

UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE FÍSICA
PÓS-GRADUAÇÃO

Erik de Oliveira Martins

*Desenvolvimento de modelos mesoscópicos assimétricos para
oligonucleotídeos e sua aplicação a defeitos tipo single-bulges em
RNAs.*

Belo Horizonte

2018

UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE FÍSICA
PÓS-GRADUAÇÃO

Erik de Oliveira Martins

*Desenvolvimento de modelos mesoscópicos assimétricos para
oligonucleotídeos e sua aplicação a defeitos tipo single-bulges em
RNAs.*

Apresentada aqui a Tese desenvolvida no programa de Pós-Graduação em Física da Universidade Federal de Minas Gerais, buscando a obtenção do grau de Doutor em Física.

Orientador: Gerald Weber

Belo Horizonte

2018

Martins, Erik

Desenvolvimento de modelos mesoscópicos assimétricos para oligonucleotídeos e sua aplicação a defeitos tipo single-bulges em RNAs.

/ Erik de Oliveira Martins - 2018

110.p

1.Física Computacional 2.Física Estatística 3.Métodos Matemáticos 4.Bioinformática 5.Biofísica.. III.Título.

CDU 335

Erik de Oliveira Martins

*Desenvolvimento de modelos mesoscópicos assimétricos para
oligonucleotídeos e sua aplicação a defeitos tipo single-bulges em
RNAs.*

Tese apresentada ao Colegiado de Pós-Graduação da
UFMG, como requisito para a obtenção do título de
Doutor em Física.

BANCA EXAMINADORA

Gerald Weber

Doutor em Física - UFMG

Ubirajara Agero Batista

Doutor em Física - UFMG

Simone Silva Alexandre

Doutora em Física - UFMG

Jorge Chahine

Doutor em Física - UNESP

Rita Maria Cunha de Almeida

Doutora em Física - UFRGS

Aos meus pais, pelo apoio e companheirismo.

Sumário

1	Introdução	11
2	As moléculas de DNA e RNA	13
2.1	Estrutura do DNA	14
2.2	Funções biológicas e RNA	16
2.3	Propriedades físicas da molécula de RNA	18
2.4	Experimento de desnaturação térmica	22
3	Modelos de desnaturação de DNA e RNA	31
3.1	O modelo de primeiros vizinhos	31
3.2	O modelo Peyrard-Bishop	34
3.3	O modelo PB para sequências heterogêneas	38
3.3.1	Equivalência térmica	39
3.3.2	Parametrização do modelo PB	41
3.4	Conclusão	42
4	Empilhamento assimétrico no modelo PB	43
4.1	Introdução	43
4.2	Stacking harmônico	45
4.2.1	Verificação numérica	47
4.3	Stacking anarmônico	52
4.3.1	Aplicação blocos de funções de pulso (2D-BPFs) a núcleos de 4 variáveis genéricos	55

4.3.2	Aplicação das 2D-BPFs à núcleos similares ao modelo Peyrard-Bishop original	59
4.4	Conclusões	61
5	Modelo mesoscópico assimétrico para o estudo de <i>single-bulges</i> em RNAs	63
5.1	Introdução	63
5.2	Bulges em RNAs	64
5.3	Modelo PB no estudo de <i>bulges</i>	70
5.4	Conjunto de sequências e suas respectivas temperaturas de melting	74
5.5	Otimização dos parâmetros	74
5.6	Resultados e discussões	78
5.7	Conclusões	85
6	Conclusões e perspectivas futuras	87
A	Artigo publicado	89

Resumo

O entendimento de estruturas de RNAs não canônicas como despareamentos, protuberâncias, bases triplas, laços internos, hairpins e pseudo-nós, vem crescendo devido a seus papéis biológicos em que são submetidas e as amplitudes de movimento realizadas. Em particular, estruturas contendo protuberâncias (*bulges*) tem sido importantes em interações RNA-RNA, RNA ligantes e também são sítios específicos para as ligações de proteínas. Diversos modelos físicos estão sendo utilizados nesse tipo de estudo. O modelo termodinâmico mais utilizado no estudo de estruturas de RNAs não canônicas desse tipo é o modelo de primeiros vizinhos (NN). Entretanto, apesar da quantidade de trabalhos publicados utilizando o modelo NN, o modelo não consegue fazer uma ligação física das energias termodinâmicas calculadas com as propriedades estruturais que essas moléculas assumem. É necessário a procura de modelos alternativos para um melhor entendimento de duplexos com esses defeitos.

Modelos mesoscópicos tipo Peyrard-Bishop (PB) ainda não foram utilizados no estudo das protuberâncias. O modelo PB consegue descrever de forma intuitiva as interações contidas em duplexos. Ao longo dos últimos anos, variações do modelo PB tem sido utilizados em diversas aplicações físicas e alcançando resultados importantes no estudo das propriedades termodinâmicas de duplexos canônicos. Entretanto, o mesmo não foi aplicado a estruturas assimétricas antes. Neste trabalho, mostramos que é possível o uso do modelo Peyrard-Bishop no estudo de estruturas assimétricas de DNAs e RNAs com protuberâncias de uma base (*single-bulges*). Na primeira parte do trabalho, adaptamos a Hamiltoniana do modelo PB de forma que ela descreva o empilhamento (*stacking*) para diferentes monômeros. Através de um tratamento algébrico adequado, mostramos que o modelo Peyrard-Bishop com empilhamento harmônico descreve estruturas assimétricas com uma constante elástica efetiva, de natureza similar a de um oscilador harmônico com molas em série. A função de partição para esse sistema também pode ser resolvida numericamente pela técnica de integral de transferência (IT). Buscamos o mesmo tratamento para o potencial com empilhamento anarmônico. Nesse caso, chegamos a uma equação de Fredholm de segunda ordem, que torna o método IT impraticável. Para a

solução da equação de Fredholm de segunda ordem, sugerimos o uso de blocos de funções de pulso (2D-BPFs) e as aplicamos ao modelo Peyrard-Bishop em um polímero de DNA contendo somente pares de base CG. Em seguida, usamos com sucesso o resultado da assimetria com potenciais harmônicos para o estudo de *single-bulges* do Grupo I em moléculas de RNA. Para isso, criamos uma pseudo-base para denotar a posição do *bulge* no duplexo na fita oposta e dividimos as interações locais próximas ao bulge em subgrupos. Os parâmetros do modelo PB pertencentes a cada subgrupo foram calculados através de uma minimização entre as temperaturas preditas e experimentais. Para uma melhor compreensão dos resultados, estudamos os perfis de aberturas de sequências experimentais que mostram uma forte perturbação para *bulges guanosinas*, não relatada em estudos anteriores. Os cálculos feitos para as sequências experimentais contendo *single-bulges adenosinas* e *uridinas* estão em acordo com medidas de ressonância magnética nuclear (NMR). Como no caso experimental, notamos uma certa preferência de *bulges uridinas* possuírem a configuração para fora da hélice (*flipped out*). Os resultados encontrados aqui no estudo de *bulges* demonstram a correlação entre propriedades termodinâmicas e propriedades estruturais físicas.

Abstract

There is a growing understanding of the biological role of non-canonical RNAs such as mismatches, bulges, triple bases, internal loops hairpins and pseudo-knots, and the nonlinear behavior associated with these processes. In particular, structures containing bulges are important for RNA-RNA interactions, RNA ligands and specific sites of protein bondings. Several physical models are being used for this kind of study. The most frequently used thermodynamic model for non-canonical RNAs of this kind is the nearest-neighbor model (NN model). However, despite the large number of studies using this model they do not obtain a physical relationship between the thermodynamic energies and the structural properties that these molecules assume. Therefore, it is necessary to find alternative models for a better understanding these kind of defects.

The mesoscopic models such as the Peyrard-Bishop (PB) model have not been used in the study of bulges. The PB model can describe in an intuitive way the interactions contained in these duplexes. In the last few years, several variants of this model were used in different physical applications and achieved important results with respect to the thermodynamic properties of canonical duplexes. However, the PB model has not been applied to asymmetric structures before. In this work, we show that it is possible to use the Peyrard-Bishop model in asymmetric structures of DNA and RNA containing single-bulges. In the first part of this work, we adapted the Hamiltonian of the Peyrard-Bishop model to describe the stacking of different monomers. We show through the algebraic treatment that the PB model with harmonic stacking describes asymmetrical structures with an effective elastic constant, similar to harmonic oscillators in series. The partition function of this system can be solved numerically by the transfer integral (TI) method, as in the original model. We tried the same treatment with the Hamiltonian with an anharmonic stacking. In this case however, we obtained a Fredholm equation of the second type, that makes it impossible to use the TI technique. For the solution of the Fredholm equation of the second type, we suggest the use of two-dimensional block pulse functions (2D-BPFs) and we apply these functions to the original model for a polymer of the DNA

containing only CG base pairs. After that, we apply successfully the asymmetry results to the study of Group I single-bulges in RNA molecules. For this purpose, we created a pseudo-base to describe the absence of the base in the opposite direction of the bulge and divide the local interactions into subgroups of parameters. The parameters of each subgroup were calculated through the minimization between the predicted and experimental temperatures. For a better understanding of the results, we study the opening profiles of the experimental sequences and we found a strong perturbation of the guanosine bulges in the sequences that were not related before. The calculations for the experimental sequences containing adenosine and uridine bulges are in agreement with the nuclear magnetic resonance (NMR) measurements. As in the experiment, we note a preference of the uridine bulges for a flipped out configuration. The results found here in the study of bulges showed the correlation of thermodynamic properties with the structural properties of these molecules.

Agradecimentos

Gostaria de agradecer aos meus pais, Rubens e Norma, pelo apoio incondicional em toda minha vida e em especial, nesses 10 anos de estudo. O incentivo, carinho, dedicação, companherismo e amor ao longo desses anos foram fundamentais para que eu chegasse aqui hoje e realizasse minha defesa de doutorado. Foram muitas as dificuldades ao longo desses anos e vocês sempre fizeram o melhor para que eu alcançasse meus objetivos. Sem a ajuda de vocês, eu não teria conseguido. Muito obrigado por tudo e espero ter retribuído o que fizeram por mim. Amo vocês demais!

Gostaria de agradecer também a minha namorada, companheira e meu amor, Bruna. Agradeço o apoio, carinho, companherismo, dedicação e principalmente, por ter trazido mais felicidade a minha vida com nosso namoro. Você mudou completamente a minha maneira de enxergar o amor. Você me completa!

Gostaria de agradecer ao meu orientador e amigo, professor Gerald Weber, pela sincera orientação ao longo desses 4 anos. Agradeço pela paciência, discussões científicas e ensinamentos. Aprendi bastante com você nesse tempo. Espero que nossa colaboração e essa maravilhosa amizade continue após o término desse ciclo.

Gostaria de agradecer aos meus amigos de João Monlevade, em especial ao Júlio e Ivair. Vocês sabem o tanto que estudei nesses últimos anos e me apoiaram em todas as horas. Muito obrigado!

Gostaria de agradecer também ao meu aprendiz Juvenil Siqueira (New). Foram anos de bastante conversas, sarcasmos e bullying. Tenho certeza que você aprendeu bastante comigo nesses últimos anos, ainda que algumas vezes eu ache que estou perdendo meu tempo te ensinando sobre a vida hahahahaha... ;-)

Gostaria de agradecer também aos meus amigos de pós-graduação da UFMG, em especial ao Alberto (Timão), Elton (Pumba), Filomeno (Filozão da massa), Fábio, Leandra, Ronaldo (olha o golfinho!), Leonardo (Léo), Mateus (pior discípulo e amigo do Nene), Gustavo

(Gu) e Zé Roberto (Zezinho). Obrigado pela amizade nesses anos de mestrado e doutorado, pelas discussões científicas e conversas fiadas.

Gostaria de agradecer aos meus amigos de Ouro Preto, em especial ao Thales e Tião. Mesmo longe, a amizade persistiu nesses anos após me graduar.

Gostaria de agradecer ao meu amigo Frank Longford. Obrigado pela amizade e ajuda em minha breve passagem por Southampton. Foram boas conversas e tive uma grande experiência ao conhecer a Inglaterra.

Um agradecimento também a todos os professores da Universidade Federal de Minas Gerais (UFMG), pelos ensinamentos tanto acadêmicos quanto pessoais. Agradeço também aos funcionários da UFMG pela ajuda e orientação durante esses 6 anos de pós-graduação (mestrado e doutorado).

Não poderia esquecer também dos professores da Universidade Federal de Ouro Preto (UFOP), em especial, ao professor e orientador de graduação, Romuel Figueiredo Machado. Obrigado pelos ensinamentos em minha graduação.

