69	58.86
CGCGTACGCGTACGCG	[143, 144]	100	100	57.9	57.45	55.67	79.50	74.42	74.54
CGCAAGCCGGCTTGC	[143, 144]	100	100	64.2	63.29	63.72	77.36	78.97	78.87
GGCAGTTC	[144]	100	100	30.8	31.49	31.08	41.45	45.27	44.66
GGTTCAGC	[144]	100	100	29.4	31.52	31.1	41.45	45.28	44.67
CGCTGTAG	[144]	100	100	31.7	31.54	31.24	46.88	47.75	48.07
CGTGCTAG	[144]	100	100	31.9	31.55	31.25	46.88	47.76	48.08
AGTAACGCCAT	[144]	100	100	38.3	38.35	38.35	55.23	55.75	55.49
AATGCCGTAGT	[144]	100	100	37.8	38.35	38.34	55.23	55.75	55.48
CCATCGCTACC	[144]	100	100	43.4	43.87	43.76	57.75	60.32	59.87
CGATGGCCTAC	[144]	100	100	44.1	43.86	43.78	57.75	60.32	59.88
CGTTGTTAC	[144]	100	100	35.2	33.39	33.19	49.42	49.18	48.74
CCGTAACGTTGG	[144]	100	100	44.2	44.91	44.92	59.34	61.01	49.74
ACTGACTGACTG	[144]	100	100	42.2	42.25	42.46	57.17	57.49	50.74
ACTGACTGACTGACTG	[144]	100	100	51.0	50.41	51.04	65.30	65.68	51.74

**Tabela 23**

Sequências de DNA em PEG200/40%wt utilizados no trabalho nucleotídeos em ambiente que simula o meio celular e suas respectivas sequências controle (DNA-W/40), temperaturas calculadas se estivessem apenas em H<sub>2</sub>O. Fitas principais estão no sentido 5' para 3'.  $T_{exp}$  são as temperaturas de desnaturação medidas,  $T_{calc}$  são as temperaturas calculadas se as sequências estivessem inseridas apenas em H<sub>2</sub>O utilizando o modelo de próximos vizinhos descrito na subseção 3.2,  $T'_{PB}$  as calculadas utilizando o modelo PB descrito na subseção 3.1 e  $T'_{HMS-}$  as calculadas utilizando o modelo HMS com sinal  $s=-1$  descrito na subseção 5.3. Concentrações de sal [Na<sup>+</sup>] estão em mM e concentrações totais das fitas  $C_t$  em  $\mu$ M. Todas as temperaturas estão em °C.

Sequence	Ref.	$C_t$	[Na <sup>+</sup> ]	RNA-P200/20			RNA-W/20		
				$T_{exp}$	$T'_{PB}$	$T'_{HMS-}$	$T_{calc}$	$T'_{PB}$	$T'_{HMS-}$
UUAUCGAUAA	[147]	100	1000	43.4	43.22	43.09	48.96	49.37	49.41
UAUCGAUA	[147]	100	1000	37.7	36.33	36.45	41.45	41.08	41.16
UCAUGA	[146]	100	1000	22.9	23.79	23.62	29.46	30.00	30.07
ACUGC	[146]	100	1000	38.1	39.01	39.05	46.98	46.62	46.65
AUGGAC	[146]	100	1000	33.2	33.82	33.66	39.60	39.49	39.50
GCGAUA	[146]	100	1000	30.0	30.58	30.58	36.63	36.16	36.13
GCUAUG	[146]	100	1000	30.0	30.94	30.93	36.12	36.39	36.28
ACCGGU	[146]	100	1000	45.6	47.15	47.23	51.83	52.57	52.70
AGCGCU	[146]	100	1000	43.6	43.55	43.59	51.64	52.33	52.41
CACGUG	[146]	100	1000	37.8	36.22	36.30	42.41	42.22	42.22
CAGCUG	[146]	100	1000	38.0	39.10	39.16	46.60	46.45	46.46
CCAUGG	[146]	100	1000	41.4	40.83	40.93	46.60	46.72	46.76
CCUAGG	[146]	100	1000	44.2	43.87	44.05	48.11	47.83	47.81
CUGCAG	[146]	100	1000	40.0	39.10	39.16	46.60	46.45	46.46
GACGUC	[146]	100	1000	42.1	40.34	40.36	45.11	44.71	44.67
GAGCUC	[146]	100	1000	43.7	43.20	43.17	48.96	48.87	48.85
GCAUGC	[146]	100	1000	43.3	41.00	40.98	48.27	48.35	48.33
CGCGCG	[146]	100	1000	49.7	49.68	49.62	58.49	58.53	58.45
CGGCCG	[146]	100	1000	54.0	55.13	55.12	62.24	61.84	61.86
GCCGGC	[146]	100	1000	60.1	60.38	60.19	66.61	66.72	66.66
UAUAUAUA	[146]	100	1000	21.1	21.09	20.99	21.67	21.57	21.63
UAAUAUUA	[146]	100	1000	18.2	18.48	18.49	16.99	17.56	17.56
AACUAGUU	[146]	100	1000	38.9	37.09	37.02	41.22	42.07	42.01
ACUAUAGU	[146]	100	1000	37.7	39.23	39.10	44.19	44.39	44.35
AGAUAUUCU	[146]	100	1000	37.2	37.65	37.74	43.66	44.06	44.06
GAUAUAUC	[146]	100	1000	36.1	35.15	35.12	39.39	38.83	38.80
GAAUAUUC	[146]	100	1000	32.6	32.78	32.84	36.59	35.94	35.92
AACCGGUU	[146]	100	1000	54.1	52.39	52.39	58.65	58.38	58.36
ACUGCAGU	[146]	100	1000	52.1	53.67	53.70	62.32	61.44	61.48
GCAAUUGC	[146]	100	1000	46.9	48.21	48.29	54.95	54.97	54.95
GAACGUUC	[146]	100	1000	46.0	47.25	47.27	52.48	51.92	51.90
AGCGCGCU	[146]	100	1000	64.7	64.21	64.18	73.90	73.50	73.50
AGCCGGCU	[146]	100	1000	69.7	67.96	67.87	76.78	76.26	76.29
GCGAUCGC	[146]	100	1000	60.0	60.74	60.85	67.08	68.67	68.68
GACCGGUC	[146]	100	1000	64.8	65.56	65.50	70.25	70.92	70.92

**Tabela 24**

Sequências de RNA em PEG200/20%wt utilizados no trabalho nucleotídeos em ambiente que simula o meio celular e suas respectivas sequências controle (RNA-W/20), temperaturas calculadas se estivessem apenas em H<sub>2</sub>O. Fitas principais estão no sentido 5' para 3'.  $T_{exp}$  são as temperaturas de desnaturação medidas,  $T_{calc}$  são as temperaturas calculadas se as sequências estivessem inseridas apenas em H<sub>2</sub>O utilizando o modelo de próximos vizinhos descrito na subseção 3.2,  $T'_{PB}$  as calculadas utilizando o modelo PB descrito na subseção 3.1 e  $T'_{HMS-}$  as calculadas utilizando o modelo HMS com sinal  $s=-1$  descrito na subseção 5.3. Concentrações de sal [Na<sup>+</sup>] estão em mM e concentrações totais das fitas  $C_t$  em  $\mu$ M. Todas as temperaturas estão em °C.

Sequence	Ref.	$C_t$	[Na <sup>+</sup> ]	RNA-P200/40			RNA-W/40		
				$T_{exp}$	$T'_{PB}$	$T'_{HMS-}$	$T_{calc}$	$T'_{PB}$	$T'_{HMS-}$
CAUGCC	[145]	100	100	31.0	31.27	31.28	38.72	38.49	38.56
CUAGGC	[145]	100	100	34.3	31.40	31.28	43.45	43.71	43.68
CAGCGG	[145]	100	100	36.4	36.79	36.63	48.26	48.24	48.30
CUACGC	[145]	100	100	27.8	28.90	28.67	40.14	40.99	40.91
GAACUCC	[145]	100	100	34.1	34.20	34.22	46.07	45.78	45.75
GGCUGUCC	[145]	100	100	54.3	54.81	54.97	66.05	65.70	65.68
GGCUGUUC	[145]	100	100	45.6	46.04	46.19	56.40	56.46	56.44
GGUUCUGC	[145]	100	100	44.9	46.04	46.19	56.40	56.46	56.44
AGCUGUCU	[145]	100	100	42.4	42.71	42.81	54.34	54.11	54.18
AGUCUGCU	[145]	100	100	42.6	42.71	42.81	54.34	54.11	54.18
GGCAGUUC	[145]	100	100	46.2	46.04	46.19	56.40	56.46	56.44
GGUUCAGC	[145]	100	100	45.1	46.04	46.19	56.40	56.46	56.44
CGCUGUCG	[145]	100	100	49.8	50.55	50.68	59.38	59.49	59.41
CGUCUGCG	[145]	100	100	50.3	50.55	50.68	59.38	59.49	59.41
UGCUGUCA	[145]	100	100	43.0	43.05	43.21	53.14	52.83	52.83
UGUCUGCA	[145]	100	100	42.9	43.05	43.21	53.14	52.83	52.83
UAUGAGGA	[145]	100	100	34.7	35.55	35.73	45.18	44.79	44.82
UAGAUGGA	[145]	100	100	35.2	35.55	35.73	45.18	44.79	44.82
GUGCCGAG	[145]	100	100	54.3	52.72	52.84	62.77	62.66	62.61
GCCGAGUG	[145]	100	100	54.1	52.71	52.84	62.77	62.66	62.61
CGCUGUAG	[145]	100	100	44.1	45.19	45.34	54.82	54.99	54.85
CGUGCUAG	[145]	100	100	45.8	45.19	45.34	54.82	54.99	54.86
AUUGGAUACAAA	[145]	100	100	42.3	41.71	41.60	49.74	49.73	49.73
AUACAUUGGAAA	[145]	100	100	41.6	41.71	41.60	49.74	49.73	49.74
GGCUCAAUUGAC	[145]	100	100	56.3	57.55	57.41	63.64	64.07	64.10
UACAUGUA	[145]	100	100	28.2	26.36	26.42	36.37	36.94	36.88
GGUAUACC	[145]	100	100	39.2	40.09	40.19	51.80	51.35	51.29
GGUUAACC	[145]	100	100	36.5	36.78	37.11	49.42	49.34	49.39
CGAAUUCG	[145]	100	100	31.4	32.77	33.08	39.16	38.33	38.36
GCUUAAGC	[145]	100	100	38.6	37.74	37.81	47.99	47.74	47.72
CGUAUACG	[145]	100	100	35.7	35.53	35.64	44.11	44.56	44.45
CGCAUGCG	[145]	100	100	52.7	51.58	51.88	57.18	56.94	56.93
CGGUACCG	[145]	100	100	51.0	50.46	50.82	62.72	62.13	62.14
CAAGCUUG	[145]	100	100	35.8	36.33	36.61	43.98	44.35	44.36
CUUGCAAG	[145]	100	100	38.3	36.33	36.61	43.98	44.35	44.36
GAACGUUC	[145]	100	100	36.8	34.57	34.69	45.33	45.03	44.99
GUUCGAAC	[145]	100	100	35.4	34.57	34.69	45.33	45.03	44.99
GAUCCGGAUC	[145]	100	100	58.4	57.11	57.24	62.28	62.74	62.74
GGAUCGAUCC	[145]	100	100	56.0	57.11	57.24	62.28	62.74	62.74
AUGAGCUCAU	[145]	100	100	48.8	47.94	47.96	54.28	54.28	54.30
AUCAGCUGAU	[145]	100	100	46.7	47.94	47.96	54.28	54.29	54.30
UUACGCGUAA	[145]	100	100	42.2	42.90	43.05	54.06	54.28	54.26
AUCGCUAGCGAU	[145]	100	100	60.0	59.88	59.68	66.82	65.92	65.97
CAUAGGCCUAUG	[145]	100	100	59.1	59.42	59.29	66.01	65.28	65.29
CUAUGGCCUAUG	[145]	100	100	59.4	59.42	59.29	66.01	65.28	65.30

**Tabela 25**

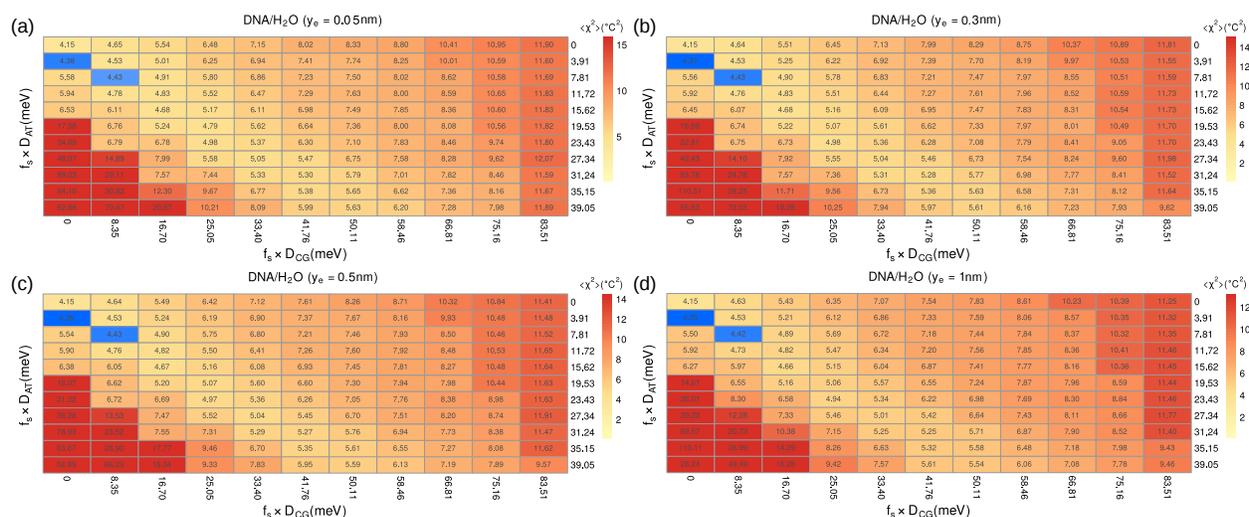
Sequências de RNA em PEG200/40%wt utilizados no trabalho nucleotídeos em ambiente que simula o meio celular e suas respectivas sequências controle (RNA-W/40), temperaturas calculadas se estivessem apenas em H<sub>2</sub>O. Fitas principais estão no sentido 5' para 3'.  $T_{exp}$  são as temperaturas de desnaturação medidas,  $T_{calc}$  são as temperaturas calculadas se as sequências estivessem inseridas apenas em H<sub>2</sub>O utilizando o modelo de próximos vizinhos descrito na subseção 3.2,  $T'_{PB}$  as calculadas utilizando o modelo PB descrito na subseção 3.1 e  $T'_{HMS-}$  as calculadas utilizando o modelo HMS com sinal  $s=-1$  descrito na subseção 5.3. Concentrações de sal [Na<sup>+</sup>] estão em mM e concentrações totais das fitas  $C_t$  em  $\mu$ M. Todas as temperaturas estão em °C.