Por fim, gostaria de agradecer as agências de fomento CNPq, CAPES e FAPEMIG pela ajuda financeira ao longo desses anos de estudo.

“The significance of our lives and our fragile planet is then determined only by our own wisdom and courage. We are the custodians of life’s meaning. We long for a Parent to care for us, to forgive us our errors, to save us from our childish mistakes. But knowledge is preferable to ignorance. Better by far to embrace the hard truth than a reassuring fable. If we crave some cosmic purpose, then let us find ourselves a worthy goal.”

Carl Sagan

1 Introdução

Desejamos estudar condições para as quais as assimetrias encontradas em RNAs possam ser consideradas utilizando-se um modelo mesoscópico. Em particular, nosso objetivo é o entendimento das propriedades físicas de dupla hélices contendo protuberâncias de uma base (*single-bulges*). Esperamos com o estudo aqui realizado poder utilizar o modelo Peyrard-Bishop para a predição de estruturas terciárias futuramente. Para isso, precisamos fazer uma relação entre as propriedades físicas termodinâmicas e estruturais dessas estruturas de RNAs contendo essas protuberâncias. Tomamos um conjunto experimental de oligonucleotídeos contendo *single-bulges* para as quatro bases nitrogenadas em RNAs junto a suas respectivas temperaturas de melting [1–4]. Usaremos as temperaturas experimentais na parametrização do modelo Peyrard-Bishop. Após o processo de minimização na determinação dos parâmetros, estudaremos os perfis de abertura dado pela medida de comprimento médio das interações de hidrogênio do modelo e o compararemos com os resultados experimentais. Mostraremos o efeito causado pelos *bulges* nos duplexos de RNA e nos pares de base próximos a eles.

Começamos a tese discutindo um pouco sobre as estruturas dos duplexos de DNAs e RNAs e suas respectivas interações associadas. Mostramos a orientação das bases ao longo das moléculas de RNAs e as conformações que elas podem assumir no caso canônico e não canônico. Ainda no capítulo 2, falaremos das técnicas experimentais em que essas hélices são submetidas, dando ênfase ao experimento de desnaturação térmica e o método de Van Hoff na determinação das temperaturas de melting.

No capítulo 3, apresentamos os dois modelos termodinâmicos mais utilizados nos estudos de duplexos, o modelo de primeiros vizinhos e o modelo Peyrard-Bishop. Apresentamos como utilizar esses modelos no estudo de duplexos e as técnicas de minimizações associadas a esses modelos. Discutiremos os avanços feitos pelo modelo Peyrard-Bishop no cálculo de duplas hélices e como esse modelo se mostra uma ferramenta mais completa que o modelo de primeiros vizinhos.

No capítulo 4, modificamos a Hamiltoniana do modelo Peyrard-Bishop para explorar

a assimetria de duplexos. No caso de um potencial harmônico, conseguimos uma relação simples entre as constantes elásticas dos monômeros que constituem a molécula sendo ainda possível o emprego da técnica de integral de transferência (IT) apresentada no capítulo 3. Ao utilizarmos o potencial de Dauxois *et al.* [5] no estudo de assimetria na Hamiltoniana do modelo, vemos que a técnica numérica IT é impraticável na solução da função de partição. O problema se resume na solução de uma equação de Fredholm de segunda ordem, onde sugerimos o uso de blocos de funções de pulso (2D-BPFs) para tal tarefa. Apresentamos as propriedades das 2D-BPFs e a aplicamos ao modelo original de Peyrard-Bishop [6].

No capítulo 5, aplicamos o hamiltoniano assimétrico com potenciais harmônicos no estudo de *single-bulges* em RNAs. Definimos uma pseudo-base na fita oposta permitindo a localização do *bulge* e também como recurso para o uso do modelo Peyrard-Bishop. Separamos as interações contidas nas sequências em três subgrupos que serão definidos ao longo do capítulo 5. Cada subgrupo tem sua importância no cálculo dos parâmetros do modelo e no estudo das propriedades estruturais. Em seguida, minimizamos os parâmetros do modelo e discutimos os resultados obtidos. Para uma análise completa, analisamos os perfis de aberturas de algumas sequências do conjunto e as comparamos com os experimentos de ressonância magnética nuclear (NMR). Os resultados apresentados no capítulo são consistentes com os experimentos de NMR para as sequências estudadas e fica evidente uma relação entre as propriedades termodinâmicas dos duplexos contendo *bulges* com as conformações que essas moléculas podem assumir.

Por fim, no capítulo 6, fazemos as considerações finais sobre toda a tese aqui descrita, como também as perspectivas futuras com relação ao método aqui apresentado e aplicações subsequentes.

2 As moléculas de DNA e RNA

Há duas grandes classes de moléculas biológicas: proteínas e ácidos nucleicos. Proteínas realizam a maioria das funções nos organismos vivos. Elas são máquinas ativas que catalisam reações químicas, transportam outras moléculas, como por exemplo, o oxigênio transportado pela hemoglobina, fazem controle dos canais de íons que cruzam a membrana plasmática, se juntam na formação de estruturas rígidas como o citoesqueleto, atuam como motores moleculares responsáveis pelas contrações dos músculos e muitas outras tarefas [7].

O mapa de construir estas proteínas é dado pelo código genético armazenado na estrutura do DNA (*deoxyribose nucleic acids*), portanto, o DNA é o responsável por formar o genoma de todos os seres vivos. Devido à sua complexidade, vários grupos de pesquisa nos Estados Unidos e na Europa no final da década de 1940 e no início dos anos 1950, buscavam entender como os átomos da molécula de DNA são mantidos unidos por ligações covalentes e como as moléculas resultantes estão organizadas no espaço tridimensional. Isso despertou interesse não somente de biólogos, mas também de físicos, químicos e matemáticos. Não é de surpreender que existissem ideias de que o DNA apresentasse uma estrutura muito complicada que diferia radicalmente de um gene para o outro [7].

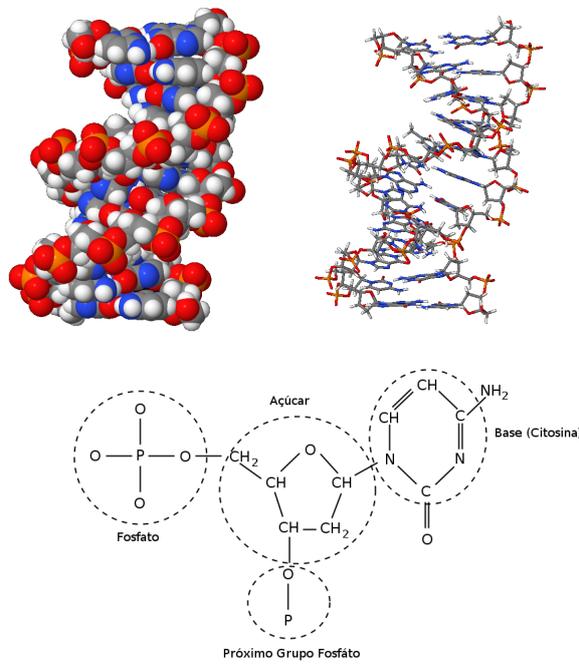
A procura pela estrutura da molécula, despertou a atenção de grandes cientistas da época, como por exemplo, Linus Pauling [8]. Foi então que em 1953 Watson e Crick propuseram o modelo correto para a dupla hélice do DNA baseados nos trabalhos do bioquímico austríaco Erwin Chargaff, do matemático John Griffiths e dos físicos Maurice H. F. Wilkins e Rosalind Franklin [8–10]. Essa estrutura demonstrou que todos os genes apresentam basicamente a mesma forma tridimensional e que as diferenças entre dois genes encontram-se na ordem e no número de seus quatro nucleotídeos construtores ao longo das fitas complementares. Isso levou a premiação do nobel de fisiologia para James D. Watson, Francis Crick e Maurice H. F. Wilkins em 1962.

Sabemos hoje que a descrição estática de moléculas biológicas (incluindo o DNA) não é suficiente para explicar as suas funções. Isso é devido às grandes mudanças conforma-

cionais que as moléculas biológicas estão sujeitas. No caso do DNA isso acontece durante os processos de transcrição e replicação. O entendimento dessas mudanças durante os processos biológicos é a motivação para o estudo da dinâmica da molécula de DNA. Devido as grandes amplitudes de movimento que envolvem a molécula, a dinâmica do DNA apresenta um caráter não linear que não pode ser ignorado. Os efeitos de flutuações térmicas locais desempenham um papel fundamental na funcionalização do DNA. Portanto, é imprescindível o uso da mecânica estatística no estudo de moléculas de DNAs.

2.1 Estrutura do DNA

A molécula de DNA é um polímero, ou mais precisamente, um conjunto de polímeros entrelaçados. A topologia básica dos grupos e interações que compõem a molécula, levam a conclusão que a estrutura do DNA possui a forma helicoidal. Entretanto, essas considerações não são suficientes para determinar todos os detalhes da estrutura, e em particular, se a molécula é canhota ou destra. O fato é que as duas situações podem ser encontradas na natureza e dentro dessas duas classes, podemos encontrar diferentes formas dependendo do número de grupos fosfatos ao longo de suas estruturas [11]. Focaremos aqui nas formas de estruturas A e B, mais comuns para duplexos de DNAs e RNAs, respectivamente. A forma de estrutura A contém em média cerca de 11 a 12 grupos fosfatos por volta tomando-se o maior e o menor sulco ao longo da hélice do duplexo. A forma de estrutura B de duplexos contém cerca de 10 grupos fosfatos por volta. Notamos que a forma A de duplexos possui uma estrutura mais compacta que a forma B. É sabido que a molécula de DNA em solução pode ser encontrada nas formas A e B. Ambas levam a uma hélice destra. Em particular, algumas estruturas de DNAs podem mudar suas conformações por desidratação [12], mas o mesmo não acontece para estruturas de RNAs que apesar de conterem uma estrutura muito semelhante aos DNAs, conseguem manter a sua configuração estrutural na forma A. Uma vez que a forma B da molécula de DNA é a estrutura mais comum na natureza, assumiremos daqui em diante essa configuração nos exemplos e discussões da tese ao falarmos de DNA e a ilustramos na figura 2.1. Ainda que o foco da tese seja o estudo de RNAs, o conhecimento da estrutura do DNA facilita o entendimento das interações contidas nessas moléculas.

**Figura 2.1**

Estruturas atomística (a esquerda) e esquemática (a direita) para uma molécula de DNA com 12 pares de base e 8-oxaguanina na quarta posição. Figura da referência 13.

Figura 2.2

Um exemplo de nucleotídeo em DNA (a esquerda) e a visão esquemática dos nucleotídeos ao longo das duas cadeias poliméricas (a direita) enfatizando as ligações fosfato-açúcar (P-A-P) no *backbone* e a interação entre as bases (B-B) de cada monômero.

Os monômeros da molécula de DNA são formados por blocos construtores chamados nucleotídeos, onde cada nucleotídeo é formado por 3 elementos: um grupo fosfato PO_4^- , um anel de açúcar (formado por um grupo cíclico de 5 átomos) e uma base nitrogenada que é um complexo orgânico que pode ter um ou dois ciclos [14, 15]. A espinha dorsal (*backbone*) da dupla hélice é formada por uma sequência de grupos de açúcar e fosfato, e este é orientado devido ao grupo fosfato ser ligado a um átomo de carbono que não pertence ao anel do açúcar, mas o próximo grupo fosfato é ligado a um átomo de carbono que pertence ao anel orgânico, orientação conhecida como $5' \rightarrow 3'$. Essas interações se repetem ao longo de toda a molécula como visto na figura 2.2 e o outro monômero que compõe a dupla hélice possui uma orientação contrária, isto é, $3' \rightarrow 5'$.

Os açúcares e fosfatos ao longo da molécula são os mesmos, assim os nucleotídeos se diferem pela base que interage com o anel de açúcar. Para o DNA temos quatro tipos de base *adenina* (A), *timina* (T), *citocina* (C) e *guanina* (G). Essas bases são divididas em duas categorias, purinas e pirimidinas, onde são definidas pelas quantidades de ciclos orgânicos que cada uma possui. As purinas possuem dois ciclos orgânicos e as bases que constituem esse grupo são adenina e guanina. As pirimidinas possuem somente um ciclo orgânico, e as bases

que a constituem são timina e citosina.

Uma observação importante que levou Watson e Crick a descoberta da estrutura do DNA, foi o fato das bases nitrogenadas tenderem a se formar ligações de hidrogênio e os pares de base serem sempre formados por uma purina e uma pirimidina. No emparelhamento de Watson-Crick para o DNA, temos as interações A-T descrevendo duas pontes de hidrogênio e C-G que são ligadas por três pontes de hidrogênio, sendo então ligações mais robustas. É importante frisar aqui que a palavra robusta foi usada no sentido da ligação C-G ser mais estável que a ligação A-T, pois sabemos que ambos os tipos de ligações são fracas quando comparadas as ligações covalentes.