## APÊNDICE B - Procura das energias de barreira iniciais para os modelos HMS, PCLA, TRMF e PCLJ

Neste apêndice, apresentamos as figuras de varredura do  $\langle \chi^2 \rangle$  dos modelos HMS, PCLA, TRMF e PCLJ. O processo de varredura consiste em calcular os valores de  $\chi^2$  para cada valor de energia de barreira em um intervalo fixo. Por exemplo, no intervalo entre 0 e 1 do parâmetro  $f_s$  do modelo HMS (página 79), com um passo de 0.1 teremos 11 pontos. Para cada ponto, teremos um valor de  $\chi^2$  diferente. Deste modo, teremos uma ideia de como o parâmetro  $f_s$  afeta as predições de temperatura. O DNA canônico apresenta dois tipos de pares de bases, AT e CG, que interagem diferente com o solvente devido à sua estrutura química. Logo, o valor de  $\chi^2$  será determinado por um valor de  $f_s$  (AT) para AT e um  $f_s$  (CG) para CG. Ao final, teremos uma matriz 11x11 de  $\chi^2$  do espectro da região analisada. A função  $\chi^2$  é uma soma quadrática da diferença entre as temperaturas experimentais e preditas, conforme a equação (30), ou seja,  $\chi^2$  aumenta à medida que aumentamos o número total  $N$  de sequências. Deste modo, utilizamos o valor de  $\langle \chi^2 \rangle$ , que é o valor de  $\chi^2$  dividido pelo número de sequências  $N$  presente em cada base de dados, DNA-P200/40 e DNA-W/40. Utilizamos os 11 valores de energia do modelo HMS para construir a matriz de  $\langle \chi^2 \rangle$  dos outros modelos. Os demais parâmetros dos modelos de barreira foram mantidos constantes, cujo os valores estão na tabela 12.

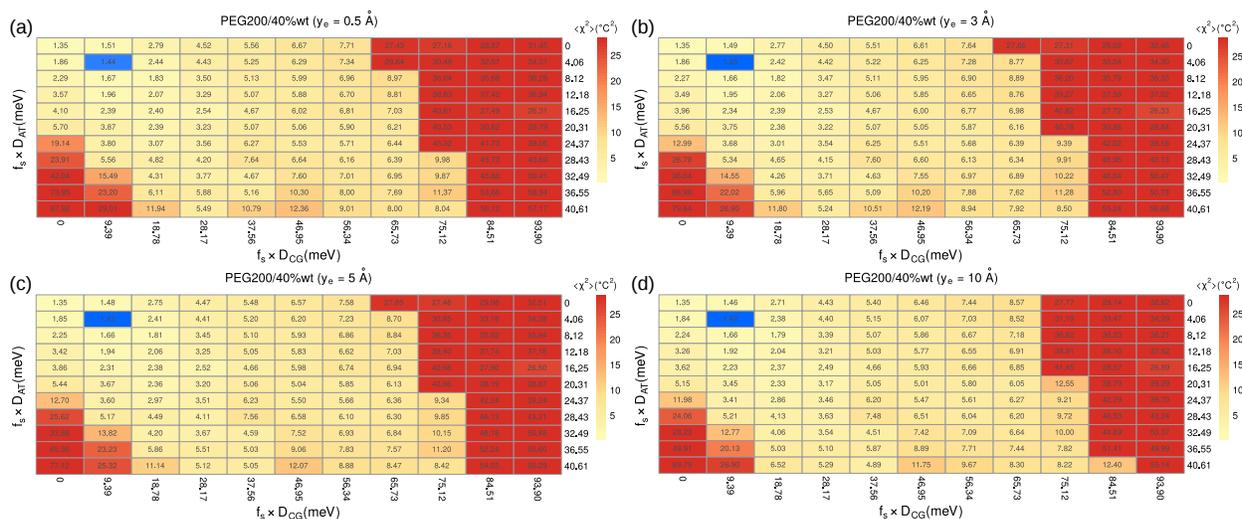
**Modelo HMS** Variamos a faixa entre 0 e 1 do parâmetro  $f_s$  com passo de 0.1. O parâmetro que determina a energia da barreira é  $f_s D$ . Assim, ao invés de montarmos o gráfico em função de  $f_s$  (CG) *versus*  $f_s$  (AT), montamos o gráfico  $f_s D_{CG}$  *versus*  $f_s D_{AT}$  para que todos os modelos sejam analisados em unidades de energia. As matrizes de  $\langle \chi^2 \rangle$  para as sequências DNA-W/40 estão mostradas nas figuras 45 e 47, enquanto as DNA-P200/40, são mostradas nas figuras 46 e 48. Destacamos os pontos com tonalidades em azul que representam os menores valores de  $\langle \chi^2 \rangle$  para o modelo HMS<sup>-</sup>. O ponto onde  $f_s D_{CG} = f_s D_{AT} = 0$  representa o  $\langle \chi^2 \rangle$  do modelo PB.

Primeiramente, analisarmos as figuras 45 e 46, que são referentes ao modelo HMS<sup>-</sup>, equação (34) com  $s=-1$ . Notamos, em todos os gráficos, dois pontos em azul na figura 45 e um ponto em azul na figura 46. Descartaremos o ponto em que  $f_s D_{CG}$  é zero, enquanto  $f_s D_{AT}$  não, pois não faz sentido o solvente ter efeito apenas no par de base AT. AO analisarmos quimicamente o par CG e o par AT, o solvente, no caso a H<sub>2</sub>O, afetaria mais o par CG do que o par AT. Logo, a barreira energética do par CG deve ser um pouco maior do que o par AT. Isso é evidente no outro ponto



**Figura 45**

Modelo HMS<sup>-</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $14 \text{ }^{\circ}C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm, equação (34).



**Figura 46**

Modelo HMS<sup>-</sup>. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $30 \text{ }^{\circ}C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm.

Parâmetro	DNA-W/40	DNA-P200/40
$D_{AT}$	39.10 meV	40.61 meV
$D_{CG}$	83.51 meV	93.90 meV
$k_{AT-AT}$	1.35 eV/nm <sup>2</sup>	3.03 eV/nm <sup>2</sup>
$k_{AT-CG}$	2.58 eV/nm <sup>2</sup>	1.80 eV/nm <sup>2</sup>
$k_{AT-GC}$	4.84 eV/nm <sup>2</sup>	4.65 eV/nm <sup>2</sup>
$k_{AT-TA}$	3.53 eV/nm <sup>2</sup>	1.22 eV/nm <sup>2</sup>
$k_{CG-AT}$	3.29 eV/nm <sup>2</sup>	4.78 eV/nm <sup>2</sup>
$k_{CG-CG}$	2.55 eV/nm <sup>2</sup>	2.30 eV/nm <sup>2</sup>
$k_{CG-GC}$	3.92 eV/nm <sup>2</sup>	4.35 eV/nm <sup>2</sup>
$k_{GC-AT}$	1.26 eV/nm <sup>2</sup>	2.55 eV/nm <sup>2</sup>
$k_{GC-CG}$	2.22 eV/nm <sup>2</sup>	1.91 eV/nm <sup>2</sup>
$k_{TA-AT}$	0.08 eV/nm <sup>2</sup>	0.05 eV/nm <sup>2</sup>
$f_s$ (AT)	0.2	0.1
$f_s$ (CG)	0.1	0.1
$\lambda_s$ (AT)	17.5 nm	17.5 nm
$\lambda_s$ (CG)	22.5 nm	22.5 nm
$y_e$ (AT) = $y_e$ (CG)	0.05, 0.3, 0.5 e 1 nm	0.05, 0.3, 0.5 e 1 nm
$s_{AT} = s_{CG}$	-1	-1

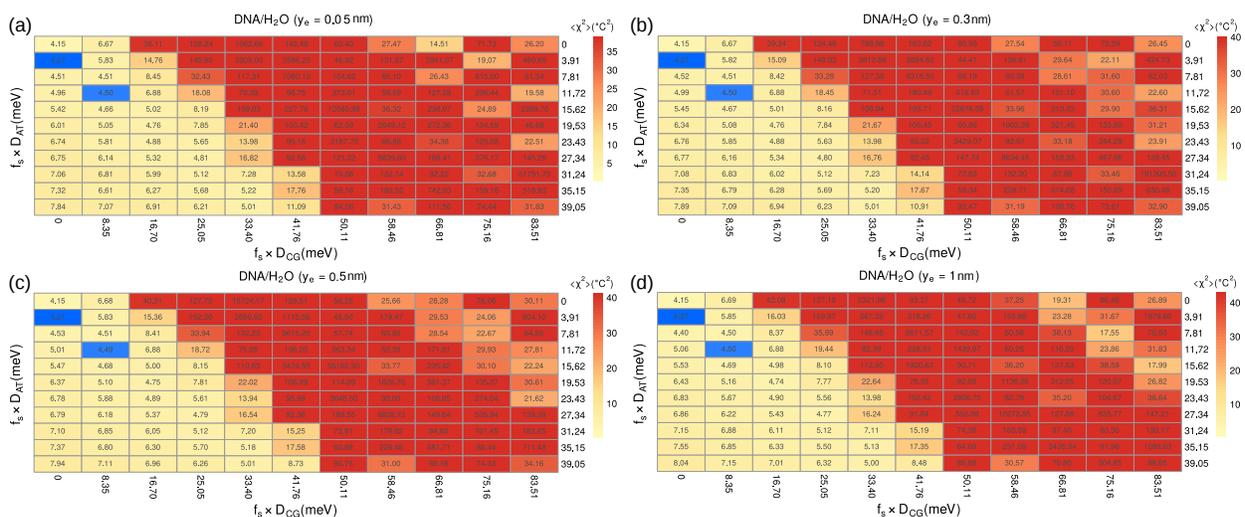
**Tabela 26**

Parâmetros iniciais utilizados para as minimizações do modelo HMS<sup>-</sup>.

em azul em DNA-W/40 (figuras 45) e no ponto em azul do conjunto DNA-P200/40 (figuras 45). Portanto, utilizaremos os valores iniciais  $f_s$  (CG) = 0.1 para DNA-W/40 e DNA-P200/40,  $f_s$  (AT) = 0.2 para DNA-W/40 e  $f_s$  (AT) = 0.1 para DNA-P200/40. Um resumo com os parâmetros iniciais das minimizações do modelo HMS<sup>-</sup> se encontra na tabela 26.

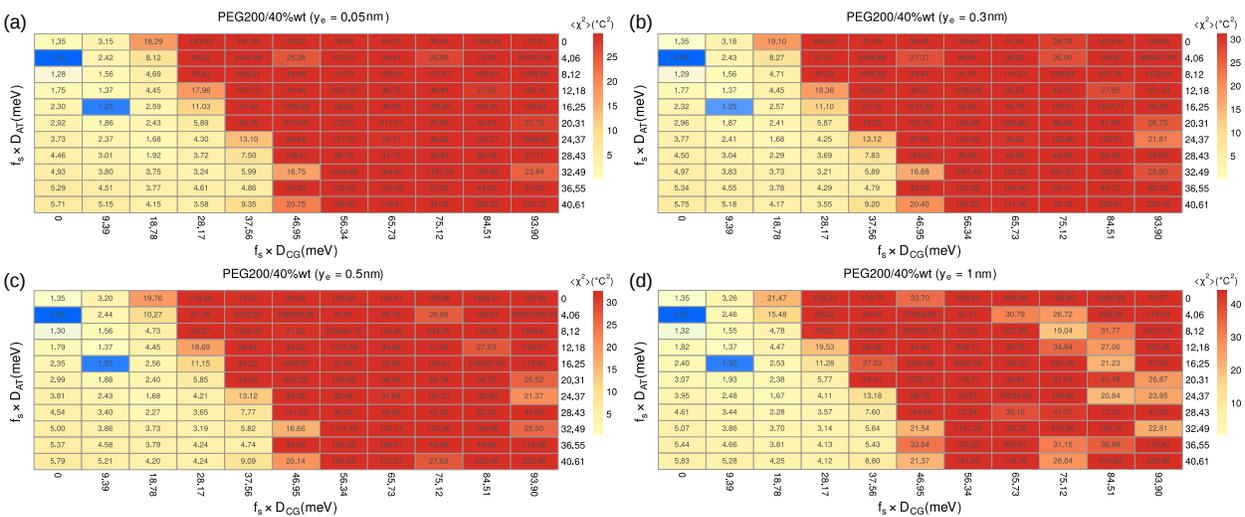
Analisando os gráficos das figuras 47 e 48 do modelo HMS<sup>+</sup>, notamos pontos melhores do que o modelo PB nos gráficos da figura 48. No entanto, um dos pontos possui valor de energia de barreira CG igual a zero, o que não faz sentido, já que o solvente atua em ambos os pares de bases CG e AT. Logo, descartamos este ponto. O mesmo processo de exclusão foi realizado para um dos pontos em azul da figura 47. Sobrou apenas um ponto em azul para as sequências DNA-W/40 (figura 47) e um ponto para DNA-P200/40 (figura 48). Portanto, utilizamos os valores iniciais  $f_s$  (CG) = 0.1 para DNA-W/40 e DNA-P200/40,  $f_s$  (AT) = 0.3 para DNA-W/40 e  $f_s$  (AT) = 0.4 para DNA-P200/40. Um resumo com os parâmetros iniciais das minimizações do modelo HMS<sup>+</sup> se encontra na tabela 27.

Como forma de validar a eficiência dos modelos HMS<sup>-</sup> e HMS<sup>+</sup>, realizamos as três minimizações em valores diferentes de  $y_e$ . Apesar de que nenhuma mudança significativa dos valores de  $\langle \chi^2 \rangle$  apareceu nos gráficos das figuras 45, 47, 46 e 48, esperamos que, através das minimizações, encontramos valores de altura de barreira que apresentem alguma relação com a distância de equilíbrio  $y_e$ . Caso isso seja confirmado, os modelos HMS<sup>-</sup> e HMS<sup>+</sup> seriam úteis em determinar a interação das moléculas de água localizadas em diferentes posições distantes do nucleotídeo.



**Figura 47**

Modelo HMS+. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $40 \text{ }^{\circ}C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm.



**Figura 48**

Modelo HMS+. Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $f_s D$  para AT e CG em diferentes distâncias de equilíbrio  $y_e$ . Colorimos todos os pontos maiores que  $30 \text{ }^{\circ}C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_e = 0.05$  nm, (b)  $y_e = 0.3$  nm, (c)  $y_e = 0.5$  nm, (d)  $y_e = 1$  nm.

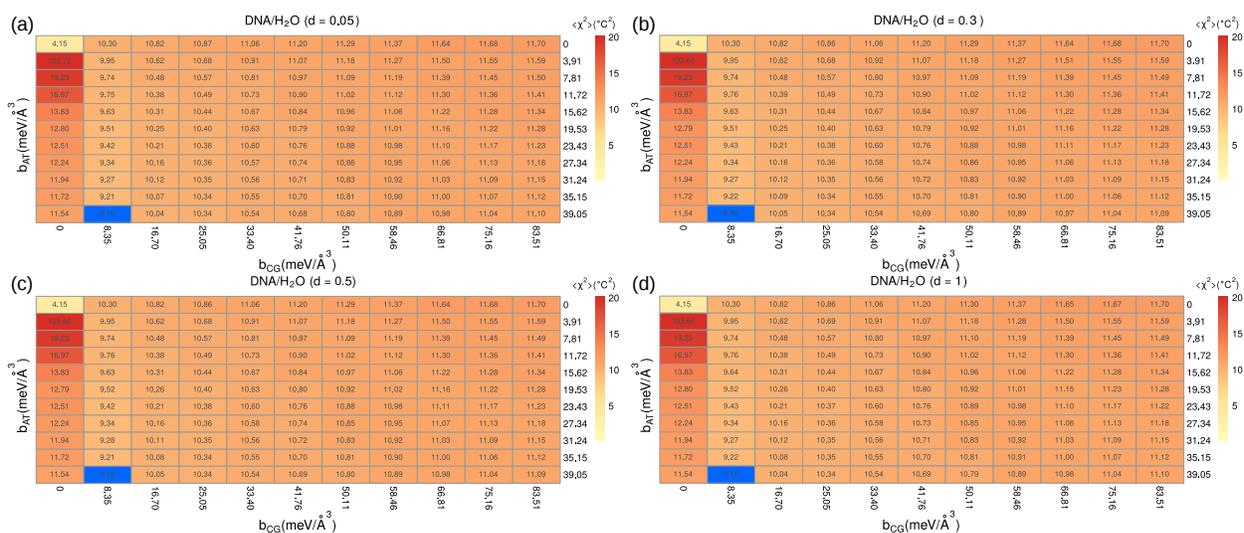
Parâmetro	DNA-W/40	DNA-P200/40
$D_{AT}$	39.10 meV	40.61 meV
$D_{CG}$	83.51 meV	93.90 meV
$k_{AT-AT}$	1.35 eV/nm <sup>2</sup>	3.03 eV/nm <sup>2</sup>
$k_{AT-CG}$	2.58 eV/nm <sup>2</sup>	1.80 eV/nm <sup>2</sup>
$k_{AT-GC}$	4.84 eV/nm <sup>2</sup>	4.65 eV/nm <sup>2</sup>
$k_{AT-TA}$	3.53 eV/nm <sup>2</sup>	1.22 eV/nm <sup>2</sup>
$k_{CG-AT}$	3.29 eV/nm <sup>2</sup>	4.78 eV/nm <sup>2</sup>
$k_{CG-CG}$	2.55 eV/nm <sup>2</sup>	2.30 eV/nm <sup>2</sup>
$k_{CG-GC}$	3.92 eV/nm <sup>2</sup>	4.35 eV/nm <sup>2</sup>
$k_{GC-AT}$	1.26 eV/nm <sup>2</sup>	2.55 eV/nm <sup>2</sup>
$k_{GC-CG}$	2.22 eV/nm <sup>2</sup>	1.91 eV/nm <sup>2</sup>
$k_{TA-AT}$	0.08 eV/nm <sup>2</sup>	0.05 eV/nm <sup>2</sup>
$f_s$ (AT)	0.3	0.4
$f_s$ (CG)	0.1	0.1
$\lambda_s$ (AT)	17.5 nm	17.5 nm
$\lambda_s$ (CG)	22.5 nm	22.5 nm
$y_e$ (AT) = $y_e$ (CG)	0.05, 0.3, 0.5 e 1 nm	0.05, 0.3, 0.5 e 1 nm
$s_{AT} = s_{CG}$	+1	+1

**Tabela 27**

Parâmetros iniciais utilizados para as minimizações do modelo HMS+.

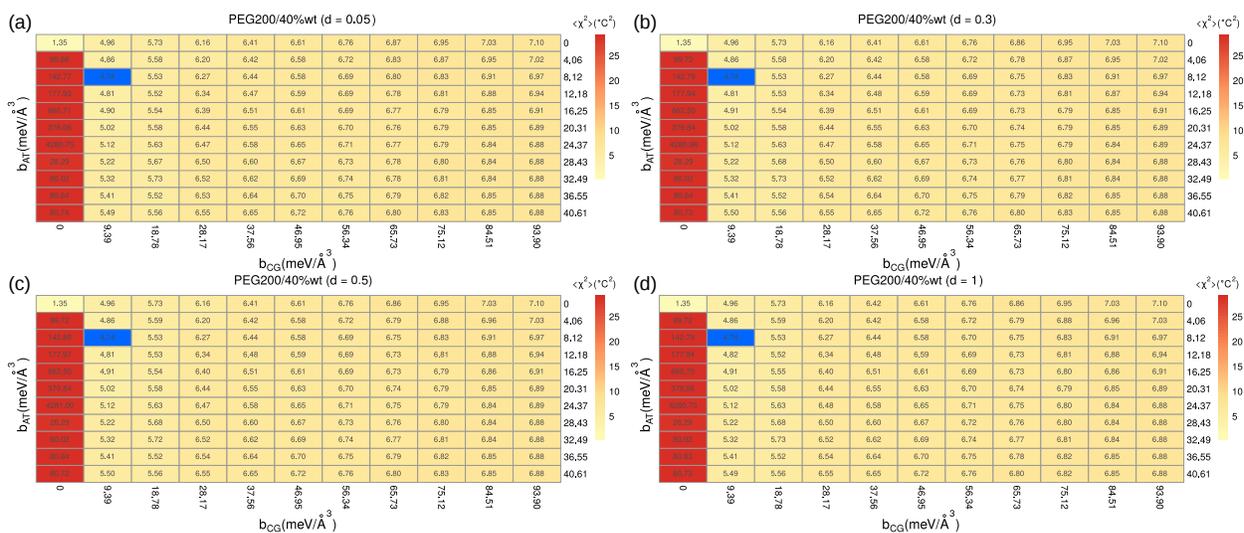
**Modelo PCLA** Para analisar todos os modelos sob as mesmas condições, utilizamos as unidades de energia de barreira do modelo HMS<sup>-</sup>, ou HMS<sup>+</sup> para observar o comportamento das energias  $b_{AT}$  e  $b_{CG}$  do modelo PCLA. Uma parte do Hamiltoniano do modelo PCLA consiste em  $V_{Morse} + V_{barreira}$ , que é similar ao modelo HMS. Assim, aplicamos o mesmo princípio de que a energia da barreira não deve exceder o valor  $D$  da profundidade do poço do potencial de Morse.

As figuras 49 e 50 consistem em mapas de calor do  $\langle \chi^2 \rangle$  de  $b_{CG}$  versus  $b_{AT}$  com os demais parâmetros constantes, iguais aos valores da tabela 12. Os valores de  $\langle \chi^2 \rangle$  do modelo apresentam pouca variação com o parâmetro  $d$ . Além disso, todos os valores de  $\langle \chi^2 \rangle$  do modelo PCLA são muito piores do que os do modelo PB. Para ter uma ideia de quão ruim o modelo PCLA é nas predições de temperatura em comparação com o modelo PB, observamos a razão entre o melhor valor de  $\langle \chi^2 \rangle$  do modelo PCLA (ponto em azul) e o modelo PB (ponto  $b_{CG} = b_{AT} = 0$ ). O modelo PCLA é aproximadamente  $2.2 \times$  e  $3.5 \times$  pior que o modelo PB para prever temperaturas de desnaturação. Entretanto, faremos o parecer final em relação ao modelo PCLA após a rodada de teste das minimizações discutida na seção 5.6.4. De acordo com os gráficos das figuras 49 e 50, precisamente o ponto em azul, escolheríamos os seguintes valores iniciais  $b_{CG} = 8.35 \text{ meV}/\text{Å}^3$  para DNA-W/40,  $b_{CG} = 9.39 \text{ meV}/\text{Å}^3$  para DNA-P200/40,  $b_{AT} = 39.05/\text{meV}\text{Å}^3$  para DNA-W/40 e  $b_{AT} = 8.12 \text{ meV}/\text{Å}^3$  para DNA-P200/40. Os valores de  $b_{AT}$  das sequências DNA-W/40 e DNA-P200/40 são muito diferentes. Uma barreira de  $b_{AT}$  muito maior do que  $b_{CG}$  não faria sentido, pois o par CG é mais hidratado que o par AT, ou AU no caso do RNA [34, 167]. Observamos que os valores de  $\langle \chi^2 \rangle$  na coluna  $b_{CG} = 8.35 \text{ meV}/\text{Å}^3$  apresentados na figura 49 das sequências DNA-W/40 são



**Figura 49**

Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $20 \text{ } ^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $d = 0.05$ , (b)  $d = 0.3$ , (c)  $d = 0.5$ , (d)  $d = 1$ .



**Figura 50**

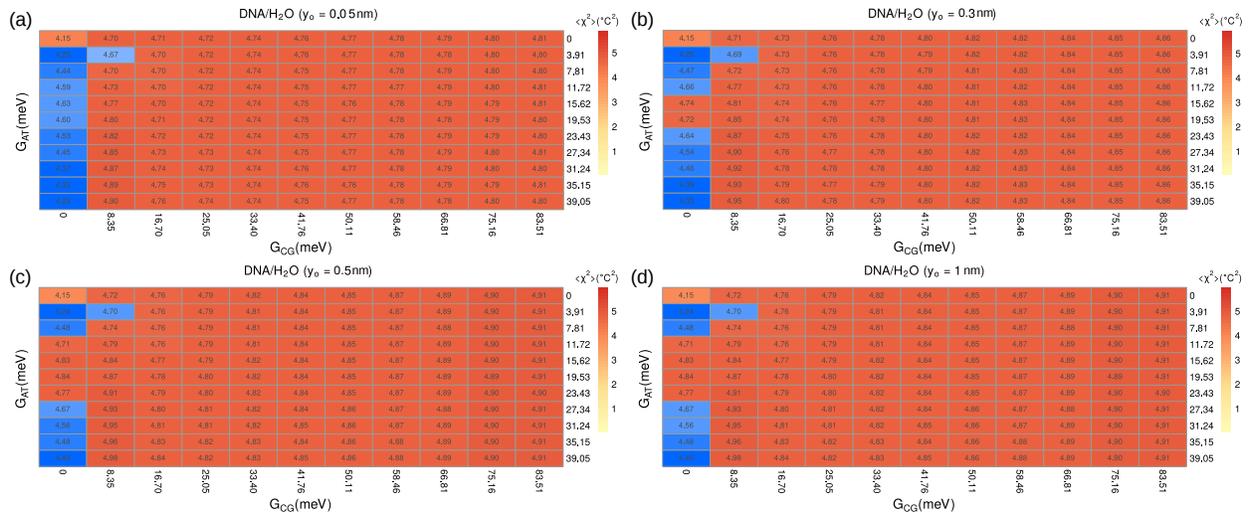
Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $b$  para AT e CG em diferentes distâncias de equilíbrio  $d$ . Colorimos todos os pontos maiores que  $14 \text{ } ^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $d = 0.05$ , (b)  $d = 0.3$ , (c)  $d = 0.5$ , (d)  $d = 1$ .