Além da estabilidade da molécula dada pelas ligações C-G ao longo da dupla hélice, um outro fator importante para a forma compacta da molécula é o empilhamento (*stacking*) das bases ao longo da dupla hélice. Essas interações possuem uma origem complexa, uma parte dada pela sobreposição dos orbitais π das bases e a outra parte das interações hidrofóbicas [14]. As interações impõem uma distância bem definida entre as bases ao longo do eixo da molécula, levando a um aumento da rigidez ao longo do eixo da dupla hélice. As bases podem entretanto deslizar umas sobre as outras nas direções perpendiculares ao eixo, e as energias de *stacking* são altamente dependentes das sequências de base. Com o objetivo de exemplificar o que foi dito, teremos diferentes valores de energia para os diferentes empilhamentos ao longo da molécula como mostrado em 2.1

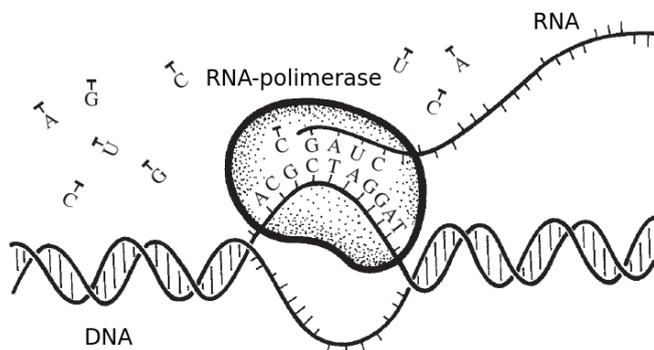
$$\begin{array}{ccc}
 \begin{array}{c} \uparrow \\ A-T \\ \downarrow \end{array} & \neq & \begin{array}{c} \uparrow \\ A-T \\ \downarrow \end{array} & \neq & \begin{array}{c} \uparrow \\ C-G \\ \downarrow \end{array} \\
 0.14\text{eV} & & 0.23\text{eV} & & 0.63\text{eV}
 \end{array} \tag{2.1}$$

2.2 Funções biológicas e RNA

O estudo do DNA é importante para o entendimento das funções biológicas presentes nos seres vivos. A sequência dos pares de base no DNA codificam a informação para a síntese das proteínas, que são polímeros feitos de 20 aminoácidos diferentes. Devido ao fato de termos

somente quatro pares de base, uma base não é o suficiente para codificar um aminoácido. A codificação acontece a cada três letras consecutivas (códon) na fita para indicar um aminoácido. Por exemplo, uma série CTT é a codificação para o aminoácido leucina. Com três letras por aminoácido o código tem algumas redundâncias, por exemplo, CTC também codifica o aminoácido leucina. Entretanto, algumas sequências específicas não são usadas para codificar aminoácidos, como é o caso do *stop codon* TAA responsável pelo fim de uma codificação de um gene para uma particular proteína. Além dos genes, o DNA possui regiões que não são lidas (genes não codificantes) onde seus papéis biológicos ainda não são completamente entendidos. Alguns deles poderiam ser somente remanescentes da evolução dos organismos.

A leitura do código genético é uma operação complexa porque as bases que contém a informação estão ocultas na dupla hélice e não estão favoráveis a uma reação química. Portanto, o acesso aos códons requeridos provocam grandes distorções na estrutura do DNA que expõem as bases para fora do empilhamento. Durante a *replicação* do DNA na divisão celular, a molécula é distorcida e suas pontes de hidrogênio são rompidas pela enzima helicase (processo semelhante a um zíper) separando as duas hélices. Ao fazer isso, as enzimas polimerases ligam nucleotídeos que estavam dispersos no meio aos nucleotídeos de uma das fitas, obedecendo as regras de pareamento. A outra fita é formada aos poucos devido a alta quantidade de nucleotídeos no meio durante a replicação. No fenômeno de *transcrição* do DNA a distorção é um pouco diferente. A transcrição do DNA é a leitura de um simples gene para a sintetização de uma proteína. Esse também é feito por uma enzima polimerase, mas de outro tipo, a RNA-polimerase. A transcrição é controlada pela enzima que escaneia a molécula de DNA até encontrar o sítio de iniciação, isto é, o começo do gene. Então, a enzima distorce a molécula de forma a se desenrolar em um comprimento de cerca de 15 à 20 pares de base. As bases expostas serão copiadas pela enzima ao longo da molécula e transcritas em uma molécula de RNA (mRNA ou RNA mensageiro) que irá ser modelo para a síntese da proteína. Na transcrição DNA-RNA, a enzima substitui os pares de base timina (T), pela base uracila (U) que também possui um ciclo orgânico. O emparelhamento Watson-Crick é respeitado no processo biológico, assim para o caso de termos uma dupla hélice de RNA, teremos adenina ligada com uracila (A-U) e citosina ligada com guanina (C-G). Falaremos mais sobre essas interações na próxima seção. O processo de transcrição DNA-RNA é conhecido como “transcrição tipo bolha” onde a

**Figura 2.3**

Esquema de uma fita de RNA sendo formada pela enzima RNA-polimerase. Note que ao ler a base adenina (A) o RNA passa a ter uma base uracila (U). A fita de DNA serve de modelo para a síntese precisa do RNA sempre respeitando o emparelhamento Watson-Crick. Figura retirada do capítulo 4, referência 15.

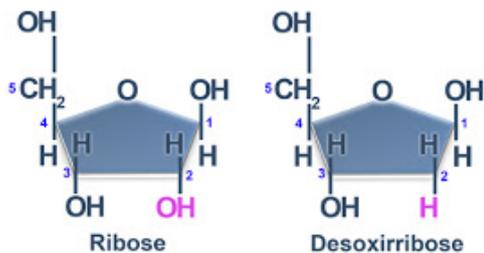
enzima se move ao longo de todo DNA fechando os pares de base que estavam abertos e abrindo outros até que o fim do gene seja alcançado. Como se trata de um processo bastante dinâmico, se pudéssemos tirar uma foto do fenômeno, o processo seria algo parecido como a figura 2.3.

As interações construídas no processo de formação do RNA merecem uma atenção especial, pois estão envolvidas diretamente nos processos biológicos no qual a molécula participa. A geometria associada a essas interações nos dão pistas sobre a estrutura que esses duplexos assumem. A seguir, discutiremos essas interações para um pareamento de duas pontes de hidrogênio e apresentaremos a nomenclatura de famílias para essas interações.

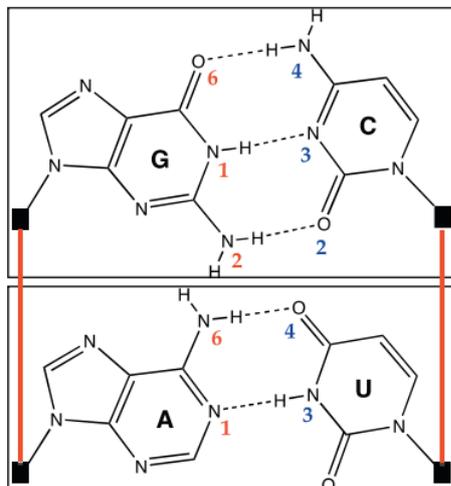
2.3 Propriedades físicas da molécula de RNA

Assim como o DNA, a dupla hélice de RNA é um polímero composto de monômeros entrelaçados, mas possui em sua estrutura diferentes componentes. Como dito anteriormente, a molécula de RNA possui as bases adenina (A), citosina (C), guanina (U) e uracila (U). Mas esta não é a única diferença comparada a molécula de DNA. Os nucleotídeos que formam a molécula de RNA possuem em sua estrutura um outro tipo de açúcar: a ribose. Como mostrado na figura 2.4, a pentose presente no DNA é a desoxirribose, enquanto que no RNA encontramos a ribose. Esses açúcares possuem cinco carbonos em sua estrutura e se diferenciam pela ausência de um grupo hidroxila (-OH) no carbono (C2') da desoxirribose. Essa configuração dos nucleotídeos na dupla hélice de RNA faz com que a mesma tenha sua estrutura mais compacta, a forma A, contendo em média 11 fosfatos por volta.

Essas base se combinam formando pares como a molécula de DNA. No pareamento Watson-Crick para a dupla hélice de RNA, temos o mesmo comportamento de uma base pirimidina ligado com uma purina. O RNA nesse contexto possui interações entre as bases

**Figura 2.4**

Açúcares presentes nas estruturas de RNA e DNA. Notamos que as pentoses se diferem por um grupo hidroxila (-OH). Figura da referência 16.

**Figura 2.5**

Interações no contexto Watson-Crick entre as bases nitrogenadas em RNAs. Na parte de cima da figura está representada as interações CG com suas três ligações de hidrogênio e abaixo da figura as interações AU com suas duas ligações de hidrogênio. As barras vermelhas da figura representam a isostericidade dos pares de base AU e CG. Figura da referência 17.

nitrogenadas adenina e uracila (AU) formando duas ligações de hidrogênios e interações entre as bases citosina e guanina (CG) que formam três pontes de hidrogênios como podemos ver na figura 2.5. Entretanto, esse tipo de interação é apenas uma fração das interações que a molécula de RNA pode assumir.

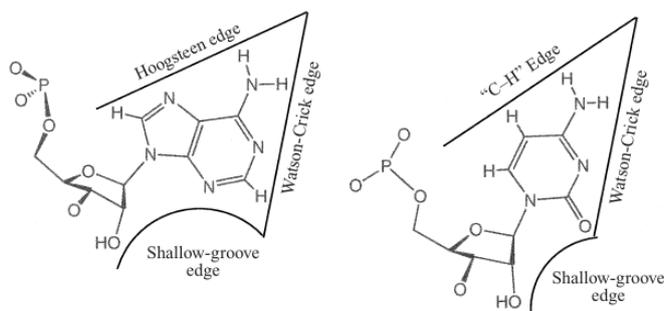
No contexto não canônico, as moléculas de RNA se dobras em uma grande e complexa variedade de estruturas tridimensionais (3D). Ainda que no contexto mais básico ao estudar as interações de pareamento de bases, novas interações continuam sendo descobertas [18]. A complexidade dessas configurações dificulta a predição de estruturas 3D e o entendimento de suas funções biológicas. Ao olharmos interações com duas pontes de hidrogênio, todos os pares são organizados pela simetria desses arranjos e tipos da base aromática em um total de 12 famílias [19]. As bases que compõem a molécula de RNA interagem de três maneiras possíveis ao longo de suas pontas: Watson-Crick (W-C), Shallow-groove (Sh-G) e Hoogsteen ou C-H edge dependendo se a base for uma purina ou pirimidina, como mostrado na figura 2.6. As interações glicosídicas, levam as bases a se orientarem de duas formas possíveis: *cis* ou *trans* em relação a um eixo paralelo as ligações de hidrogênio dos pares de base. Junto com as ligações glicosídicas e as interações de pontas, a conformação da base (*Anti* ou *Syn*) e a orientação local da fita (paralela e antiparalela), finalizam os arranjos gerados e as famílias. Três desses quatro

N	Orientação da ligação glicosídica	Fronteiras interagentes	Orientação da fita
1	Cis	Watson-Crick/Watson-Crick	Antiparalela
2	Trans	Watson-Crick/Watson-Crick	Paralela
3	Cis	Watson-Crick/Hoogsteen	Paralela
4	Trans	Watson-Crick/Hoogsteen	Antiparalela
5	Cis	Watson-Crick/Shallow-groove	Antiparalela
6	Trans	Watson-Crick/Shallow-groove	Paralela
7	Cis	Hoogsteen/Hoogsteen	Antiparalela
8	Trans	Hoogsteen/Hoogsteen	Paralela
9	Cis	Hoogsteen/Shallow-groove	Paralela
10	Trans	Hoogsteen/Shallow-groove	Antiparalela
11	Cis	Shallow-groove/Shallow-groove	Antiparalela
12	Trans	Shallow-groove/Shallow-groove	Paralela

Tabela 2.1

Geometrias possíveis no pareamento de bases com duas ligações de hidrogênios.

parâmetros definem o quarto, logo representamos as 12 famílias pela tabela 2.1.

**Figura 2.6**

As três fronteiras de interação de bases pirimidinas e purinas estão sujeitas. Hoogsteen (purinas) e C-H (pirimidinas) são equivalentes geometricamente. Figura adaptada da referência 19.

Uma vez introduzido as orientações das bases e da fita, podemos descrever melhor o pareamento Watson-Crick comum em duplas hélices de RNAs. Estruturas canônicas de RNAs possuem duas interações de fronteira W-C, fitas localmente antiparalelas, pontes glicosídicas orientadas como cis e conformações de base-açúcar no sentido anti. Os quatro pares de base no contexto Watson-Crick (AU, UA, CG e GC) geometricamente são caracterizados por mútua isostericidade, tendo como consequência biológica a permutabilidade desses pareamentos sem qualquer distorção da hélice (forma A) que eles compõem. Este tipo de configuração define a estrutura secundária de uma molécula de RNA.

Um outro tipo de pareamento muito comum em RNA capaz de descrever a estrutura secundária da molécula de RNA são as interações do tipo guanina-uracila (wobble GU pairs) mostrada na figura 2.7. Diferente do pareamento canônico, ligações do tipo GU não são auto-isostéricas, isto é, a geometria da configuração GU é diferente da geometria da configuração UG. Esse tipo de ligação frequentemente substitui pares de base canônicos em hélices de RNAs,

tendo como consequência o mínimo de distorção na molécula. Entretanto, a literatura mostra que o contexto de pareamentos do tipo GU e UG são muito mais amplos e complexos que uma simples substituição de pares de base canônicos como mostrado por Amarante e Weber [20].

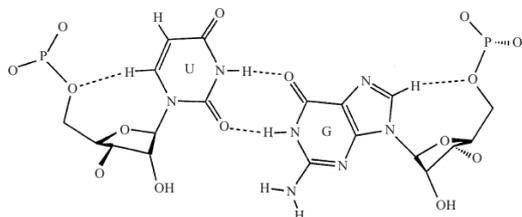


Figura 2.7

Esquema de ligações entre a uracila e a guanina do par de base GU em RNA. Figura retirada e adaptada da referência 19

As pontes de hidrogênio que interagem em RNAs podem assumir diferentes configurações. As interações mais comuns encontradas em RNAs canônicos são as $N \cdots H-N$ e as bifurcadas do tipo $N-H \cdots O \cdots H-N$, onde esse mesmo oxigênio faz também uma ligação do tipo covalente com um átomo de carbono ($O=C$), mas as mesmas podem ser encontradas em RNAs não canônicos. Entretanto, o contexto de RNAs não canônicos é mais rico e complexo ao se tratar de interações das bases. Em algumas geometrias, a molécula de água possui participação direta nas interações da estrutura. Essas interações são conhecidas como *água inserida* (water-inserted) ou *abertas* (open) uma vez que é possível a inserção de uma molécula de água em um dos sulcos da hélice. O oxigênio da molécula de água serve como uma espécie de ponte entre os monômeros da forma $N-H \cdots O_w \cdots H-N$. Também em interações não canônicas de RNAs é comum o aparecimento de ligações de hidrogênio, envolvendo átomos de hidrogênio polarizado covalentemente agrupados com um átomo de carbono do anel aromático da base, ligações que podem tomar as formas $C-H \cdots O$ ou $C-H \cdots N$. Estes tipos de interações são muito comuns em fronteiras do tipo Shallow-groove/Shallow-groove.

Uma última e não menos importante interação de pareamentos não canônicos é a configuração side-by-side, que envolve nucleotídeos adjacentes na mesma cadeia. Também conhecidas como plataformas adenosinas (adenosine platforms), essas interações foram previstas teoricamente para DNAs em 1995, mas foram encontradas em uma estrutura de RNA contendo bulges [19]. Essas configurações podem ser encontradas com as fronteiras interagindo da forma *trans* W-C/Hoogsteen e *cis* Hoogsteen/Shallow-groove, onde seu arranjo molecular é da forma $N-H \cdots O2'-C2'$ onde o oxigênio desse arranjo pertence a uma hidroxila, logo também faz uma ligação com um átomo de hidrogênio. A descrição dessas interações já nos fornece o conhecimento necessário para o objetivo dessa tese nos estudos de RNAs. Esses tipos de arranjos

moleculares foram obtidos principalmente por imagens cristalográficas. Para o encerramento do capítulo, descreverei o experimento óptico de desnaturação térmica. Dentre as técnicas experimentais existentes no estudo de duplexos, focaremos no fenômeno da desnaturação, pois utilizamos dados experimentais desse experimento no ajuste de parâmetros do nosso modelo.

2.4 Experimento de desnaturação térmica

Para o modelamento da dinâmica não-linear de duplas hélices de DNA e RNA, precisamos de um conjunto de dados experimentais para estabelecer um modelo mais significativo. Estes dados são fornecidos por diversos experimentos físicos, sendo obtido por métodos padrões da matéria condensada e também por métodos baseados em experimentos com simples moléculas. Para a dinâmica desses duplexos focaremos nos dados de medidas de desnaturação térmica de um conjunto de sequências para inferir suas propriedades mecânicas e termodinâmicas.

O fenômeno de desnaturação térmica de moléculas de DNA e RNA é a separação dos dois monômeros que constituem essas duplas hélices submetidas a um aumento de temperatura. Este tipo de técnica é fácil de monitorar, pois o rompimento das ligações de hidrogênio que constituem essas hélices é acompanhado de um grande aumento na absorvância de luz ultravioleta (UV) em 260 nm das sequências [21]. Dizemos que a molécula está desnaturada quando há separação total dos dois monômeros. Experimentos utilizando homopolímeros, isto é, sequências que contenham somente um tipo de ligação (AT ou CG para DNAs e AU ou CG para RNAs), apresentam uma transição mais acentuada que sequências que contenham combinações dessas interações como visto na figura 2.8. Entretanto as duplas hélices naturais onde as sequências possuem combinações dessas interações ocorrem em múltiplos passos. Experimentos utilizando a técnica cryomicroscopy demonstram que a desnaturação térmica dessas sequências começa como aberturas locais, gerando formação de uma espécie de bolhas em diferentes locais dessas sequências (denaturation bubbles) e a medida em que se aumenta a temperatura, essas bolhas invadem toda a molécula causando a separação dos monômeros como na simulação da figura 2.9. É preciso enfatizar que o fenômeno de desnaturação térmica da molécula é um processo reversível. O resfriamento da amostra com as sequências desnaturadas faz com que os monômeros procurem seus pares e voltam ao estado de dupla hélice. Assim, no caso de aque-

cermos duplexos de DNAs e RNAs todas em uma mesma amostra até uma temperatura onde todas as sequências estão desnaturadas, poderemos ter a formação de híbridos de DNA/DNA, RNA/RNA e também RNA/DNA [22].

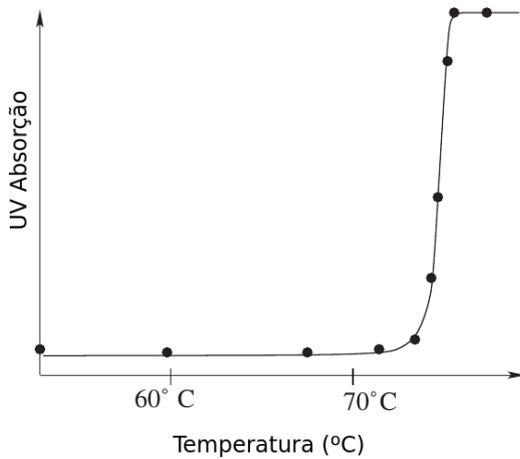


Figura 2.8

Curva de desnaturação de um homopolímero de DNA contendo somente pares de base CG. Figura retirada da referência 14

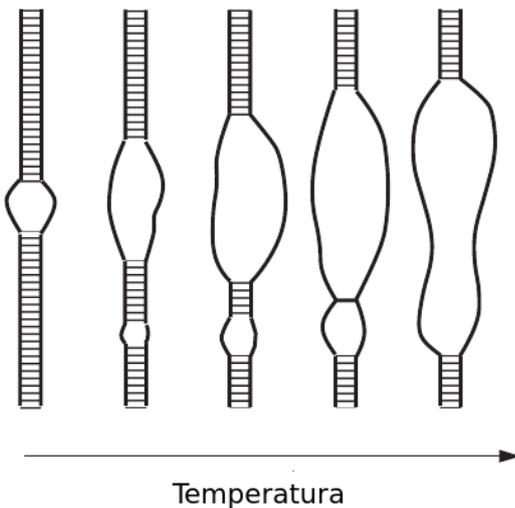


Figura 2.9

Desenho esquemático da desnaturação da molécula de DNA mostrando as aberturas locais que a fita sofre (*denaturation bubbles*) no processo de desnaturação. Figura retirada da referência 14

Em meio ao processo de desnaturação térmica da molécula a formação de bolhas acontece devido ao rompimento das interações de hidrogênio e orbitais π entre os anéis das bases dado o desenrolar da hélice. Nota-se então que gradualmente o aumento da temperatura modifica o estado eletrônico da molécula possibilitando o aumento de absorção de radiação. Devido à diferenças de gradientes de temperatura na solução e interações com o meio salino onde as sequências se encontram, algumas dessas duplas hélices se desnaturam antes que outras. Define-se então a temperatura de desnaturação ou temperatura de melting T_m onde metade dos duplexos presentes na solução estão dissociados. Nesse tipo de experimento, a temperatura de melting pode ser determinada por dois métodos. O primeiro método consiste em um fitting da curva de desnaturação da molécula onde se toma a sua primeira derivada encontrando o ponto de inflexão da curva de absorção e seu respectivo máximo [23]. O segundo método é feito

utilizando o método de van't Hoff tomando uma análise termodinâmica do sistema. O método gráfico de van't Hoff mostra a clara dependência da temperatura de melting e concentração da fita com as quantidades termodinâmicas ΔH° , ΔS° e ΔG° .

Nesse tipo de experimento é necessário ter o controle da concentração salina em que as sequências se encontram. Já foi mostrado por Ferreira *et al.* [24] em uma análise teórica que a solução salina nesses experimentos afetam as extremidades dessas hélices [24]. Sabe-se também que a dissociação da molécula também pode ocorrer em altas concentrações salinas [25, 26]. Então o controle do *buffer* (solução no qual as amostras se encontram contendo na maior parte das vezes NaCl, NaH_2PO_4 pH7 e EDTA dissódico) nessas soluções é essencial na realização do experimento.

As curvas de melting de dupla hélices são comumente descritas utilizando a teoria de transição *standard helix-to-coil*. Nesse caso, *helix* está associada ao estado de dupla hélice da molécula que representarei por X·Y e *coil* se refere aos monômeros desordenados após a desnaturação total da molécula como estado final descrito pela soma dos monômeros X e Y. Como o fenômeno de desnaturação se trata de um processo reversível, a equação de equilíbrio desse fenômeno pode ser descrita como



A constante de equilíbrio K_{eq} associada a equação (2.2) para a dissociação é relatada para o número total de concentração de oligonucleotídeos na solução C_t . O conceito de C_t é facilmente entendido quando pensamos em todas as hélices da solução dissociadas, isto é,

$$C_t = [\text{X}]_{\text{tot}} + [\text{Y}]_{\text{tot}},$$

onde os colchetes indicam a concentração dos monômeros. Entretanto, em meio ao processo de desnaturação podemos ter frações de hélices dissociadas e ligadas na solução e a identidade dessas sequências é importante no fenômeno.

A relação entre constante de equilíbrio K_{eq} e concentração total é conceitualmente diferente em três casos comuns: desnaturação de duplexos não-auto-complementares, duplexos

auto-complementares e em sondas que hibridizam com uma dada sequência, como em um reconhecimento de sítio. Descreverei aqui somente o caso de fitas auto-complementares e não auto-complementares em altas concentrações. O terceiro caso que argumentei acima, pode ser entendido como uma fita não-auto-complementar. Entretanto, em baixas concentrações a equação de balanço nesse terceiro caso, fração das duplexos, concentrações e constante de equilíbrio se tornam totalmente diferentes e deixo como leitura complementar o artigo de Schreiber-Gosche e Edwards [27].

Fitas não auto-complementares: Nesse contexto, a fração de duplexos dissociados é dado por

$$f = \frac{[Y]}{[X \cdot Y]_{\text{init}}} = \frac{[X]}{[X \cdot Y]_{\text{init}}}, \quad (2.3)$$

e a concentração total de fitas em uma dada temperatura descrita por $C_t = [X] + [Y] + 2[X \cdot Y]$.

Uma vez definindo a constante de equilíbrio da forma

$$K_{\text{eq}} = \frac{[X][Y]}{[X \cdot Y]}, \quad (2.4)$$

encontramos a relação entre a constante de equilíbrio, concentração total e fração de hélices dissociadas

$$K_{\text{eq}} = f^2 \frac{C_t}{2(1-f)}. \quad (2.5)$$

Na temperatura de desnaturação T_m onde a fração de duplexos dissociados é $1/2$, temos a constante de equilíbrio $K_{\text{eq}} = C_t/4$ em soluções com sequências não-auto-complementares.