Parâmetro	DNA-W/40	DNA-P200/40
$D_{AT}$	39.10 meV	40.61 meV
$D_{CG}$	83.51 meV	93.90 meV
$k_{AT-AT}$	1.35 eV/nm <sup>2</sup>	3.03 eV/nm <sup>2</sup>
$k_{AT-CG}$	2.58 eV/nm <sup>2</sup>	1.80 eV/nm <sup>2</sup>
$k_{AT-GC}$	4.84 eV/nm <sup>2</sup>	4.65 eV/nm <sup>2</sup>
$k_{AT-TA}$	3.53 eV/nm <sup>2</sup>	1.22 eV/nm <sup>2</sup>
$k_{CG-AT}$	3.29 eV/nm <sup>2</sup>	4.78 eV/nm <sup>2</sup>
$k_{CG-CG}$	2.55 eV/nm <sup>2</sup>	2.30 eV/nm <sup>2</sup>
$k_{CG-GC}$	3.92 eV/nm <sup>2</sup>	4.35 eV/nm <sup>2</sup>
$k_{GC-AT}$	1.26 eV/nm <sup>2</sup>	2.55 eV/nm <sup>2</sup>
$k_{GC-CG}$	2.22 eV/nm <sup>2</sup>	1.91 eV/nm <sup>2</sup>
$k_{TA-AT}$	0.08 eV/nm <sup>2</sup>	0.05 eV/nm <sup>2</sup>
$b_{AT}$	7.81 meV/Å <sup>3</sup>	8.12 meV/Å <sup>3</sup>
$b_{CG}$	8.35 meV/Å <sup>3</sup>	9.39 meV/Å <sup>3</sup>
$c_{AT}$	0.057	0.057
$c_{CG}$	0.044	0.044
$d_{AT} = d_{CG}$	0.05, 0.3, 0.5 e 1	0.05, 0.3, 0.5 e 1
$\alpha_{AT}$	2.78 nm <sup>-1</sup>	2.78 nm <sup>-1</sup>
$\alpha_{CG}$	9.39 nm <sup>-1</sup>	9.39 nm <sup>-1</sup>

**Tabela 28**

Parâmetros iniciais utilizados para as minimizações do modelo PCLA. Os valores diferentes de  $d$  são relativos a posições de equilíbrio entre o solvente e os nucleotídeos.

bem próximos. Esta pequena diferença nos valores  $\langle \chi^2 \rangle$  leva a praticamente aos mesmos valores de predição de temperaturas. Por isso, escolhemos o ponto com  $\langle \chi^2 \rangle = 9.74 \text{ } ^\circ C^2$ , que é exatamente o mesmo ponto das sequências DNA-P200/40 na figura 50. Um resumo com os parâmetros iniciais das minimizações do modelo PCLA se encontra na tabela 28.

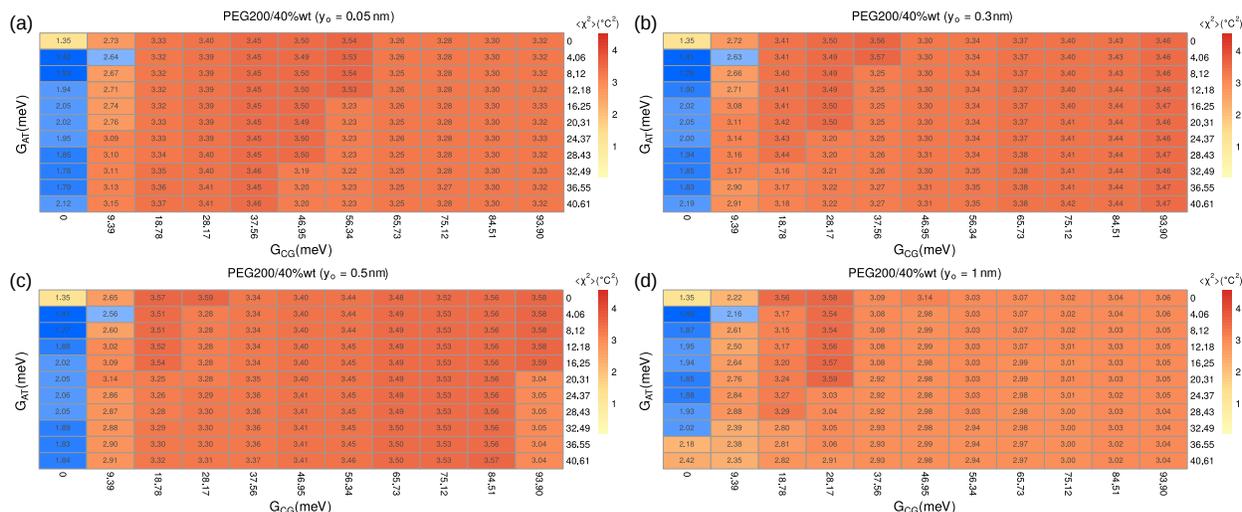


**Figure 51**

Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $G$  para AT e CG em diferentes distâncias de equilíbrio  $y_o$ . Colorimos todos os pontos maiores que  $6 \text{ } ^\circ\text{C}^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_o = 0.05$  nm, (b)  $y_o = 0.3$  nm, (c)  $y_o = 0.5$  nm, (d)  $y_o = 1$  nm.

**Modelo TRMF** O mapa de  $\langle \chi^2 \rangle$  do modelo TRMF para DNA-W/40 (figura 51) e DNA-P200/40 (figura 52) apresenta tonalidades em vermelho e laranja. Colorimos os pontos com os menores valores  $\langle \chi^2 \rangle$  de tonalidades azul, ou seja, aqueles que mais se aproximam das temperaturas experimentais. A maioria dos melhores valores se encontra na região em  $G_{CG} = 0$  e  $G_{AT} \neq 0$ . Isto sugere que ao ajustarmos apenas a altura de barreira  $G$ , tenderemos ao valor zero. Mas, talvez isso não seja verdade para os outros testes, como variar  $G$  e  $D$ , ou variar  $G$  e  $k$ . Logo, sobra apenas um ponto fora dessa região em cada um dos gráficos das figuras 51 e figura 52. Dessa forma, escolhemos os valores iniciais  $G_{CG} = 8.35$  meV para DNA-W/40,  $G_{CG} = 9.39$  meV para DNA-P200/40,  $G_{AT} = 3.91$  para DNA-W/40 e  $b_{AT} = 4.06$  meV para DNA-P200/40. Um resumo com os parâmetros iniciais das minimizações do modelo TRMF se encontra na tabela 29.

Um ponto interessante desse modelo é que, à medida que variamos  $y_o$ , notamos uma pequena melhora nos valores de  $\langle \chi^2 \rangle$ . Talvez o modelo sirva para distinguir águas próximas de águas distantes do DNA. Embora as mudanças nos valores de  $\langle \chi^2 \rangle$  sejam pequenas, como mostrado nas figuras 51 e 47, esperamos que, através das minimizações, encontremos valores de altura de barreira que apresentem alguma relação com a distância de equilíbrio  $y_o$ . Caso isso se confirme, o modelo TRMF seria útil em determinar a interação das moléculas de água localizadas em diferentes posições distantes do nucleotídeo.



**Figura 52**

Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $G$  para AT e CG em diferentes distâncias de equilíbrio  $y_o$ . Colorimos todos os pontos maiores que  $5 \text{ } ^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$  excluindo o ponto (0,0). (a)  $y_o = 0.05$  nm, (b)  $y_o = 0.3$  nm, (c)  $y_o = 0.5$  nm, (d)  $y_o = 1$  nm.

Parâmetro	DNA-W/40	DNA-P200/40
$D_{AT}$	39.10 meV	40.61 meV
$D_{CG}$	83.51 meV	93.90 meV
$k_{AT-AT}$	1.35 eV/nm <sup>2</sup>	3.03 eV/nm <sup>2</sup>
$k_{AT-CG}$	2.58 eV/nm <sup>2</sup>	1.80 eV/nm <sup>2</sup>
$k_{AT-GC}$	4.84 eV/nm <sup>2</sup>	4.65 eV/nm <sup>2</sup>
$k_{AT-TA}$	3.53 eV/nm <sup>2</sup>	1.22 eV/nm <sup>2</sup>
$k_{CG-AT}$	3.29 eV/nm <sup>2</sup>	4.78 eV/nm <sup>2</sup>
$k_{CG-CG}$	2.55 eV/nm <sup>2</sup>	2.30 eV/nm <sup>2</sup>
$k_{CG-GC}$	3.92 eV/nm <sup>2</sup>	4.35 eV/nm <sup>2</sup>
$k_{GC-AT}$	1.26 eV/nm <sup>2</sup>	2.55 eV/nm <sup>2</sup>
$k_{GC-CG}$	2.22 eV/nm <sup>2</sup>	1.91 eV/nm <sup>2</sup>
$k_{TA-AT}$	0.08 eV/nm <sup>2</sup>	0.05 eV/nm <sup>2</sup>
$G_{AT}$	3.91 meV	4.06 meV
$G_{CG}$	8.35 meV	9.39 meV
$b_G$ (AT)	17.5 nm	17.5 nm
$b_G$ (CG)	22.5 nm	22.5 nm
$y_o$ (AT) = $y_o$ (CG)	0.05, 0.3, 0.5 e 1 nm	0.05, 0.3, 0.5 e 1 nm

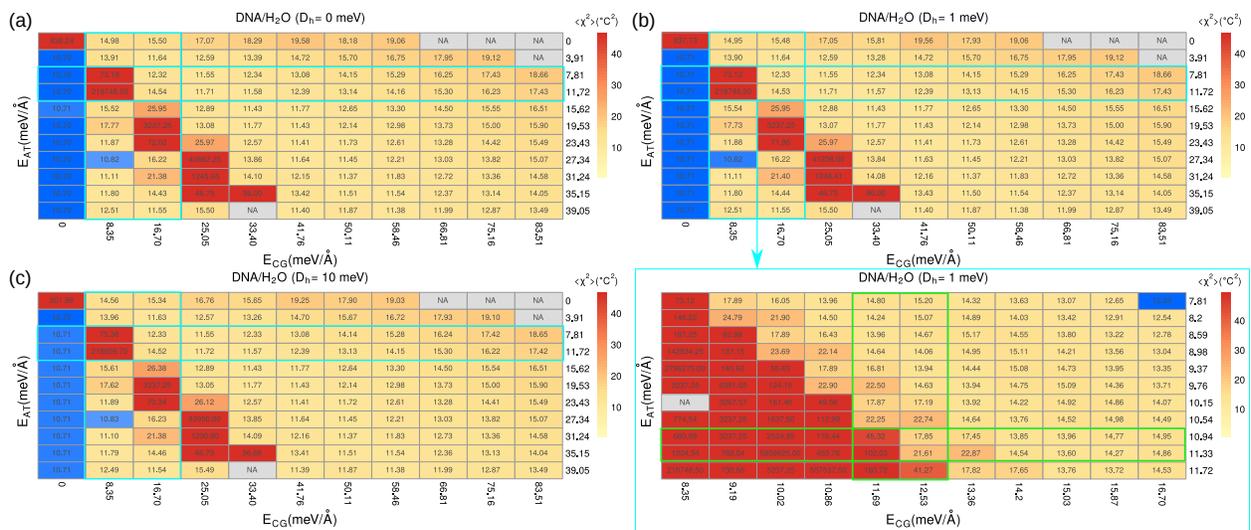
**Tabela 29**

Parâmetros iniciais utilizados para as minimizações do modelo TRMF.

**modelo PCLJ** O modelo PCLJ é completamente diferente dos demais modelos. Suas equações do potencial de Morse e do potencial barreira são acopladas, conforme as expressões (36) e (37). Cada conjunto de variações de parâmetros,  $E$ ,  $D_h$ ,  $\lambda_h$  e  $\gamma$ , produz valores de profundidade de poços  $A$  diferentes. A fim de normalizar os modelos, procuramos os parâmetros que mais se assemelhem ao model PB padrão. Para isso, fixamos os valores  $\gamma_{AT}$  e  $\gamma_{CG}$  nos valores 17.5 nm e 22.5 nm, respectivamente. Colocamos  $D_h=0$ , pois após a dissociação dos pares de base a energia é zero (modelo PB). Após a separação dos pares de base, as moléculas de  $H_2O$  começam a mediar a interação entre os dois nucleotídeos. Observando os pares AT e CG na figura 2c (página 29), notamos apenas uma diferença nos grupos funcionais entre os nucleotídeos. Na posição C(4), especificamente, a guanina possui um grupo  $-NH_2$  e a adenina, um grupo  $-CH$ . O grupo  $-NH_2$ , além de ser mais eletronegativo, forma mais ligações de hidrogênio com a  $H_2O$  em comparação ao grupo  $-CH$ . Portanto, o par CG deve possuir uma energia de barreira um pouco maior do que o par AT. Em soluções onde o solvente é apenas água, não observamos uma terceira ligação de hidrogênio no par AT. A terceira ligação de hidrogênio, realizada pelo grupo funcional  $-CH$  da adenina e  $-C=O$  da timina, pode apreecer apenas quando perturbamos as moléculas de  $H_2O$  próximas ao par AT [4, 168]. No entanto, o grupo  $-CH$  da adenina e  $-C=O$  da timina podem realizar ligações de hidrogênio com as moléculas de  $H_2O$  após a dissociação. Logo, escolhemos o valor de  $E_{AT} \sim 11 \text{ meV}/\text{Å}$  e  $E_{CG} \sim 12 \text{ meV}/\text{Å}$  para a  $H_2O$ . Em seguida, variamos os parâmetros  $\lambda_h$  (AT) e  $\lambda_h$  (CG) até encontrarmos valores de  $A_{AT}$  e  $A_{CG}$  próximos aos valores  $D_{AT}$  e  $D_{CG}$  do modelo PB, mostrados na figura 36. Os valores de  $\lambda_h$  (AT) e  $\lambda_h$  (CG) encontrados foram  $5.87 \times 10^{-3} \text{ nm}$  e  $7.18 \times 10^{-3} \text{ nm}$ , respectivamente. Após determinarmos os valores de  $\lambda_h$  para DNA-W/40, variamos os parâmetros  $E_{AT}$  e  $E_{CG}$  para as sequências de PEG200 de forma a obter valores de profundidade de poço  $A_{AT}$  e  $A_{CG}$  próximos aos valores  $D_{AT}$  e  $D_{CG}$  do modelo PB, conforme a figura 36. Os valores encontrados foram  $E_{AT} \sim 11.2 \text{ meV}/\text{Å}$  e  $E_{CG} \sim 13.4 \text{ meV}/\text{Å}$ .