Fitas auto-complementares: Nesse caso, podemos tomar a equação (2.3) como antes, mas tomamos $[Y] = [X]$. Isso leva a fração de duplexos dissociados a forma

$$f = \frac{[X]}{2[X \cdot X]_{\text{init}}}. \quad (2.6)$$

Entretanto, a concentração total de fitas para qualquer temperatura passa a ser escrita como $C_t = [X] + 2[X \cdot X]$. A partir da definição da constante de equilíbrio na equação (2.4), encontramos uma relação semelhante a equação (2.5) no caso de duplexos auto-complementares da

forma

$$K_{\text{eq}} = 2f^2 \frac{C_t}{1-f}. \quad (2.7)$$

Essa relação na temperatura de melting onde o fator $f = 1/2$, temos que a constante de equilíbrio para essa identidade de duplexos se torna igual a concentração total de oligonucleotídeos dissociados, isto é, $K_{\text{eq}} = C_t$.

Uma vez obtida a constante de equilíbrio K_{eq} é possível realizar a análise termodinâmica desse tipo de sistema via o método de van't Hoff. Jacobus Henricus van't Hoff considerou em seu estudo de pressão osmótica de duas soluções (passagem do solvente de um meio menos concentrado para um meio mais concentrado) o valor da pressão em uma membrana semipermeável suficiente para impedir a ocorrência da osmose. Podemos mapear esse problema no caso de dupla hélices, analisando a blindagem eletrônica dos duplexos com relação a troca de íons entre o solvente e as sequências. O arranjo dos duplexos de DNA e RNA canônicos, não permitem a entrada de partículas de águas ou solventes no meio em sua estrutura, sendo portanto uma molécula hidrofóbica.

A equação de van't Hoff para a pressão osmótica é baseada nas equações de gases ideais, tal que a pressão osmótica Π sentida pela membrana é da forma

$$\Pi = k'RTi, \quad (2.8)$$

onde R é a constante universal dos gases, T a temperatura do meio, k' é a concentração molar de soluto em solução (N/V) e i é o fator de van't Hoff. O fator de van't Hoff i é um parâmetro que contém a quantidade de partículas que serão ionizadas ou dissociadas em solução, após o alcance do equilíbrio químico e pode ser equacionado como

$$i = \alpha(q - 1) + 1,$$

onde α é o grau de ionização e q o número total de íons liberados na ionização de um composto.

A equação de van't Hoff (2.8) dessa maneira acaba sendo um pouco confusa por conter uma constante de gás ideal na equação de pressão osmótica. Manter uma constante de gás ideal em um sistema como esse pode gerar um mal entendimento do mapeamento do

problema. Sugiro reescrever a equação de van't Hoff na forma

$$\Pi = k'\eta T, \quad (2.9)$$

onde $\eta = Ri$ contém toda informação de ionização do sistema. A termodinâmica molar desse sistema pode ser escrita em função das densidades de energia e constante de equilíbrio da forma

$$du = Tds + \frac{\Pi}{k'^2} dk', \quad (2.10)$$

e minimizando a constante de equilíbrio teremos

$$dh = Tds + \frac{1}{k'} d\Pi, \quad (2.11)$$

$$dg = -sdT + \frac{1}{k'} d\Pi, \quad (2.12)$$

subtraindo a equação (2.12) da equação (2.11), concluímos que

$$dg = dh - d(sT) \quad (2.13)$$

que para uma temperatura fixa a equação (2.13) se resume a forma

$$dg = dh - Tds, \quad (2.14)$$

fechando a descrição termodinâmica para o método de van't Hoff.

Na descrição termodinâmica para as equações de densidades de estados da energia de Gibbs desse sistema (2.12), obtemos para uma temperatura fixa

$$\left(\frac{\partial g}{\partial \Pi} \right)_T = \frac{1}{\eta T} \left(\frac{\partial g}{\partial k'} \right)_T, \quad (2.15)$$

e dado que $\partial g / \partial \Pi = 1/k'$ a temperatura constante, obtemos a equação termodinâmica para a energia livre de Gibbs desse sistema

$$\left(\frac{\partial G}{\partial k'} \right)_T = \frac{N\eta T}{k'}. \quad (2.16)$$

Portanto, a energia livre de Gibbs desse sistema dada pelo processo de reação e produto é escrita da forma final como

$$\Delta G = N\eta T \ln k' \bigg|_{k'_r}^{k'_p} = -N\eta T \ln K_{eq} = -NRT i \ln K_{eq}. \quad (2.17)$$

Acontece que no caso de duplas hélices canônicas, não há ionização da molécula. Portanto, $\alpha = 0$ o que implica em $i = 1$ e a equação (2.17) se resume a forma

$$\Delta G = -NRT \ln K_{eq}. \quad (2.18)$$

Através da equação (2.14), concluímos a dependência da temperatura de melting com a constante de equilíbrio, chegando nas relações

$$\frac{1}{T_m} = \begin{cases} -\frac{R}{\Delta H} \ln C_t + \frac{\Delta S - R \ln 4}{\Delta H} & \text{fitas não auto-complementares,} \\ -\frac{R}{\Delta H} \ln C_t + \frac{\Delta S}{\Delta H} & \text{fitas auto-complementares.} \end{cases} \quad (2.19)$$

Assim, através de uma regressão linear utilizando o método de van't Hoff, podemos extrair os parâmetros termodinâmicos de entalpia ΔH° e entropia ΔS° e com isso ΔG° pela equação (2.14) a uma temperatura fixa. A visualização do método de van't Hoff pode ser vista na figura 2.10, no caso de uma dupla hélice de DNA. É conhecido na literatura que a quantidade de ligações de hidrogênio CG influencia no aumento da temperatura de melting. Entretanto, isso não é suficiente para a determinação dessa temperatura. O sequenciamento de pares de base AU e CG ao longo da molécula definem melhor o valor da temperatura de melting.

Em duplas hélices maiores, foi notado um problema na distinção dos estados das sequências nas amostras onde foram medidas as curvas de absorção UV [28]. O aumento na intensidade de absorbância das sequências indica a quantidade de pares de base abertos. Contudo, não é capaz de discernir em quais sequências o par de base aberto corresponde, pois a medida é uma propriedade do conjunto de todas as sequências que contém a amostra. Assim, não é possível distinguir medidas onde em uma intensidade de 50% pode corresponder a situação em que metade das sequências da solução se encontram no estado simples e metade com fitas duplas, das amostras em que as moléculas estão parcialmente abertas. Muitas vezes nesses tipos de sequên-

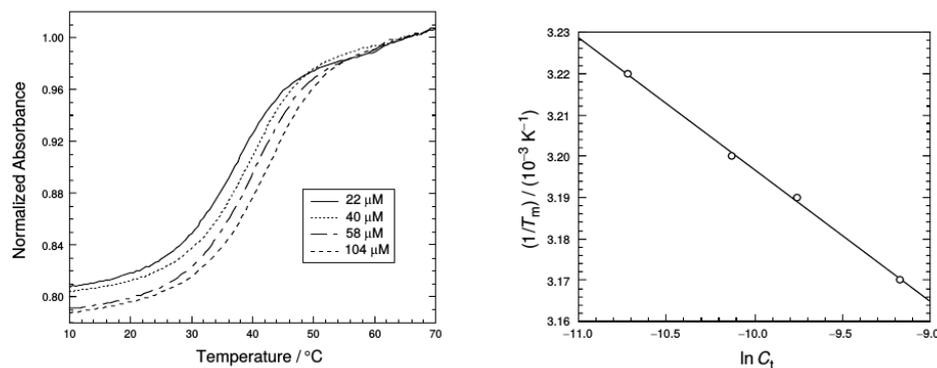


Figura 2.10

Curvas experimentais em um estudo do fenômeno de desnaturação térmica das sequências de DNA 5'CAAAAAAAG3' e 5'GTTTTTTC3', todas elas auto-complementares. À esquerda temos o perfil das duas sequências em diferentes concentrações e à direita o gráfico de van't Hoff onde cada ponto do gráfico representa uma das sequências. Figura retirada da referência 21 .

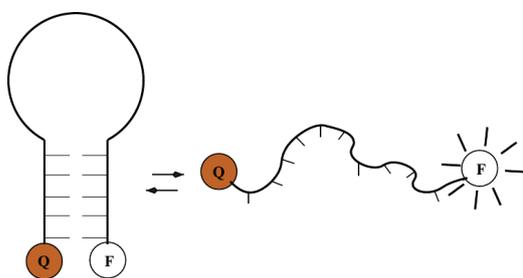


Figura 2.11

Molécula de DNA *hairpin* nas configurações fechada (esquerda) e aberta (direita), onde Q e F são *quencher* e *fluorophore*, respectivamente. Figura retirada e adaptada da referência 31.

cias buscando diferenciar essas situações, utilizam-se fitas auto-complementares que se dobram nelas mesmas formando um grampo de cabelo, conhecido como *estrutura hairpin* [29, 30]. As bolhas não pareadas formadas durante a desnaturação da molécula, se assemelham muito a estrutura do duplexo *hairpin*. Logo esses estudos se complementam, pois o entendimento do aumento de bolhas nas sequências através do *hairpin*, como também as diferentes conformações que podem levar a formação de um *hairpin* são possíveis em ambos estudos [31]. Nas extremidades de duplexos *hairpins* é comum se encontrar agentes químicos (*quencher* e *fluorophore*) que fluorecem ao ocorrer a desnaturação total do *hairpin*, como mostrado na figura 2.11.

Neste capítulo, vimos as propriedades de estruturas dos duplexos canônicos de DNA e RNA e algumas de suas funções biológicas. Focamos nos experimentos de desnaturação térmica onde conseguimos obter propriedades físicas energéticas desses duplexos. Entretanto, discutimos propriedades físicas de DNAs e RNAs canônicos. Os duplexos não canônicos possuem funções biológicas de extrema importância, em particular, iremos estudar conformações contendo protuberâncias (*bulges*). Introduzirei o conceito de *bulges* no capítulo 5. Mas antes, iremos estudar modelos termodinâmicos capazes de descrever estruturas e propriedades desses

duplexos.

3 Modelos de desnaturação de DNA e RNA

No experimento de desnaturação apresentado no capítulo 2, a determinação das propriedades físicas ΔH° e ΔS° quanto a temperatura de melting são importantes para a análise termodinâmica das sequências que contém as amostras. É bastante comum em estudos teóricos o uso dessas propriedades físicas no entendimento da estabilidade de sequências e predições de estruturas terciárias de DNAs e RNAs [32, 33].

Apresentamos neste capítulo dois modelos termodinâmicos comumente utilizados no estudo de propriedade de duplexos: o modelo de primeiros vizinhos (nearest neighbour model ou simplesmente, NN model) e o modelo Peyrard-Bishop (PB model). Ambos os modelos são excelentes alternativas nos estudos desses duplexos e possuem baixo custo computacional.

3.1 O modelo de primeiros vizinhos

O modelo de primeiros vizinhos é um modelo mais simples que associa quantidades de energia livre a cada par de interações na sequência e vem sendo extensivamente usado por bioquímicos e físicos na discussão de estabilidade de sequências de duplexos de DNA e RNA [33, 34].

Os incrementos de energias de cada interação da hélice dependem do primeiro par adjacente a partir da regressão de dados da concentração de duplexos versus temperatura de desnaturação para diferentes sequências. Esse modelo tem como premissa a suposição de uma energia inicial para que ocorra o processo de desnaturação ou hibridização das sequências [35], isto é, uma energia mínima para a ocorrência desses fenômenos.

O conceito do modelo de primeiros vizinhos é subdividir as sequências estudadas contabilizando as interações presentes em cada par de base adjacente. Assim, no caso de estudarmos uma duplo qualquer, como por exemplo a sequência canônica de RNA 5'CGAUUCG3', as

contribuições de energia da interação de cada par adjacente ficam da forma

$$\begin{aligned}
 & \begin{array}{l} 5' - \text{CGAUUCG} - 3' \\ 3' - \text{GCUAAGC} - 5' \end{array} = \\
 & \begin{array}{l} 5' - \text{CG} - 3' \\ 3' - \text{GC} - 5' \end{array} + \begin{array}{l} 5' - \text{GA} - 3' \\ 3' - \text{CU} - 5' \end{array} + \begin{array}{l} 5' - \text{AU} - 3' \\ 3' - \text{UA} - 5' \end{array} + \\
 & + \begin{array}{l} 5' - \text{UU} - 3' \\ 3' - \text{AA} - 5' \end{array} + \begin{array}{l} 5' - \text{UC} - 3' \\ 3' - \text{AG} - 5' \end{array} + \begin{array}{l} 5' - \text{CG} - 3' \\ 3' - \text{GC} - 5' \end{array}.
 \end{aligned}$$

Notamos que a sequência acima contém 7 pares de base, levando a 6 contribuições de energia dos pares adjacentes. Assim, dado uma sequência com N pares de base, teremos $N - 1$ contribuições de energia dada as interações relacionadas aos pares adjacentes. Se contarmos todas as configurações possíveis de pares de base, em um conjunto de sequências ou em uma sequência muito longa, poderemos ter 16 interações possíveis. Entretanto por questões de simetria de DNAs e RNAs canônicos, essas interações se resumem a 10 possibilidades.