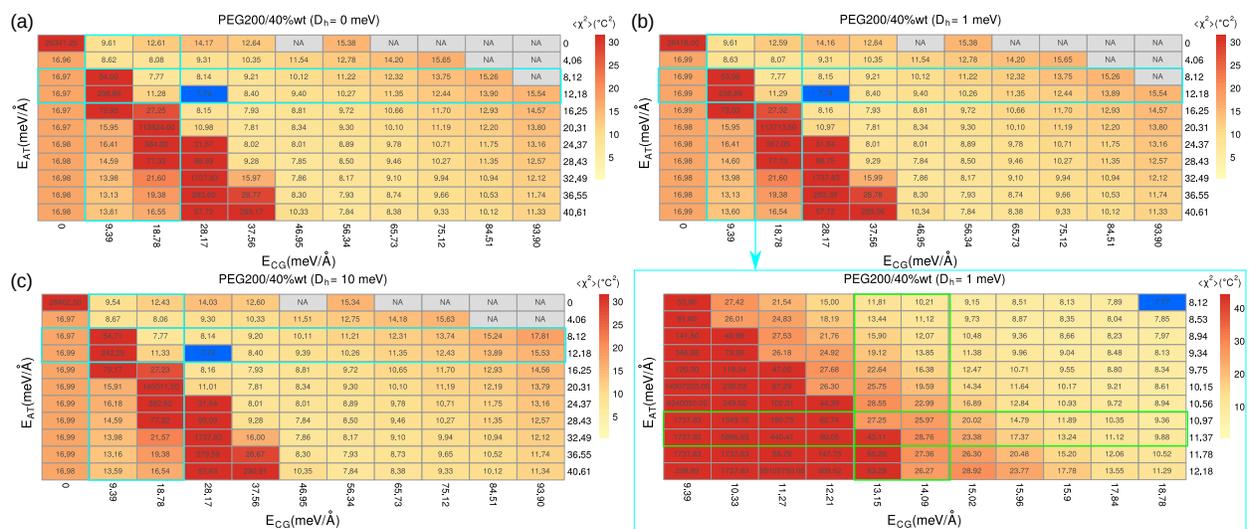
O próximo passo foi avaliar o mapa de  $\langle \chi^2 \rangle$  da altura de barreira  $E$ , mantendo os demais parâmetros nos valores mencionados anteriormente. Análogo aos demais modelos, os parâmetros de empilhamento  $k$  foram mantidos fixos nos valores calculados pelo modelo PB para DNA-W/40 e DNA-P200/40, como discutido na seção 5.5.2. De forma a seguir as mesmas unidades dos demais modelos, avaliamos as mesmas faixas de energia: entre 0 e 83.51 meV para  $E_{CG}$ (DNA-W/40); entre 0 e 93.90 meV para  $E_{CG}$ (DNA-P200/40); entre 0 e 39.05 meV para  $E_{AT}$ (DNA-W/40); e entre 0 e 40.61 meV para  $E_{AT}$ (DNA-P200/40). Apesar de o modelo PB possuir energia zero após as bases se separarem, a existência de uma pequena energia remanescente da interação entre o solvente e as fitas simples faz sentido fisicamente. Logo, analisamos a influência do parâmetro  $D_h$  nas predições de temperatura para três valores diferentes: (a)  $D_h = 0 \text{ meV}$ , (b)  $D_h = 1 \text{ meV}$  e (c)  $D_h = 10 \text{ meV}$ .

Nos gráficos da figura 53, observamos que as sequências com água apresentam os melhores



**Figura 53**

Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $E$  para AT e CG em diferentes energias  $D_h$ . Destacamos a região de intersecção entre os retângulos nas figuras. Esta região representa o intervalo de valores de  $E_{CG}$  e  $E_{AT}$  que reproduzem valores de profundidades de poço  $A$  similares ao modelo PB. Colorimos todos os pontos maiores que  $40 \text{ } ^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$ . (a)  $D_h = 0$  meV, (b)  $D_h = 1$  meV, (c)  $D_h = 10$  meV. O gráfico abaixo da figura (b) é da região entre a intersecção dos retângulos ampliada.



**Figura 54**

Variação do  $\langle \chi^2 \rangle$  em função da altura da barreira  $E$  para AT e CG em diferentes energias  $D_h$ . Destacamos a região de intersecção entre os retângulos nas figuras. Esta região representa o intervalo de valores de  $E_{CG}$  e  $E_{AT}$  que reproduzem valores de profundidades de poço  $A$  similares ao modelo PB. Colorimos todos os pontos maiores que  $30 \text{ } ^\circ C^2$  com a cor vermelha. As tonalidades de amarelo até vermelho se aplicam aos demais pontos. Os pontos com tonalidades em azul são dos menores valores  $\langle \chi^2 \rangle$ . (a)  $D_h = 0$  meV, (b)  $D_h = 1$  meV, (c)  $D_h = 10$  meV. O gráfico abaixo da figura (b) é da região entre a intersecção dos retângulos ampliada.

valores de  $\langle \chi^2 \rangle$  na região em que  $E_{CG} = 0$ . Já, nos gráficos com PEG, o melhor  $\langle \chi^2 \rangle$  está ao lado da intersecção dos retângulos ciano na figura 54. O primeiro ponto azul nos gráficos do conjunto DNA-W/40, com valores diferentes de zero para ambas as alturas de barreira, apresenta um valor de  $E_{AT} = 27.34 \text{ meV}/\text{Å}$ . Ao calcular a profundidade do poço  $A$ , encontramos um valor próximo de  $99.07 \text{ meV}$  para a profundidade do poço do par AT. Devido ao valor pequeno de  $E_{CG}$ , a profundidade do poço  $A_{CG}$  será bem menor que  $A_{AT}$ . Para permanecer dentro do regime próximo do modelo PB, usamos as figuras 53 e 54 apenas de referência. Os parâmetros de Morse similares aos do modelo PB são bons parâmetros para o modelo PCLJ? A intersecção dos retângulos em verde nas figuras com  $D_h = 1 \text{ meV}$  nos dá detalhes dos valores de  $\langle \chi^2 \rangle$ . Em ambos os casos, o  $\langle \chi^2 \rangle$  para iniciar às minimizações é bastante alto. Para DNA-W/40, o valor inicial está entre  $17.85^\circ C < \langle \chi^2 \rangle < 102.3^\circ C$  e para DNA-P200/40, o valor inicial está entre  $25.97^\circ C < \langle \chi^2 \rangle < 43.11^\circ C$ . Em outras palavras, as minimizações começam com valores de temperaturas preditas bem distantes das temperaturas experimentais. Entretanto, para manter os mesmos padrões adotados para os outros modelos, iniciaremos nossas análises a partir deste ponto. Não observamos nenhuma melhora (ou apenas uma melhora insignificante) nos valores de  $\langle \chi^2 \rangle$  ao variar o parâmetro  $D_h$ . Contudo, esse parâmetro possui um significado físico importante, relacionada à energia da fita simples após a separação do par de bases. A energia entre a fita simples e o solvente deve ter valores um pouco diferentes para AT e CG. Assim, partiremos do mesmo valor inicial de  $D_h = 1 \text{ meV}$  para AT e CG, mas o avaliaremos de forma distinta para cada par de bases. Os valores iniciais utilizados forma  $E_{(AT)} = 11 \text{ meV}/\text{Å}$ ,  $E_{(CG)} = 12 \text{ meV}/\text{Å}$  e  $D_{h(AT)} = D_{h(CG)} = 1 \text{ meV}$  para DNA-W/40 e  $E_{(AT)} = 11.2 \text{ meV}/\text{Å}$ ,  $E_{(CG)} = 13.4 \text{ meV}/\text{Å}$  e  $D_{h(AT)} = D_{h(CG)} = 1 \text{ meV}$  para DNA-P200/40. Um resumo com os parâmetros iniciais das minimizações do modelo PCLJ se encontra na tabela 30.

Parâmetro	DNA-W/40	DNA-P200/40
$k_{AT-AT}$	1.35 eV/nm <sup>2</sup>	3.03 eV/nm <sup>2</sup>
$k_{AT-CG}$	2.58 eV/nm <sup>2</sup>	1.80 eV/nm <sup>2</sup>
$k_{AT-GC}$	4.84 eV/nm <sup>2</sup>	4.65 eV/nm <sup>2</sup>
$k_{AT-TA}$	3.53 eV/nm <sup>2</sup>	1.22 eV/nm <sup>2</sup>
$k_{CG-AT}$	3.29 eV/nm <sup>2</sup>	4.78 eV/nm <sup>2</sup>
$k_{CG-CG}$	2.55 eV/nm <sup>2</sup>	2.30 eV/nm <sup>2</sup>
$k_{CG-GC}$	3.92 eV/nm <sup>2</sup>	4.35 eV/nm <sup>2</sup>
$k_{GC-AT}$	1.26 eV/nm <sup>2</sup>	2.55 eV/nm <sup>2</sup>
$k_{GC-CG}$	2.22 eV/nm <sup>2</sup>	1.91 eV/nm <sup>2</sup>
$k_{TA-AT}$	0.08 eV/nm <sup>2</sup>	0.05 eV/nm <sup>2</sup>
$E_{(AT)}$	11 meV/Å	11.2 meV/Å
$E_{(CG)}$	12 meV/Å	13.4 meV/Å
$\gamma_{AT}$	17.5 nm	17.5 nm
$\gamma_{CG}$	22.5 nm	22.5 nm
$D_h (AT) = D_h (CG)$	1 meV	1 meV
$\lambda_h (AT)$	$5.87 \times 10^{-3}$ nm	$5.87 \times 10^{-3}$ nm
$\lambda_h (CG)$	$7.18 \times 10^{-3}$ nm	$7.18 \times 10^{-3}$ nm

**Tabela 30**

Parâmetros iniciais utilizados para as minimizações do modelo PCLJ.

## **APÊNDICE C - Artigo Publicado**



Research paper

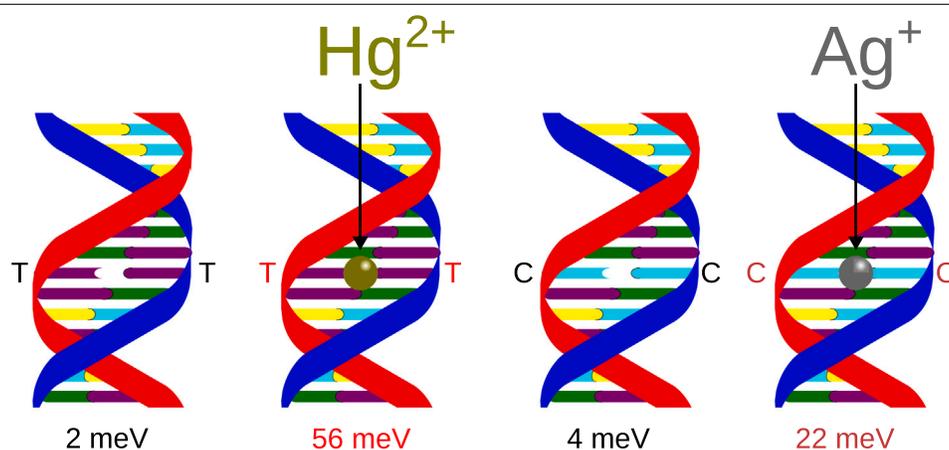


## Mesoscopic model confirms strong base pair metal mediated bonding for T–Hg<sup>2+</sup>–T and weaker for C–Ag<sup>+</sup>–C

Luciano Gabriel Silva, Gerald Weber\*

Departamento de Física, Universidade Federal de Minas Gerais, 31270-901, Belo Horizonte - MG, Brazil

### GRAPHICAL ABSTRACT



### HIGHLIGHTS

- Metal mediated bond potentials were obtained for T–Hg<sup>2+</sup>–T and C–Ag<sup>+</sup>–C.
- T–Hg<sup>2+</sup>–T bond potential is of similar strength as a hydrogen bond of a CG base pair.
- C–Ag<sup>+</sup>–C is much weaker supporting the hypothesis of a bond of type N3–Ag<sup>+</sup>–N3.
- Stacking interactions are increased only for GC–TT and CC–AT.

### ARTICLE INFO

#### Keywords:

Metal-mediated DNA  
Metallo-base-pair  
Peyrard-Bishop DNA model  
DNA stability  
Hydrogen bonds  
Stacking interaction

### ABSTRACT

Metal mediated DNA are mismatches where a metal ion bridges the base pairs and are of interest for biosensors. Here, we study T–Hg<sup>2+</sup>–T, which is known to be stabilized by base pair metal mediated bonds, and C–Ag<sup>+</sup>–C whose stabilization mechanism is less well understood. We use a mesoscopic model and published melting temperatures of sequences containing CC or TT mismatches in the presence of the metal ions. For T–Hg<sup>2+</sup>–T we obtain a strong base pair bond potential and moderate one for C–Ag<sup>+</sup>–C, only one stacking potential for each configuration is stronger than usual.

\* Corresponding author.

E-mail address: [gweber@ufmg.br](mailto:gweber@ufmg.br) (G. Weber).
<https://doi.org/10.1016/j.cplett.2022.139847>

Received 1 March 2022; Received in revised form 20 June 2022; Accepted 1 July 2022

Available online 5 July 2022

0009-2614/© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

Metal-mediated DNA, sometimes called M-DNA, where a metal ion occupies the hydrogen bond, typically in mismatched base pairs, has a long history, reaching back to the early 1950–60s [1–3]. Today, its continued interest lies in its potential to be used as a highly sensitive and specific metal sensor [4–7] as well as in the formation of nanowires [8–10]. The pioneering work by Ono and Miyake [11] revealed two M-DNA structures, CC mismatches mediated by  $\text{Ag}^+$ , and TT mismatches mediated by  $\text{Hg}^{2+}$ , which are still the main M-DNA structures studied. For T- $\text{Hg}^{2+}$ -T, it was found that it yields a higher melting temperature than a canonical AT base pair [11], and its structure obtained from NMR measurements is now considered well understood [12–14]. In particular it was shown that two imino protons are released when T- $\text{Hg}^{2+}$ -T is formed [12] and that C1'-C1' distances are shorter than those of Watson-Crick base pairs [13,15]. Miyake et al. [16] determined that in a T- $\text{Hg}^{2+}$ -T structure,  $\text{Hg}^{2+}$  binds directly to N3 of thymidine in place of the imino proton and bridges the thymidine residues. However, for C- $\text{Ag}^+$ -C a number of questions regarding its structure still remain. For instance, there appear to be conflicting results suggesting either a N3- $\text{Ag}^+$ -N3 [17], N3- $\text{Ag}^+$ -O2 [18], or a N3- $\text{Ag}^+$ -N3 linkage [19]. On the other hand, DFT calculations by Espinosa Leal et al. [20] suggests the existence of interplanar bonds which were also observed in crystal structures of  $\text{N}^1$ -hexylcytosine [19]. There are evidences of C- $\text{Ag}^+$ -C base pair formation in backbone-less  $[\text{Ag}(\text{cytidine})_2]^+$  complexes [21] which were shown to form a cisoid configuration in density-functional theory (DFT) calculations by Kliuev et al. [22], however it is unclear how these finding would translate to fully formed duplexes. Other DFT calculations explore the effect of  $\text{Zn}^{2+}$  [23], which are not covered here.

Mesoscopic models, such as the Peyrard-Bishop (PB) model, uses a set of effective potentials to describe the main interactions of the DNA double helix, namely the hydrogen bonds and the stacking interactions [24]. This type of model can be used to calculate melting temperatures with similar accuracy as nearest-neighbour models [25]. These models depend critically on adjusted parameters, and we showed that these parameters can be obtained from experimental melting temperatures [26]. An additional bonus of obtaining the parameters in this way is that they can be directly related to hydrogen bonds, or base pair metal mediated bonds, and the stacking interactions. In other words, combining the mesoscopic model with experimental melting temperatures provides a way to evaluate these interactions. The method has now been validated for several oligonucleotide systems, including a recent study for all possible combinations of mismatches for which we obtained overall good correspondence to the results from other experimental methods such as NMR [27]. The new mismatch parametrization now opens an opportunity to apply the method to other systems containing mismatches in more challenging configurations, notably T- $\text{Hg}^{2+}$ -T and C- $\text{Ag}^+$ -C. It is of particular interest to understand if existing melting temperatures of T- $\text{Hg}^{2+}$ -T and C- $\text{Ag}^+$ -C would result in increased base pair Morse potentials which is the central hypothesis of metal-ion stabilization in these mismatches [16]. It would be of especial interest to determine if existing melting temperatures could advance our understanding concerning the conflicting results related to base-pair metal mediated bonding of C- $\text{Ag}^+$ -C.

The major challenge in studying the metal mediated T- $\text{Hg}^{2+}$ -T and C- $\text{Ag}^+$ -C with mesoscopic models is the relative small number of melting temperatures that are scattered over a number of different studies. It is unclear if the mesoscopic model would be effective in providing a clear description of the base pair bond interactions for such a limited set of data. Furthermore, DNA mismatches are very sensitive to changes in buffer conditions and it is not clear if the existing parameters can be used as reliable starting point for the optimization of metal mediated potentials. On the other hand a successful optimization of T- $\text{Hg}^{2+}$ -T and C- $\text{Ag}^+$ -C would broaden the potential applications of DNA mismatch existing parameters and provide a critical validation of the method for a small melting temperature data set.

## 2. Materials and methods

### 2.1. Notation

We used the notation CC and TT for non-canonical DNA base pairs. The stacking interactions are represented by a (base pair 1)-(base pair 2) scheme. For example, the mismatch CC and the canonical pair AT, corresponding to the dimer 5'-CA-3'/3'-CT-5', is represented by CC-AT. Note that due to symmetry considerations, as 5'-CA-3'/3'-CT-5' is the same as 5'-TC-3'/3'-AC-5', therefore CC-AT is same as TA-CC.

The main goal is to observe the changes in base pair bond and stacking interactions, next to the mismatch due to the presence of a metal ion. To differentiate between the presence or absence of  $\text{Ag}^+$ , or  $\text{Hg}^{2+}$ , in solution we used the following notation: sequences without metal were called control, represented by CC-CTRL and TT-CTRL; for sequences with metal we used the notation C- $\text{Ag}^+$ -C and T- $\text{Hg}^{2+}$ -T that is commonly used in the field.

### 2.2. Melting temperature data set

For C- $\text{Ag}^+$ -C we used 22 sequences containing a concentration of 1 equivalent of  $\text{Ag}^+$  per mismatch CC from Refs. [11,28–38] and their corresponding control sequences (CC-CTRL) without  $\text{Ag}^+$  (21 melting temperatures). For T- $\text{Hg}^{2+}$ -T we used 34 sequences containing 1 equivalent of  $\text{Hg}^{2+}$  per mismatch TT from Refs. [13,16,28,29,31,32,38–44] and 24 control sequences (TT-CTRL) without  $\text{Hg}^{2+}$ .

The majority of melting temperatures in these sets are at a salt concentration,  $[\text{Na}^+]$ , of 100 mM and total strand concentration of 2  $\mu\text{M}$ . For different conditions we applied a salt correction factor, equation (22) from Owczarzy et al. [45], and recalculated the strand concentrations to 2  $\mu\text{M}$  [46] from the reported experimental total enthalpy and entropy variations. Supplementary tables S1 and S2 show the melting temperatures,  $T_{exp}$ , that were adjusted in this way. The remaining buffer conditions were in general very similar among the various datasets.

### 2.3. Mesoscopic models

The mesoscopic Peyrard-Bishop (PB) model describes the intramolecular interactions of DNA through simple potentials [24]. For example, stacking between the nearest-neighbours base pairs is described by a harmonic potential, while the base pair bond, usually a hydrogen bond, is represented by a Morse potential. For the  $i$ th base pair the Morse potential is

$$V(y_i) = D_\alpha (e^{-y_i/\lambda_\alpha} - 1)^2, \quad (1)$$

where the parameter  $D_\alpha$  tells us how strongly the  $\alpha$  base pair are bonded, whereas  $\lambda_\alpha$  is related to the potential width. For the nearest-neighbour stacking interaction we use a modified harmonic potential which includes a small twist angle  $\theta$  [26]

$$W(y_i, y_{i-1}) = \frac{k_\beta}{2} (y_i^2 - 2y_i y_{i-1} \cos \theta + y_{i-1}^2), \quad (2)$$

where the parameter  $k_\beta$  is the coupling constant of nearest-neighbours of type  $\beta$ , and the angle is fixed at  $\theta = 0.01$  rad, which is introduced to overcome well known numerical problems with the transfer-integral technique [26].

With the two potentials of Eqs. (1) and (2) we can calculate the classical partition function from which we obtain an adimensional index  $\tau_i(P)$ , which is proportional to the melting temperatures [26]. Additionally, we can calculate the average strand displacement  $\langle y_i \rangle$  for each base pair  $i$  by calculating the expected value of  $y$  from the partition function [26].

## 2.4. Calculation of melting temperatures

The index  $\tau_i(P)$  is used to predict melting temperatures for any sequence  $i$ , where  $P$  is a given set of parameters. Experimental melting temperatures and  $\tau_i(P)$  are linearly dependent within groups of sequences of same length  $N$ . This relation is shown in the following equation

$$T'_i(P) = a_0(N) + a_1(N)\tau_i(P), \quad (3)$$

where  $a_{0,1}(N)$  are the linear regression coefficients. These coefficients have a linear dependence with  $N^{1/2}$

$$a_k(N) = b_{0,k} + b_{1,k}N^{1/2}, \quad k = 0, 1. \quad (4)$$

## 2.5. Thermal equivalence optimization

The parameter optimization is carried out by minimizing the merit function

$$\chi_j^2(P_j) = \sum_{i=1}^N [T_i - T'_i(P_j)]^2, \quad (5)$$

where  $P_j$  is the  $j$ th tentative set of parameters. The goal is to find the best set of  $P_j$  parameters such that the calculated temperatures,  $T'_i(P_j)$ , are closest to the experimental temperatures,  $T_i$ . Considering that the set  $P$  contains  $M$  parameters, the numerical problem consists of finding the  $M$ -dimensional minima of Eq. (5), which is numerically implemented with a downhill simplex algorithm [25].

Another equation considered in our discussion to evaluate the quality of the prediction is the average melting temperature deviation

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T_i - T'_i(P_j)|. \quad (6)$$

## 2.6. Minimization procedure

One of the main challenges in the numerical minimization of Eq. (5) is the unavoidable occurrence of local minima, that is, local valleys that do not coincide with the best overall minima. To overcome this problem, we repeat the minimization process multiple times, restarting from different initial parameters. For each minimization round, the initial set of parameters  $P_{\text{init}}$  are randomly sampled around the 10% of value of seed parameters  $P_{\text{seed}}$ , that is, within the interval of  $0.9P_{\text{seed}}$  to  $1.1P_{\text{seed}}$ . Only Morse potentials  $D$  and stacking potentials  $k$  were varied. The  $\lambda$  parameters were kept constant since we have found, in previous studies, that their effect on melting temperatures is negligible [26]. In total, 6 parameters were varied for each type of metal mediated mismatch: one Morse potential factor  $D$  and 5 stacking factors  $k$ . The CTRL-TT, CTRL-CC, C-Ag<sup>+</sup>-C and T-Hg<sup>2+</sup>-T were treated separately.

**CTRL optimization.** For CTRL sequences, the initial seed parameters were those from Ref. [27] (supplementary table S3). One problem that occurred in this particular study was that some parameters that would repeatedly tend to zero. Due to the particularities of the minimization algorithm, once a parameter reaches zero during the minimization it will remain so during the whole minimization process. This risks artificially skewing the remaining parameters to suboptimal values. On the other hand, one cannot discard the occurrence of null parameter as DNA mismatches tend to have very small Morse potentials in most cases as well as very small stacking potentials for some specific situations. To ensure that null value parameters are not algorithmic artefacts, we identified the parameters that went to zero and artificially raised their values and repeated the whole process with a mix of non-zero optimized and artificially raised values. If after successive major rounds these same parameters would again tend to zero, it would then be an indication that indeed this would be their optimal value and not an artefact of the minimization algorithm. CTRL-TT and CTRL-CC were

optimized separately. The CTRL minimizations were performed 1000 times for each round, with different initial parameters. This procedure was repeated two more times, always raising null parameters to their initial values as needed. A fourth, and last, round of minimization was for estimating the influence of experimental uncertainty on our new parameters. This was done by changing the experimental temperatures by small random amounts such that the resulting standard deviation corresponds to the experimental uncertainty. We used 0.6 °C uncertainty for CTRL, which were obtained from the average reported uncertainty for each of the published measurements. Final quality parameters were  $\chi^2 = 58.56$  °C<sup>2</sup> and  $\langle \Delta T \rangle = 1.24$  °C for CTRL-CC;  $\chi^2 = 144.18$  °C<sup>2</sup> and  $\langle \Delta T \rangle = 1.79$  °C for CTRL-TT.

**C-Ag<sup>+</sup>-C and T-Hg<sup>2+</sup>-T optimization.** For the preliminary round, we joined the datasets, CTRL-TT with T-Hg<sup>2+</sup>-T, and CTRL-CC with C-Ag<sup>+</sup>-C, and varied the metal mediated parameters while keeping the control parameters fixed. We call this the preliminary optimization. This has the practical effect that all melting temperatures are forced to follow the same regression parameters Eqs. (3),(4), and causes the T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C parameters to assume values that bear some reference to the control parameters. This approach is used only to obtain seed parameters for T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C which are shown in supplementary table S4. After this, we perform three rounds of optimization using the T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C separately from the CTRL sets. For the last round, we proceed as for the last CTRL, assuming a 0.6 °C to evaluate the experimental uncertainty on our new parameters. Each round was performed in 1000 local minimization steps, results shown here are from the final minimization round. Final quality parameters were  $\chi^2 = 111.07$  °C<sup>2</sup> and  $\langle \Delta T \rangle = 1.86$  °C for C-Ag<sup>+</sup>-C;  $\chi^2 = 94.21$  °C<sup>2</sup> and  $\langle \Delta T \rangle = 1.35$  °C for T-Hg<sup>2+</sup>-T.

## 2.7. Availability

The metal mediated parameters calculated were included in the latest version of our free TfReg software [25] which can be used to verify our results. The software and the parameters are available at <http://tinyurl.com/tfregufmg>, for other download options see the supplementary information.

## 3. Results and discussion

The optimization rounds were carried out independently for T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C, and their respective controls. In total, four rounds with 1000 different initial conditions were carried out as described in the methods section. Quality parameters for final  $\langle \Delta T \rangle$  was 1.86 °C for C-Ag<sup>+</sup>-C and 1.35 for T-Hg<sup>2+</sup>-T which are of the same magnitude as found in previous calculations using this method in different situations, see for instance [47,48].

In Fig. 1 we show the optimized Morse potentials. For TT-CTRL the Morse potential is somewhat lower than its seed value, in the lower range of TT mismatches that were obtained previously [27]. The CC-CTRL Morse potential is also situated at the smaller values of CC, albeit marginally larger than the TT potential. However, both potentials are very small when compared to canonical base pairs AT and CG, therefore it is not possible to say that CC is more stable than TT from these results. With the addition of Hg<sup>2+</sup>, the Morse potential increases strongly, indicating a substantial base-pair metal mediated bonding, see Fig. 1. This is much higher than any known TT configuration [27] (shaded area) and roughly 4/5 of that of a CG Morse potential (70 meV) [26], which leaves very little doubt that T-Hg<sup>2+</sup>-T is stabilized by a Hg<sup>2+</sup> mediated bond. To understand this result, let us briefly discuss the T-Hg<sup>2+</sup>-T structure. Three configurations were suggested for T-Hg<sup>2+</sup>-T: N3-Hg<sup>2+</sup>-N3, O2-Hg<sup>2+</sup>-N3 and N3-Hg<sup>2+</sup>-O4 [49]. The N3-Hg<sup>2+</sup>-N3 has the oxygen atoms in opposition to each other, yet apart far enough such that the repulsive force between them may not be so strong, which could explain the large value of the Morse potential,

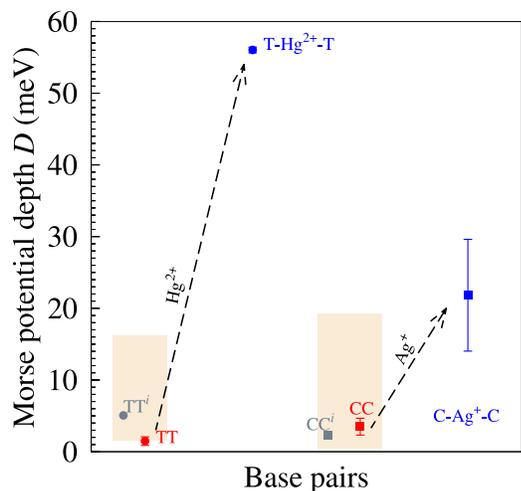


Fig. 1. Average Morse potential depths  $D$  calculated for T-Hg<sup>2+</sup>-T (blue bullet) C-Ag<sup>+</sup>-C (blue box) and their respective controls (red bullet and box). Error bars are the calculated standard deviation of the last minimization round. The seed values for mismatches, TT<sup>i</sup> and CC<sup>i</sup> (horizontally displaced for clarity), are shown as grey bullet and box, respectively. Dashed arrows indicate the changes due to the addition of cations and are only intended as guide for the eye. The shaded area shows the interval of minimum and maximum Morse potentials for all types of TT and CC mismatches from Ref. [27].