A variação de entalpia e entropia totais de uma dada sequência é obtida experimentalmente pelo método de van't Hoff mostrado no capítulo 2, equação 2.19. As contribuições totais de entalpia (ΔH_n^{tot}) e entropia (ΔS_n^{tot}) descritas pelo modelo de primeiros vizinhos, podem ser resumidas na equações 3.1 e 3.2,

$$\Delta H_n^{\text{tot}} = \sum_i N_{in} \Delta H_i + \sum_j N_j^c \Delta H^c, \quad (3.1)$$

$$\Delta S_n^{\text{tot}} = \sum_i N_{in} \Delta S_i + \sum_j N_j^c \Delta S^c, \quad (3.2)$$

para uma dada sequência n . Os termos ΔH_i e ΔS_i representam as contribuições de entalpia e entropia das interações de dois pares de base ao longo da sequência, que no caso de moléculas canônicas, temos no máximo 10 interações de pares de base tanto para entalpia, quanto para entropia. O número de ocorrências N_{in} registra a quantidade de vezes que um mesmo tipo de interação aparece no duplexo. Os segundos termos das equações 3.1 e 3.2, ΔH^c e ΔS^c , representam contribuições adicionais dependendo da característica da sequência, incluindo também as entropias e entalpias iniciais para o processo de desnaturação ou hibridização. Devido a diversidade de possibilidades de correções sugeridas por diversos autores no modelo ao longo dos anos [34], coloca-se o termo N_j^c representando o número de correções necessárias dado o tipo de sequência e predição a ser estudada.

Após a definição dos termos relevantes nas equações 3.1 e 3.2, o modelo usa os dados experimentais da regressão de van't Hoff para as entalpias e entropias totais e através de uma linearização das equações por sequências do conjunto experimental estudado, determina os valores de interações dos vizinhos para os duplexos. Os resultados se tornam cada vez melhores, quanto maior for o conjunto de sequências de duplexos utilizados na linearização.

Uma técnica para a determinação dos parâmetros de interações do modelo pode ser feito por um método de otimização entre as temperaturas experimentais e as temperaturas preditas pelo modelo, partindo de um conjunto inicial de valores para os parâmetros de interação [34]. A técnica foi chamada de *otimização de temperaturas de melting* [34] (MTO-Melting Temperature Optimization) e definido o conjunto inicial de parâmetros, basta realizar uma minimização quadrática das diferenças de temperaturas predita T_i e experimental T'_i . Em um conjunto experimental contendo l sequências, a equação que ajusta os parâmetros do modelo é da forma

$$\chi^2 = \sum_{i=1}^l (T'_i - T_i)^2. \quad (3.3)$$

O método de otimização é interessante por dois principais motivos. O primeiro é que não se acumula erros devido a linearização feita para as sequências utilizando o método padrão [35]. O segundo motivo é a diminuição do número de parâmetros do modelo. Os termos de correções podem ser adicionados nos cálculos como um parâmetro único de correção ou até mesmo pode-se não utilizar correção alguma, e assim os valores dos parâmetros determinados possuem essas correções intrinsecamente. É possível a utilização total dos termos de correções como no método padrão, entretanto isso aumenta consideravelmente o número de parâmetros nos cálculos, o que exige como no método de SantaLucia *et al.* [35], um conjunto experimental grande de sequências. Um fato importante sobre a diminuição do número de parâmetros utilizando esse tipo de modelo é a convergência dos cálculos desses parâmetros. Gasta-se então computacionalmente um menor tempo e não é necessário um conjunto tão grande de sequências, ainda que um conjunto maior de sequências leve a melhores resultados.

Contudo, embora o modelo faça boas previsões para as temperaturas das sequências em termos energéticos, o modelo não é capaz de descrever as propriedades físicas associadas as interações intramoleculares das sequências [36]. Isso dificulta a discussão de propriedades físicas relacionadas as interações dessas moléculas como também a geometria adotada dos elementos

que constituem a formação das interações. Essa dificuldade é ainda mais evidente em obter bons resultados utilizando o modelo NN em estruturas assimétricas [1]. Sendo assim, precisamos de um modelo que seja capaz de descrever essas interações junto ao fenômeno da desnaturação. A resposta vem com a utilização do modelo Peyrard-Bishop que discutiremos a seguir. Esse modelo estatístico é capaz de estudar e descrever a curva de absorvância UV por temperatura, característica desse tipo de estudo, de um jeito simples e incluí em seu formalismo parte das interações intramoleculares contidas em dupla hélices.

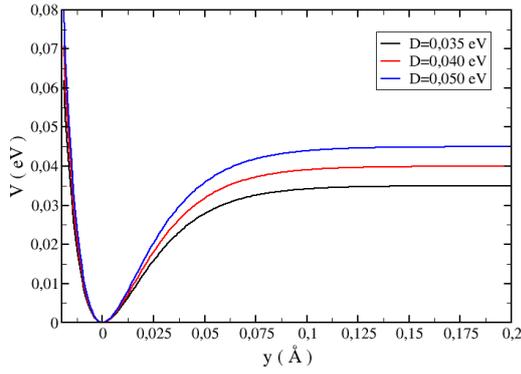
3.2 O modelo Peyrard-Bishop

O modelo Peyrard-Bishop é um modelo estatístico que nos permite modelar moléculas de DNA e RNA usando potenciais simples. Ao longo dos anos o modelo tem se mostrado um sucesso em diversos estudos de propriedades mecânicas e termodinâmicas dessas moléculas, exemplificado pelos trabalhos com análises de transcrições do tipo bolha [37, 38], as localizações de energia [39], análise da flexibilidade desses duplexos [40] e conformações de estruturas [20].

O modelo foi proposto por Peyrard e Bishop em 1989 [6]. Os autores estabeleceram que para cada par de base há dois graus de liberdade u_n e v_n , que correspondem ao deslocamento das bases com relação as posições de equilíbrio ao longo das pontes de hidrogênio que conectam as duas bases a formar o par. Assim, a dinâmica do DNA é descrita pelo Hamiltoniano

$$H = \sum_n \frac{1}{2}m(\dot{u}_n^2 + \dot{v}_n^2) + \frac{1}{2}k(u_n - u_{n-1})^2 + \frac{1}{2}k(v_n - v_{n-1})^2 + D(e^{-\frac{a}{\sqrt{2}}(u_n - v_n)} - 1)^2, \quad (3.4)$$

onde m é a massa reduzida dos pares de base (os autores assumiram que todos pares de base possuem a mesma massa reduzida), k é uma constante elástica do potencial harmônico, D é um parâmetro que representa a energia necessária para a dissociação do par de base e a representa o alcance do último potencial. O Hamiltoniano de Peyrard-Bishop é dividido basicamente em três partes: o primeiro termo representa a parte cinética da molécula. Isso nos faz lembrar que a molécula está em constante movimento sendo preciso levar em conta sua dinâmica nos cálculos de suas propriedades físicas. O segundo e terceiro termo são conjuntos de osciladores harmônicos acoplados que tem como objetivo descrever as ligações de açúcar-fosfato ao longo

**Figura 3.1**

Potencial de Morse em função do deslocamento y para $a = 4,45 \text{ \AA}$ e três valores diferentes da constante de dissociação D .

das hélices da molécula através de um empilhamento das bases com seus vizinhos (*stacking interaction*). O último termo do Hamiltoniano é o potencial de Morse [41], que é um potencial não-linear com o objetivo de descrever as ligações de hidrogênio das bases ligando os dois monômeros. Este é um potencial médio representando as duas ou três ligações de hidrogênio formadas pelas ligações das bases. O que distingue o número de ligações de hidrogênio é o parâmetro D devido a energia de dissociação do par de base. Na figura 3.1, mostramos alguns exemplos do potencial de Morse para o valor fixo de $a = 4,45 \text{ \AA}$ e diferentes valores da constante de dissociação D .

Uma vez apresentado os elementos do nosso Hamiltoniano, podemos escrever esse sistema de maneira mais conveniente fazendo mudanças de variáveis para uma melhor interpretação do sistema. Através de uma transformação canônica simples da forma

$$x_n = \frac{1}{\sqrt{2}}(u_n + v_n) \quad \text{e} \quad y_n = \frac{1}{\sqrt{2}}(u_n - v_n), \quad (3.5)$$

podemos representar a dinâmica da dupla fita de forma que o movimento possa estar em fase ou fora de fase. Como podemos ver, o primeiro termo da equação (3.5) representa a posição do centro de massa do par de base na molécula e o segundo termo associamos a dinâmica da molécula. Uma outra vantagem que obtivemos com a transformação acima é o desacoplamento das variáveis no Hamiltoniano da equação (3.4). Portanto, o Hamiltoniano pode ser reescrito da forma

$$H = H(x_n) + H(y_n) = \left[\sum_n \frac{p_{x_n}^2}{2m} + W(x_n, x_{n-1}) \right] + \left[\sum_n \frac{p_{y_n}^2}{2m} + W(y_n, y_{n-1}) + V(y_n) \right], \quad (3.6)$$

onde $p_{x_n} = m\dot{x}_n$, $p_{y_n} = m\dot{y}_n$ são os momentos canônicos para as novas variáveis x_n e y_n . Os potenciais $W(x_n, x_{n-1})$, $W(y_n, y_{n-1})$ e $V(y_n)$ são os mesmos potenciais do Hamiltoniano da equação (3.4) escritos nas novas coordenadas. Assim, para uma dada coordenada ξ_n qualquer, temos

$$W(\xi_n, \xi_{n-1}) = \frac{1}{2}k(\xi_n - \xi_{n-1})^2 \quad \text{e} \quad V(\xi_n) = D(e^{-a\xi_n} - 1)^2. \quad (3.7)$$

Para o estudo das propriedades físicas da nossa molécula, utilizaremos o formalismo de ensemble canônico e a função clássica de partição para a descrição do sistema. A escolha desse ensemble é justificada pelo caráter hidrofóbico da molécula com o solvente (não há ionização ou troca de partículas do solvente com a molécula), que consiste em um sistema do qual a molécula se encontra em um banho térmico. Seja uma molécula composta por N pares de base e homogênea, a integração nos momentos e nas variáveis de posição x_n , nos leva a função de partição da forma

$$Z = \int \prod_{n=1}^N dp_{x_n} dx_n dp_{y_n} dy_n e^{-\beta H(p_{x_n}, x_n, p_{y_n}, y_n)} = \left(\frac{8\pi^3 m^2}{k\beta^3} \right)^{N/2} Z_y, \quad (3.8)$$

onde $\beta = 1/k_B T$ sendo k_B a constante de Boltzmann e Z_y é o termo do Hamiltoniano que contém as integrações das posições da coordenada y_n , isto é

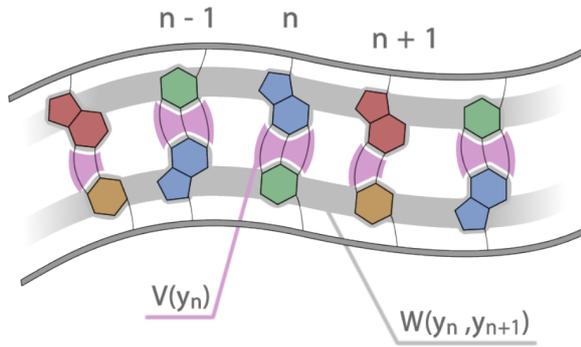
$$Z_y = \int \prod_{n=1}^N dy_n e^{-\beta(W(y_n, y_{n-1}) + V(y_n))} = \int \prod_{n=1}^N dy_n K(y_n, y_{n-1}), \quad (3.9)$$

onde

$$K(y_n, y_{n-1}) = \exp \left(-\beta \left\{ W(y_n, y_{n-1}) + \frac{1}{2}[V(y_n) + V(y_{n-1})] \right\} \right). \quad (3.10)$$

A solução das integrais anteriores contribuirá com uma constante multiplicando a parte dos potenciais da coordenada y_n . Como dito anteriormente, a dinâmica da molécula está associada a coordenada y_n , ou seja, associada a solução da função de partição Z_y . Esse é o motivo de muitas vezes encontrarmos na literatura do modelo Peyrard-Bishop somente a parte do Hamiltoniano associada a coordenada y_n , pois fisicamente é a parte mais relevante com respeito as propriedades físicas da molécula [42–44]. Um esboço das interações em função dessas coordenadas pode ser visto na figura 3.2

A função de partição Z_y pode ser avaliada de maneira exata no limite termodinâmico

**Figura 3.2**

Representação de uma duplex com as respectivas interações intramoleculares na molécula utilizando o modelo Peyrard-Bishop em função das coordenadas y_n .