Fig. 1. However, the O2-Hg<sup>2+</sup>-N3 and the N3-Hg<sup>2+</sup>-O4 structures are possible as well. While our model is not able to identify the structure, it clearly points to the one that is likely to be the most strongly bonded. The addition of Ag<sup>+</sup> increases the CC Morse potential as well, Fig. 1. While this increase is much less than for T-Hg<sup>2+</sup>-T, it is nevertheless substantial, roughly 2/3 of that of an AT Morse potential (34 meV), which suggest a single link similar to the N3-Ag<sup>+</sup>-N3 link proposed by [17,19]. Such weaker Morse potential is likely due to the proximity of the repulsive amino groups facing each other at C-Ag<sup>+</sup>-C, resulting in large propeller twist angles [17]. Other types of links that were suggested, such as N3-Ag<sup>+</sup>-O2 linkage [18] would have an extra hydrogen bond which seems unlikely given the moderate Morse potential that was found.

From the Morse potential results, there is little doubt that T-Hg<sup>2+</sup>-T is stabilized by a metal-mediated planar base pair bond, while C-Ag<sup>+</sup>-C is less so. But what then explains the large increase in melting temperature when Ag<sup>+</sup> is added? The first sequences shown in supplementary table S1 provide a clue to what might be happening, these are 6 sequences of same length with a central CC and where only the nearest-neighbours change. For ACA/TCT central trimer, there is an increase of 10 °C while for CCG/GCC it is only 2 °C. This large difference indicates that nearest-neighbour interactions have an important role to play in the stability. The nearest-neighbour interaction is modelled here by elastic constants  $k$ , Eq. (2), which in general are associated to stacking interaction. Our results for the stacking constants  $k$  are shown in Fig. 2 for both T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C. By far, the highest stacking interaction is for CC-AT nearest neighbours, exactly the type present in ACA/TCT. Interestingly, the other nearest neighbours AT-CC of the ACA/TCT trimer has a very weak interaction. CC-CC, that is, for two consecutive CC mismatches the stacking interaction is almost non-existent, which means that here the increased stability is entirely driven by in-planar base pair bonds. Inter-planar Hydrogen bonds have been predicted from DFT calculations for backbone-free C-Ag<sup>+</sup>-C tetramers [20] and observed in crystal structures for N<sup>1</sup>-hexylcytosine [19], therefore one would have expected some degree of nearest-neighbour interaction in this case. Fig. 3 shows an average displacement profile containing a double CC mismatch. This type of profile, which is calculated as described in Ref. [25], considers the

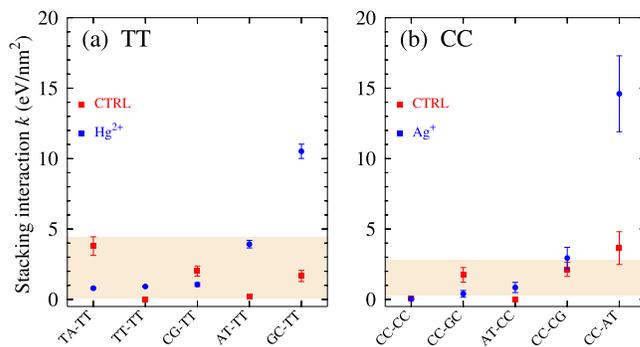


Fig. 2. Average stacking interaction  $k$  calculated for (a) T-Hg<sup>2+</sup>-T and (b) C-Ag<sup>+</sup>-C as blue bullets. Red boxes show the control (a) TT-CTRL and (b) CC-CTRL. Error bars are the calculated standard deviation of the last minimization round. The shaded area shows the interval of the minimum and maximum stacking interaction for equivalent mismatches from Ref. [27].

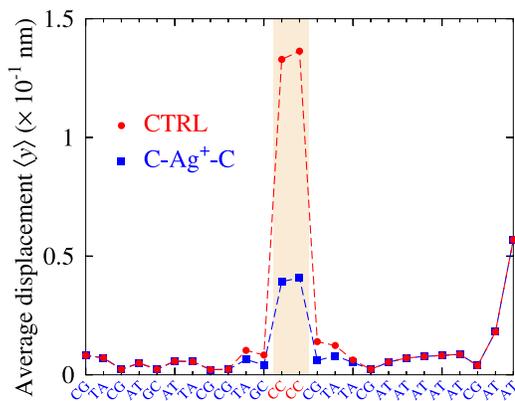


Fig. 3. Average displacement profile containing a double CC mismatch (shaded area), for the sequence from Ref. [50]. Red bullets are for CTRL parameters, and blue boxes for C-Ag<sup>+</sup>-C.

combined effects of Morse potentials and stacking, providing qualitative information regarding the regions where the base pairs are more likely to denature first. Typically, larger displacements means that the base pairs in question are less stable and will open before others. In the case of Fig. 3, the displacement profiles for C-Ag<sup>+</sup>-C (blue boxes) are much smaller, indicating increased base-pair stability as compared to the CTRL-CC (red bullets).

For T-Hg<sup>2+</sup>-T, the stacking interaction parameters shown in Fig. 2, only one strongly increased stacking for GC-TT was found, much higher than those previously seen for TT mismatches (shaded area) [27]. AT-TT is moderately increased in comparison to CTRL, while TA-TT is very small, even though its CTRL counterpart is relatively large. In Fig. 4a we show the average opening profiles for a sequence containing a double TT mismatch, flanked by AT-TT on both sides (note that AT-TT is equivalent to TT-TA). The CTRL displays very large displacements in the mismatch region (shaded area) while the addition of Hg<sup>2+</sup> causes a complete stabilization (very small displacements). By replacing AT-TT with TA-TT, as shown in Fig. 4b, we observe only a slight disturbance in the opening profiles due to the much smaller stacking, but not enough to cause major destabilization. Clearly, the T-Hg<sup>2+</sup>-T Morse potential is strong enough to compensate for the low TA-TT stacking.

The robustness of results presented here are encouraging and should enable us to apply the method to other configurations, such as for C-Hg<sup>2+</sup>-T mediated base pairs. A limited amount of melting temperatures are already available, see for instance Funai et al. [30], but not in sufficiently diverse sequence variations and lengths. Here, the smallest set was for C-Ag<sup>+</sup>-C with 22 sequences, which is in the lower

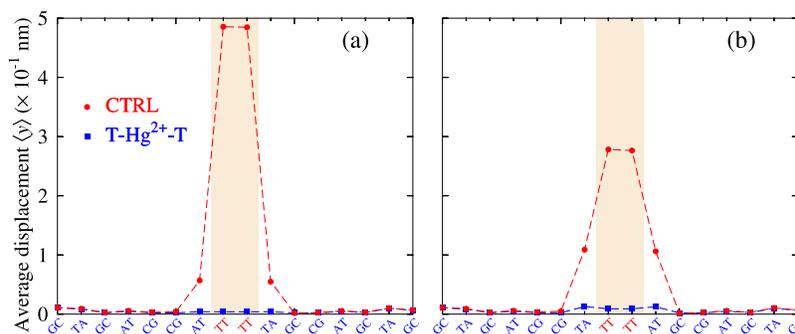


Fig. 4. Average displacement profile containing a double TT mismatch (shaded area) for the sequence from Ref. [16], panel (a). In panel (b) AT-TT was replaced with TA-TT. Red bullets are for CTRL parameters, and blue boxes for T-Hg<sup>2+</sup>-T.

limit of the method. Another aspect of interest is cation concentration, here we used the available temperatures measured at one equivalent cation per mismatch, and of course zero for the controls. It would be desirable to repeat this for a larger range of concentrations which should give us a more detailed insight of the stabilization process induced by the cations. We hope that our results may stimulate further melting temperature measurements of metal mediated base pairs.

#### 4. Conclusion

We analysed published melting temperatures for metal mediated base pairs T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C. For T-Hg<sup>2+</sup>-T we found a large Morse potential of the order of a CG base pair, while for C-Ag<sup>+</sup>-C this was much smaller, roughly 2/3 of a AT base pair. For C-Ag<sup>+</sup>-C, the weaker Morse potential favours the N3-Ag<sup>+</sup>-N3 link proposed by [17]. For a double CC mismatch the stacking interaction was found to be very weak in the presence of Ag<sup>+</sup>, but still the Morse potential was strong enough to provide some stability when compared to the absence of Ag<sup>+</sup>. Very low stacking was also found for TA-TT nearest-neighbours, however due to the large T-Hg<sup>2+</sup>-T Morse potential this does not affect the duplex stability.

Our results are generally in good agreement with the existing knowledge of T-Hg<sup>2+</sup>-T and C-Ag<sup>+</sup>-C mediated base pairs. Therefore, we believe that it should be possible to extend the method to other types of metal-mediated base pairs, for instance C-Hg<sup>2+</sup>-T, when a sufficient amount of melting temperatures become available.

#### CRedit authorship contribution statement

**Luciano Gabriel Silva:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Visualization. **Gerald Weber:** Conceptualization, Methodology, Supervision, Writing – original draft, Visualization, Project administration.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Links to the data can be found in the supplementary material.

#### Acknowledgements

Supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Brazil, Finance Code 001).

#### Appendix A. Supplementary data

Tables S1 and S2 show the sequences and temperatures used. Tables S3 and S4 show the seed parameters used.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.cplett.2022.139847>.

#### References

- [1] S. Katz, The reversible reaction of sodium thymonucleate and mercuric chloride, *J. Am. Chem. Soc.* 74 (9) (1952) 2238–2245, <http://dx.doi.org/10.1021/ja01129a023>.
- [2] T. Yamane, N. Davidson, On the complexing of deoxyribonucleic acid (DNA) by mercuric ion, *J. Am. Chem. Soc.* 83 (12) (1961) 2599–2607, <http://dx.doi.org/10.1021/ja01473a001>.
- [3] M. Daune, C. Dekker, H. Schachman, Complexes of silver ion with natural and synthetic polynucleotides, *Biopoly* 4 (1) (1966) 51–76, <http://dx.doi.org/10.1002/bip.1966.360040107>.
- [4] A. Ono, H. Togashi, Highly selective oligonucleotide-based sensor for mercury (II) in aqueous solutions, *Angew. Chem. Int. Ed. Engl.* 116 (33) (2004) 4400–4402, <http://dx.doi.org/10.1002/anie.200454172>.
- [5] W. Zhou, R. Saran, J. Liu, Metal sensing by DNA, *Chem. Rev.* 117 (12) (2017) 8272–8325, <http://dx.doi.org/10.1021/acs.chemrev.7b00063>.
- [6] V.G. Kanellis, C.G. Dos Remedios, A review of heavy metal cation binding to deoxyribonucleic acids for the creation of chemical sensors, *Biophys. Rev.* 10 (5) (2018) 1401–1414, <http://dx.doi.org/10.1007/s12551-018-0455-y>.
- [7] J. Müller, Nucleic acid duplexes with metal-mediated base pairs and their structures, *Coord. Chem. Rev.* 393 (2019) 37–47, <http://dx.doi.org/10.1016/j.ccr.2019.05.007>.
- [8] G.H. Clever, M. Shionoya, Metal-base pairing in DNA, *Coord. Chem. Rev.* 254 (19–20) (2010) 2391–2402, <http://dx.doi.org/10.1016/j.ccr.2010.04.014>.
- [9] J. Kondo, Y. Tada, T. Dairaku, Y. Hattori, H. Saneyoshi, A. Ono, Y. Tanaka, A metallo-DNA nanowire with uninterrupted one-dimensional silver array, *Nature Chem.* 9 (10) (2017) 956–960, <http://dx.doi.org/10.1038/NCHEM.2808>.
- [10] S. Vecchioni, M.C. Capece, E. Toomey, L. Nguyen, A. Ray, A. Greenberg, K. Fujishima, J. Urbina, I.G. Paulino-Lima, V. Pinheiro, J. Shih, G. Wessel, S.J. Wind, L. Rothschild, Construction and characterization of metal ion-containing DNA nanowires for synthetic biology and nanotechnology, *Sci. Rep.* 9 (1) <http://dx.doi.org/10.1038/s41598-019-43316-1>.
- [11] A. Ono, Y. Miyake, Highly selective binding of metal ions to thymine-thymine and cytosine-cytosine base pairs in DNA duplexes, *Nucleic Acids Res. Suppl.* 3 (2003) 227–228, <http://dx.doi.org/10.1093/nass/3.1.227>.
- [12] Y. Tanaka, S. Oda, H. Yamaguchi, Y. Kondo, C. Kojima, A. Ono, <sup>15</sup>N-<sup>15</sup>N J-coupling across Hg<sup>II</sup>: direct observation of Hg<sup>II</sup>-mediated T-T base pairs in a DNA duplex, *J. Am. Chem. Soc.* 129 (2) (2007) 244–245, <http://dx.doi.org/10.1021/ja065552h>.
- [13] H. Yamaguchi, J. Šbera, J. Kondo, S. Oda, T. Komuro, T. Kawamura, T. Dairaku, Y. Kondo, I. Okamoto, A. Ono, et al., The structure of metallo-DNA with consecutive thymine-Hg<sup>II</sup>-thymine base pairs explains positive entropy for the metallo base pair formation, *Nucleic Acids Res.* 42 (6) (2014) 4094–4099, <http://dx.doi.org/10.1093/nar/gkt1344>.
- [14] Y. Tanaka, J. Kondo, V. Sychrovský, J. Šbera, T. Dairaku, H. Saneyoshi, H. Urata, H. Torigoe, A. Ono, Structures, physicochemical properties, and applications of T-Hg II-T, C-Ag I-C, and other metallo-base-pairs, *Chem. Commun.* 51 (98) (2015) 17343–17360, <http://dx.doi.org/10.1039/C5CC02693H>.
- [15] J. Kondo, T. Yamada, C. Hirose, I. Okamoto, Y. Tanaka, A. Ono, Crystal structure of metallo DNA duplex containing consecutive Watson-Crick-like T-Hg<sup>II</sup>-T Base Pairs, *Angew. Chem. Int. Ed. Engl.* 53 (9) (2014) 2385–2388, <http://dx.doi.org/10.1002/anie.201309066>.

- [16] Y. Miyake, H. Togashi, M. Tashiro, H. Yamaguchi, S. Oda, M. Kudo, Y. Tanaka, Y. Kondo, R. Sawa, T. Fujimoto, et al., Mercury<sup>II</sup>-mediated formation of thymine-Hg<sup>II</sup>-thymine base pairs in DNA duplexes, *J. Am. Chem. Soc.* 128 (7) (2006) 2172–2173, <http://dx.doi.org/10.1021/ja056354d>.
- [17] T. Dairaku, K. Furuita, H. Sato, J. Šbera, K. Nakashima, J. Kondo, D. Yamanaka, Y. Kondo, I. Okamoto, A. Ono, et al., Structure determination of an Ag<sup>I</sup>-Mediated Cytosine–Cytosine Base Pair within DNA Duplex in Solution with <sup>1</sup>H/<sup>15</sup>N/<sup>109</sup>Ag NMR Spectroscopy, *Chem. — Eur. J.* 22 (37) (2016) 13028–13031, <http://dx.doi.org/10.1002/chem.201603048>.
- [18] Y. Wang, B. Ritzo, L.-Q. Gu, Silver (I) ions modulate the stability of DNA duplexes containing cytosine, methylcytosine and hydroxymethylcytosine at different salt concentrations, *RSC Adv.* 5 (4) (2015) 2655–2658, <http://dx.doi.org/10.1039/c4ra14490b>.
- [19] A. Terrón, B. Moreno-Vachiano, A. Bauzá, A. García-Raso, J.J. Fiol, M. Barceló-Oliver, E. Molins, A. Frontera, X-ray Crystal structure of a metalated double-helix generated by infinite and consecutive C<sup>+</sup>-Ag<sup>+</sup>-C<sup>+</sup> (C<sup>+</sup>:N<sup>1</sup>-Hexylcytosine) base pairs through argentophilic and hydrogen bond interactions, *Chem. — Eur. J.* 23 (9) (2017) 2103–2108, <http://dx.doi.org/10.1002/chem.201604331>.
- [20] L.A. Espinosa Leal, A. Karpenko, S. Swasey, E.G. Gwinn, V. Rojas-Cervellera, C. Rovira, O. Lopez-Acevedo, The role of hydrogen bonds in the stabilization of silver-mediated cytosine tetramers, *J. Phys. Chem. Lett.* 6 (20) (2015) 4061–4066, <http://dx.doi.org/10.1021/acs.jpclett.5b01864>.
- [21] T. Dairaku, R. Kawai, T. Kanaba, T. Ono, K. Yoshida, H. Sato, K. Nozawa-Kumada, Y. Kondo, J. Kondo, A. Ono, et al., Effect of cytosine–Ag<sup>+</sup>–cytosine base pairing on the redox potential of the Ag<sup>+</sup>/Ag couple and the chemical reduction of Ag<sup>+</sup> to Ag by tetrathiafulvalene, *Dalton Trans.* 50 (22) (2021) 7633–7639, <http://dx.doi.org/10.1039/d1dt00975c>.
- [22] P.N. Kluev, P.A. Sokolov, R.R. Ramazanov, QM/MM-MD dissociation of Ag<sup>+</sup> and H<sup>+</sup> mediated cytosine pairs: Monomers and dimers, *J. Organomet. Chem.* 919 (2020) 121333, <http://dx.doi.org/10.1016/j.jorganchem.2020.121333>.
- [23] S.S. Alexandre, B.J. Murta, J.M. Soler, F. Zamora, Stability and electronic structure of M-DNA: Role of metal position, *Phys. Rev. B* 84 (4) (2011) 045413, <http://dx.doi.org/10.1103/PhysRevB.84.045413>.
- [24] M. Peyrard, A.R. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, *Phys. Rev. Lett.* 62 (23) (1989) 2755–2757, <http://dx.doi.org/10.1103/PhysRevLett.62.2755>.
- [25] G. Weber, Tfreq: Calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models, *Bioinformatics* 29 (2013) 1345–1347, <http://dx.doi.org/10.1093/bioinformatics/btt133>, <http://bioinformatics.oxfordjournals.org/content/29/10/1345>.
- [26] G. Weber, J.W. Essex, C. Neylon, Probing the microscopic flexibility of DNA from melting temperatures, *Nat. Phys.* 5 (2009) 769–773, <http://dx.doi.org/10.1038/nphys1371>.
- [27] L.M. Oliveira, A.S. Long, T. Brown, K.R. Fox, G. Weber, Melting temperature measurement and mesoscopic evaluation of single double and triple DNA Mismatches, *Chem. Sci.* 11 (2020) 8273–8287, <http://dx.doi.org/10.1039/d0sc01700k>, <https://pubs.rsc.org/en/content/articlelanding/2020/SC/D0SC01700K>.
- [28] A. Ono, H. Torigoe, Y. Tanaka, I. Okamoto, Binding of metal ions by pyrimidine base pairs in DNA duplexes, *Chem. Soc. Rev.* 40 (12) (2011) 5855–5866, <http://dx.doi.org/10.1039/c1cs15149e>.
- [29] T. Funai, N. Adachi, M. Aotani, S. i. Wada, H. Urata, Effects of metal ions on thermal stabilities of DNA duplexes containing homo- and heterochiral mismatched base pairs: comparison of internal and terminal substitutions, *Nucleos. Nucleot. Nucl.* 39 (1–3) (2020a) 310–321, <http://dx.doi.org/10.1080/15257770.2019.1658116>.
- [30] T. Funai, M. Aotani, R. Kiriu, J. Nakamura, Y. Miyazaki, O. Nakagawa, S. i. Wada, H. Torigoe, A. Ono, H. Urata, Silver (I)-ion-mediated cytosine-containing base pairs: Metal ion specificity for duplex stabilization and susceptibility toward DNA Polymerases, *ChemBioChem* 21 (4) (2020b) 517–522, <http://dx.doi.org/10.1002/cbic.201900450>.
- [31] H. Urata, E. Yamaguchi, Y. Nakamura, S. i. Wada, Pyrimidine–pyrimidine base pairs stabilized by silver (I) ions, *Chem. Commun.* 47 (3) (2011) 941–943, <http://dx.doi.org/10.1039/C0CC04091F>.
- [32] O. Nakagawa, A. Fujii, Y. Kishimoto, Y. Nakatsuji, N. Nozaki, S. Obika, 2-O, 4-C-methylene-bridged nucleic acids stabilize metal-mediated base pairing in a DNA duplex, *ChemBioChem* 19 (22) (2018) 2372–2379, <http://dx.doi.org/10.1002/cbic.201800448>.
- [33] A. Ono, S. Cao, H. Togashi, M. Tashiro, T. Fujimoto, T. Machinami, S. Oda, Y. Miyake, I. Okamoto, Y. Tanaka, Specific interactions between silver (I) ions and cytosine–cytosine pairs in DNA duplexes, *Chem. Commun.* (39) (2008) 4825–4827, <http://dx.doi.org/10.1039/B808686A>.
- [34] N. Zimmermann, E. Meggers, P.G. Schultz, A novel silver (I)-mediated DNA base pair, *J. Am. Chem. Soc.* 124 (46) (2002) 13684–13685, <http://dx.doi.org/10.1021/ja0279951>.
- [35] I. Schönrrath, V.B. Tsvetkov, T.S. Zatssepin, A.V. Aralov, J. Müller, Silver (I)-mediated base pairing in parallel-stranded DNA involving the luminescent cytosine analog 1, 3-diaza-2-oxophenoxazine, *J. Biol. Inorg. Chem.* 24 (5) (2019) 693–702, <http://dx.doi.org/10.1007/s00775-019-01682-1>.
- [36] H. Torigoe, Y. Miyakawa, A. Ono, T. Kozasa, Thermodynamic properties of the specific binding between Ag<sup>+</sup> ions and C: C mismatched base pairs in duplex DNA, *Nucleos. Nucleot. Nucl.* 30 (2) (2011a) 149–167, <http://dx.doi.org/10.1080/15257770.2011.553210>.
- [37] S. Naskar, M. Hebenbrock, J. Müller, Light-induced formation of silver (I)-mediated base pairs in DNA: Possibilities and limitations, *Inorg. Chim. Acta* 512 (2020) 119856, <http://dx.doi.org/10.1016/j.ica.2020.119856>.
- [38] A. Fujii, O. Nakagawa, Y. Kishimoto, T. Okuda, Y. Nakatsuji, N. Nozaki, Y. Kasahara, S. Obika, A. 1, 3, 9-Triaza-2-oxophenoxazine: An artificial nucleobase forming highly stable self-base pairs with three Ag<sup>I</sup> ions in a duplex, *Chem. — Eur. J.* 25 (31) (2019) 7443–7448, <http://dx.doi.org/10.1002/chem.201900373>.
- [39] H. Torigoe, A. Ono, T. Kozasa, Hg<sup>II</sup> ion specifically binds with T:T mismatched base pair in duplex DNA, *Chem. — Eur. J.* 16 (44) (2010) 13218–13225, <http://dx.doi.org/10.1002/chem.201001171>.
- [40] O.P. Schmidt, S. Jurt, S. Johansson, A. Karimi, R.K. Sigel, N.W. Luedtke, Concerted dynamics of metallo-base pairs in an A/B-form helical transition, *Nuovo Cimento* 10 (2019) 4818, <http://dx.doi.org/10.1038/s41467-019-12440-x>.
- [41] T. Funai, C. Tagawa, O. Nakagawa, S. i. Wada, A. Ono, H. Urata, Enzymatic formation of consecutive thymine–Hg<sup>II</sup>–thymine base pairs by DNA polymerases, *Chem. Commun.* 56 (2020c) 12025, <http://dx.doi.org/10.1039/d0cc04423g>.
- [42] G. Mata, O.P. Schmidt, N.W. Luedtke, A fluorescent surrogate of thymidine in duplex DNA, *Chem. Commun.* 52 (25) (2016) 4718–4721, <http://dx.doi.org/10.1039/c5cc09552b>.
- [43] O.P. Schmidt, A.S. Benz, G. Mata, N.W. Luedtke, Hg<sup>II</sup> binds to C–T mismatches with high affinity, *Nucleic Acids Res.* 46 (13) (2018) 6470–6479, <http://dx.doi.org/10.1093/nar/gky499>.
- [44] J.H. Han, S. Hirashima, S. Park, H. Sugiyama, Highly sensitive and selective mercury sensor based on mismatched base pairing with dioxT, *Chem. Commun.* 55 (69) (2019) 10245–10248, <http://dx.doi.org/10.1039/c9cc05123f>.
- [45] R. Owczarzy, Y. You, B.G. Moreira, J.A. Manthey, L. Huang, M.A. Behlke, J.A. Walder, Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures, *Biochem* 43 (2004) 3537–3554, <http://dx.doi.org/10.1021/bi034621r>.
- [46] S. Schreiber-Gosche, R.A. Edwards, Thermodynamics of oligonucleotide duplex melting, *J. Chem. Educ.* 86 (5) (2009) 644, <http://dx.doi.org/10.1021/ed086p644>.
- [47] E.d.O. Martins, V.B. Barbosa, G. Weber, DNA/RNA Hybrid mesoscopic model shows strong stability dependence with deoxypyrimidine content and stacking interactions similar to RNA/rna, *Chem. Phys. Lett.* 715C (2019) 14–19, <http://dx.doi.org/10.1016/j.cplett.2018.11.015>.
- [48] M.I. Muniz, H.H. Lackey, J.M. Heemstra, G. Weber, DNA/TNA Mesoscopic modeling of melting temperatures suggest weaker hydrogen bonding of CG than in DNA/rna, *Chem. Phys. Lett.* (2020) 137413, <http://dx.doi.org/10.1016/j.cplett.2020.137413>, <https://www.sciencedirect.com/science/article/abs/pii/S0009261420303286>.
- [49] H. Liu, C. Cai, P. Haruehanroengra, Q. Yao, Y. Chen, C. Yang, Q. Luo, B. Wu, J. Li, J. Ma, et al., Flexibility and stabilization of Hg<sup>II</sup>-mediated C:T and T:T base pairs in DNA duplex, *Nucleic Acids Res.* 45 (5) (2017) 2910–2918, <http://dx.doi.org/10.1093/nar/gkx1296>.
- [50] H. Torigoe, Y. Miyakawa, A. Ono, T. Kozasa, Thermodynamic properties of the specific binding between Ag<sup>+</sup> ions and C:C mismatched base pairs in duplex DNA, *Nucleos. Nucleot. Nucl.* 30 (2) (2011b) 149–167, <http://dx.doi.org/10.1080/15257770.2011.553210>.