($N \rightarrow +\infty$), usando as autofunções e autovalores do operador integral de transferência (IT). O operador é definido como

$$\int dy_{n-1} K(y_n, y_{n-1}) \varphi_i(y_{n-1}) = e^{-\beta \epsilon_i} \varphi_i(y_n), \quad (3.11)$$

e as autofunções φ_i permitem calcular Z_y e os valores esperados das quantidades associadas ao problema. Essas autofunções são interpretadas como distribuições de probabilidade para a amplitude y_n . É assumido que as funções φ_i formem um conjunto completo [45]

$$\delta(r - s) = \sum_{i=1}^N \varphi_i(r) \varphi_i^*(s), \quad (3.12)$$

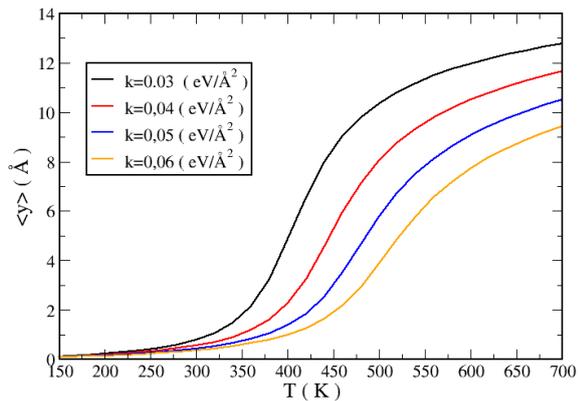
para quaisquer duas coordenadas r e s dos valores assumidos por y_n e y_{n-1} , respectivamente. Assim, a solução das integrais utilizando o operador IT no limite termodinâmico e condições de contorno apropriadas, resulta em

$$Z_y = e^{-\beta N \epsilon_0}, \quad (3.13)$$

onde ϵ_0 é o autovalor mais baixo da equação. Portanto, podemos agora fazer o estudo físico da molécula de DNA através de suas autofunções e dos seus autoestados.

No estudo da desnaturação térmica da molécula de DNA é importante determinarmos o comprimento médio das ligações de hidrogênio. Através de um cálculo similar ao feito para a equação (3.13), concluímos que

$$\langle y \rangle = \frac{1}{Z_y} \int \prod_{n=1}^N y K(y_n, y_{n-1}) dy_n \approx \langle \varphi_0(y) | y | \varphi_0(y) \rangle = \int \varphi_0^2(y) y dy, \quad (3.14)$$

**Figura 3.3**

Varição de $\langle y \rangle$ como função da temperatura T para quatro valores diferentes da constante de empilhamento k . Utilizamos os parâmetros $D = 0,04$ eV, $a = 4,45$ Å, $\theta = 0,01$ rad e variamos os valores de k .

para a autofunção normalizada $\varphi_0(y)$. Esta integral pode ser avaliada numericamente e mostra como o comprimento médio das ligações de hidrogênio $\langle y \rangle$ mudam com a temperatura T para diferentes empilhamentos como mostrado na figura 3.3.

Com o passar dos anos, o modelo Peyrard-Bishop foi sendo melhorado afim de se alcançar o melhor comportamento experimental possível. Nesse tempo, vários novos potenciais e técnicas de simulação foram implementadas [5, 44, 46–54], fazendo com que o modelo ganhasse ainda mais notoriedade no meio acadêmico. Apesar de inicialmente o modelo adotar a homogeneidade da molécula, este pode ser usado para qualquer tipo de sequência através da técnica IT [44]. A seguir, descreveremos um método capaz de fazer uma ligação direta entre o trabalho teórico de M. Peyrard e A. R Bishop com os trabalhos experimentais encontrados na literatura com o uso de sequências heterogêneas de DNA.

3.3 O modelo PB para sequências heterogêneas

Na seção anterior apresentamos o modelo PB para sequências de homopolímeros, isto é, compostos somente por um tipo de par de base. Para sequências heterogêneas, isto é, onde há ocorrência de pares tanto A-T como C-G, Zhang *et al.* [44] desenvolveram um método que usa uma expansão do núcleo nas bases de funções ortonormais para a integração da função de partição. A função de partição (3.9) agora é escrita com diferentes núcleos $K^{(i,i+1)}$, ou seja,

$$Z = \int dy_1 dy_2 \dots dy_N K^{(1,2)}(y_1, y_2) K^{(2,3)}(y_2, y_3) \dots K^{(N,1)}(y_N, y_1), \quad (3.15)$$

onde já incorporamos a condição de contorno do último par de base ligado ao primeiro par de base, isto é, $N + 1 \rightarrow 1$. Neste método, os diferentes núcleos $K^{(i,i+1)}$ são expandidos pelo mesmo conjunto de base. Dado a escolha do conjunto ortonormal $\{\varphi_m(y)\}$, podemos expandir o núcleo $K^{(i,i+1)}(x, y)$ da forma

$$K^{(i,i+1)}(x, y) = \sum_{n,m=1}^{+\infty} C_{n,m}^{(i,i+1)} \varphi_n(x) \varphi_m(y), \quad (3.16)$$

onde os coeficientes $C_{n,m}^{(i,i+1)}$ são determinados por

$$C_{n,m}^{(i,i+1)} = \int \int dx dy K^{(i,i+1)}(x, y) \varphi_n(x) \varphi_m(y). \quad (3.17)$$

Na prática, a equação (3.16) é truncada para as M primeiras bases, assim a equação é reescrita como

$$K^{(i,i+1)}(x, y) \approx \sum_{n,m=1}^M C_{n,m}^{(i,i+1)} \varphi_n(x) \varphi_m(y). \quad (3.18)$$

Substituindo a equação (3.18) na função de partição (3.15) e com um pouco de álgebra, a função de partição pode ser reescrita como

$$Z = Tr \left(\prod_{i=1}^N C^{(i,i+1)} \right). \quad (3.19)$$

ou seja, uma produtória de N matrizes onde antes tínhamos o produto de N autovalores na equação (3.13). No caso em que consideramos somente o emparelhamento tipo Watson-Crick, temos C-G como uma ligação forte (s) e A-T como uma ligação fraca (w), as matrizes C se reduzem a quatro tipos: $C^{(w,w)}$, $C^{(w,s)}$, $C^{(s,w)}$ e $C^{(s,s)}$ [44].

3.3.1 Equivalência térmica

O método usado para tratar o modelo PB em sequências heterogêneas pode ser usado para determinar um índice de equivalência que pode ser correlacionado com temperaturas de desnaturação medidas experimentalmente [40, 55]. Para isto, explora-se a produtória de matrizes da equação (3.19) separando as matrizes em componentes diagonais e não-diagonais.

Para calcular a função partição da equação (3.19) é necessário escolher um conjunto

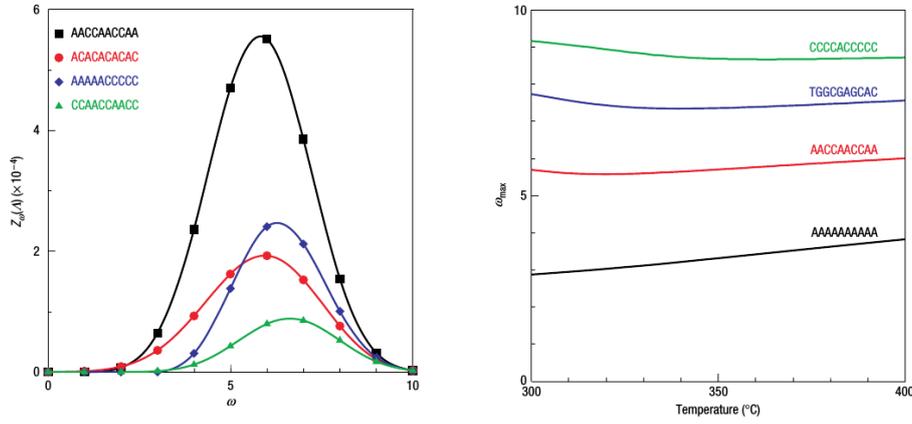


Figura 3.4

A figura a esquerda demonstra o comportamento gaussiano entre $Z_\omega(\Lambda)$ e ω para 4 sequências de 10 pares de base contendo 40% à 60% de pares de base C-G na temperatura de 370 K. A figura a direita mostra o comportamento de ω_{max} em função da temperatura para outras 4 sequências com 10 pares de base. Figura adaptada da referência 55.

de bases ortonormais. Zhang *et al.* [44] notaram que o melhor conjunto para ser usado na expansão da equação (3.16) para sequências heterogêneas são as autofunções da equação (3.11) para sequências homogêneas. Em nosso trabalho adotamos as bases para sequências tipo C-G, ou seja, a matriz $C^{(s,s)}$ é simplesmente uma matriz diagonal Λ com elementos λ_i , que são os autovalores para as sequências do tipo C-G. Portanto, escreveremos uma matriz qualquer $C^{(a,b)}$ da forma

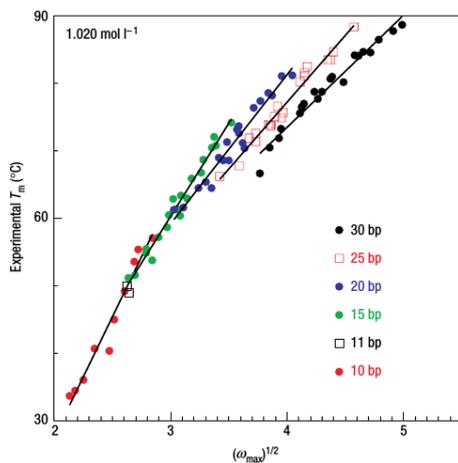
$$C^{(a,b)} = \Lambda + \Delta^{(a,b)}, \quad (3.20)$$

onde $\Delta^{(a,b)}$ representa a diferença entre a interação dos vizinhos do tipo (a,b) e os vizinhos (s,s) . Logo a função de partição (3.19) pode ser reescrita como

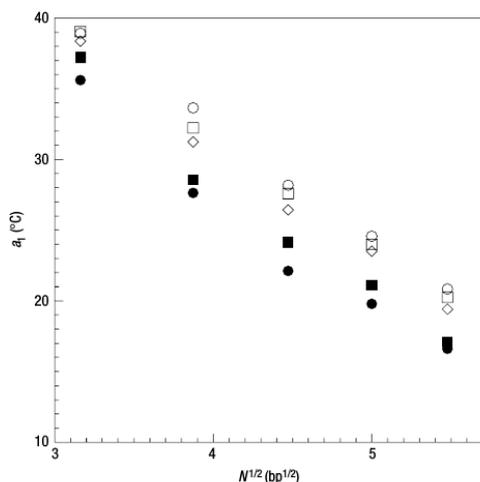
$$\begin{aligned} Z &= Tr[(\Lambda + \Delta^{(1,2)})(\Lambda + \Delta^{(2,3)}) \dots (\Lambda + \Delta^{(N,1)})] \\ &= \sum_{\omega=0}^N Z_\omega(\Lambda) = \sum_{\omega=0}^N Tr[M(\Lambda^\omega)], \end{aligned} \quad (3.21)$$

onde $M(\Lambda^\omega)$ são termos contendo ω multiplicações da matriz Λ . Para realizarmos os cálculos numericamente, avaliamos todos os termos da ordem de ω respeitando a não comutatividade das matrizes.

Como foi mostrado na referência 55 para o potencial de Dauxois *et al.* [5], o cálculo dos valores de $Z_\omega(\Lambda)$ em função de ω apresenta um perfil gaussiano em relação a ordem de pares de base C-G como se pode ver na figura 3.4. Os autores notaram que o termo ω_{max} tem

**Figura 3.5**

Temperaturas de *melting* experimentais (T_m) em função dos valores de ω_{max} para diversas sequências. Figura da referência 55.

**Figura 3.6**

Coefficiente de regressão por tamanho da sequência. O modelo considera a ordem de τ para as concentrações 0,069 mol/L (círculos), 0,119 mol/L (quadrados), 0,220 mol/L (losângulos), 0,621 mol/L (quadrado preenchido) e 1,020 mol/L (círculos preenchidos). Figura da referência 55.

pouca variação com mudanças de temperatura como mostrado a direita da figura 3.4. Isto é interessante e útil, pois agora temos um parâmetro que nos dá a assinatura de sequências que queremos analisar em termos físicos. Os autores então compararam a relação de ω_{max} com as temperaturas experimentais [56]. Podemos ver na figura 3.5 uma relação aproximadamente linear entre esses. Nota-se que o tamanho dos pares de base também são importantes no ajuste dos pontos e estes podem ser relacionados com as inclinações das retas da figura 3.5, como se pode ver na figura 3.6. Define-se então o *índice de melting* como $\tau = \omega_{max}^{1/2}$, e através de sua equivalência com as temperaturas experimentais, podemos prever temperaturas de *melting* para essas sequências com o modelo.

3.3.2 Parametrização do modelo PB

Como o modelo de primeiros vizinhos, o modelo Peyrard-Bishop também é capaz de prever a temperatura de melting. O modelo Peyrard-Bishop faz essa previsão através da curva de desvio médio das pontes de hidrogênio $\langle y \rangle$ por temperatura T . A dificuldade encontrada na

utilização do modelo PB é a definição dos valores dos parâmetros para a descrição correta da curva de desnaturação, levando a uma predição precisa da temperatura de melting. Como no modelo de primeiros vizinhos, fazemos uma minimização entre as temperaturas experimentais (T'_i) e as temperaturas preditas pelo modelo (T_i). Entretanto, as temperaturas preditas agora são determinadas pelo índice de melting τ . Efetuamos o processo de minimização onde os parâmetros da Hamiltoniana variam até reduzir o máximo essa diferença de temperaturas, descrevendo uma equação similar a 3.3, dada por

$$\chi^2 = \sum_{i=1}^l (T'_i - T_i)^2, \quad (3.22)$$

para um número de sequências l . Graficamente, podemos entender o conceito de equivalência térmica da parametrização acima, como um melhor ajuste linear da figura 3.5 [40].

Assim, obtido o menor valor de χ^2 , obtemos o melhor conjunto de parâmetros que se ajustam aos dados experimentais. Este procedimento não apenas permite uma melhor predição das temperaturas de desnaturação, mas os próprios parâmetros obtidos podem ser usados para inferir propriedades intramoleculares. Por exemplo, aplicando este método a sequências de RNAs canônicos, foi possível determinar que a ligação de hidrogênio para AU é mais forte que para AT em DNA [57]. Isso justifica um melhor empacotamento do RNA em sua estrutura comparado ao DNA.

3.4 Conclusão

Notamos nesse capítulo que o modelo Peyrard-Bishop é mais completo que o modelo de primeiros vizinhos, devido a sua descrição das interações contidas nos duplexos, sendo possível a determinação e descrição de propriedades físicas das mesmas. Entretanto, a aplicação do modelo PB fica limitada a estruturas canônicas, que são simétricas por natureza. No próximo capítulo, mostraremos que o modelo Peyrard-Bishop pode ser estendido a estruturas secundárias assimétricas, resultado que obtive durante o meu doutorado.

4 Empilhamento assimétrico no modelo PB

4.1 Introdução

A variedade de estruturas que duplexos de DNA e RNA podem assumir ainda não foi explorada utilizando o modelo PB. No caso de RNAs que possuem maiores amplitudes de movimentos em suas estruturas devido aos processos biológicos que são submetidas, que leva a um grande número de conformações algumas exemplificadas na figura 4.1, faz com que o modelo Peyrard-Bishop fique limitado a um grupo específico de estruturas secundárias simétricas. Ainda que o modelo Peyrard-Bishop vem sendo melhorado com o passar dos anos (capítulo 3), o modelo tem sua aplicação somente em estruturas canônicas simétricas.

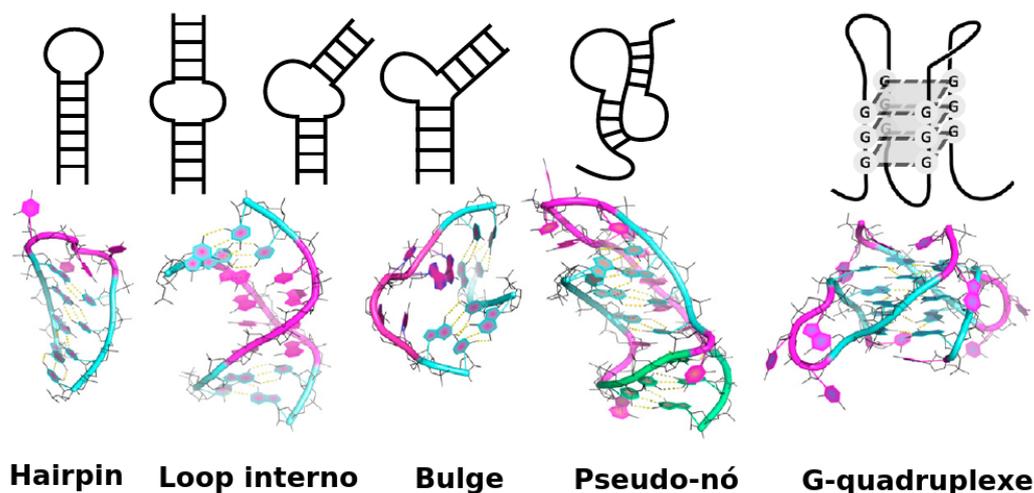


Figura 4.1

Estruturas assimétricas mais conhecidas em RNA. Da esquerda para a direita temos as conformações haste-loop ou hairpins, loops internos (internal-loop, protuberâncias (bulges), pseudo-nó (pseudo-knot) e quadruplexes (quadruplex). Figura adaptada da referência [58].

As estruturas simétricas de duplexos assumem uma forma compacta de forma que as bases nitrogenadas nessas configurações não permitem que as moléculas de água interajam com os átomos contidos na cadeia do duplexo. As posições das bases nitrogenadas ao longo da configuração direcionam a torção da molécula e com isso blindam os átomos da estrutura a

interagirem com as moléculas de água no meio. Ao longo da estrutura simétrica, teremos sempre um arranjo de 10 a 12 grupos fosfatos descrevendo um grande e um pequeno sulco nas hélices. Esse tipo de arranjo é repetido ao longo de cada volta. Estruturas assimétricas em duplexos, não formam esse tipo de arranjo. As distorções ao longo dessas estruturas podem ser laços ou/e protuberâncias. No caso de *loops* contidos nas estruturas, o despareamento das bases causam uma não uniformidade ao longo dos sulcos dessas moléculas, tornando a região mais ampla e expondo mais as bases e ligações contidas nos duplexos. Esse tipo de configuração perturba uma região maior no duplexo. Por outro lado, a ocorrência de protuberâncias na dupla hélice pode variar. Protuberâncias simples causam distorções menores quando comparadas a *bulges* de maiores ordens. Em ambos os casos, os vizinhos próximos ao *bulge* são afetados levando a uma assimetria da estrutura. Tanto em laços quanto as protuberâncias contidas em duplexos, levam a diferentes tamanhos de arranjos de grupos fosfatos por volta e afetam toda a molécula.

Começaremos adaptando o Hamiltoniano do modelo Peyrard-Bishop a descrever *stacking* para diferentes monômeros para potenciais harmônicos [6]. Mostramos que através de um tratamento algébrico adequado que o modelo Peyrard-Bishop com *stacking* harmônico descreve estruturas assimétricas e pode ser resolvido numericamente pelo tratamento padrão utilizando a técnica da integral de transferência (IT) estudado no capítulo 3. O resultado obtido se resume a uma constante efetiva definindo o empilhamento do modelo. Verificamos numericamente o resultado obtido e a concordância do mesmo para estruturas de DNAs.

Em seguida, buscamos o mesmo tratamento para potenciais anarmônicos [5]. Diferentemente do que se adota na literatura, incorporamos o potencial sugerido por Dauxois *et al.* no Hamiltoniano, tomando estes com constantes diferentes. Nesse caso, o tratamento algébrico nos leva a uma equação de Fredholm de segunda ordem, o que torna o método IT impraticável. Para a solução da função de partição associada a esse tipo de potencial, sugerimos o método numérico utilizando as funções de blocos pulsos bidimensionais (2D-BPFs). Explicamos brevemente o que são essas funções e suas propriedades matemáticas. Com o objetivo de demonstrar que o formalismo de 2D-BPFs se aplica ao modelo PB, aplicamos a técnica ao modelo original tomando uma duplice de DNA homogênea. Por fim, chegamos as conclusões e discutimos perspectivas futuras com a utilização de 2D-BPFs no estudo de duplexos assimétricos de DNAs e RNAs.

4.2 Stacking harmônico

Partindo do modelo sugerido por Peyrard e Bishop [6], modificamos os valores da constante de empilhamento para cada monômero na equação (3.4). Podemos ver essa modificação de forma intuitiva na figura 4.2 para uma molécula de DNA, onde o mesmo pode ser feito para uma molécula de RNA.

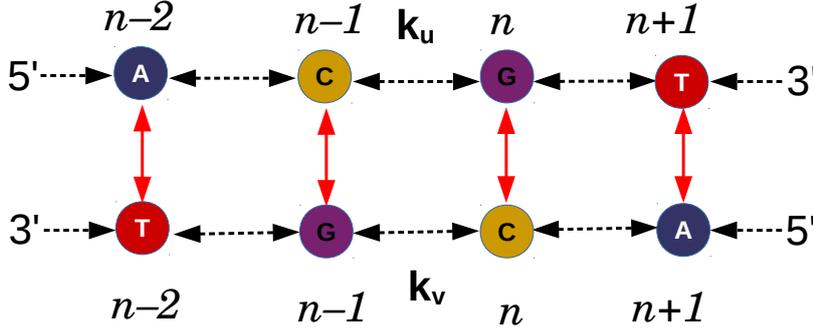


Figura 4.2

Esquemática da dupla hélice de DNA segundo o modelo Peyrard-Bishop com constantes assimétricas u e v .

O Hamiltoniano para N pares de base com a modificação proposta passa a ser da forma

$$H = \sum_{n=1}^N \frac{1}{2} m (\dot{u}_n^2 + \dot{v}_n^2) + \frac{1}{2} k_u (u_n - u_{n-1})^2 + \frac{1}{2} k_v (v_n - v_{n-1})^2 + D \left(e^{-\frac{a}{\sqrt{2}}(u_n - v_n)} - 1 \right)^2, \quad (4.1)$$

onde k_u e k_v correspondem as constantes dos monômeros u e v , respectivamente. Portanto, a função de partição integrada nos momentos para o novo Hamiltoniano resulta em

$$Z = (2\pi m k_B T)^N Z_{xy}, \quad (4.2)$$

onde

$$Z_{xy} = \int \prod_{n=1}^N dx_n dy_n \exp \left[-\beta \left\{ \frac{1}{2} k_a [(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2] + k_s (x_n - x_{n-1})(y_n - y_{n-1}) + V(y_n) \right\} \right], \quad (4.3)$$

com

$$k_a = \frac{1}{2}(k_u + k_v) \quad \text{e} \quad k_s = \frac{1}{2}(k_u - k_v). \quad (4.4)$$

Na tentativa de resolver analiticamente a função de partição nas coordenadas x_n , fatoramos a função de partição da equação (4.3) com um esforço algébrico. Reescrevemos a função de partição Z_{xy} na forma

$$\begin{aligned} Z_{xy} = & \prod_{n=1}^N \int dy_n \exp \left\{ -\beta \left[\frac{1}{2} \gamma (y_n - y_{n-1})^2 + V(y_n) \right] \right\} \\ & \times \int dx_n \exp \left\{ -\beta \left[\sqrt{\frac{k_a}{2}} (x_n - x_{n-1}) + \frac{k_s}{\sqrt{2k_a}} (y_n - y_{n-1}) \right]^2 \right\}, \end{aligned} \quad (4.5)$$

onde usamos a definição

$$\gamma = \frac{k_a^2 - k_s^2}{k_a}. \quad (4.6)$$

Introduzimos uma mudança de variável para a solução do conjunto de integrais na segunda parte da equação (4.5), onde chamamos a nova coordenada de r_n definida como

$$r_n = \sqrt{\frac{k_a}{2}} (x_n - x_{n-1}) + \frac{k_s}{\sqrt{2k_a}} (y_n - y_{n-1}), \quad (4.7)$$

com

$$dr_n = \sqrt{\frac{k_a}{2}} dx_n, \quad (4.8)$$

reduzindo o segundo grupo de integrais da equação (4.5)

$$\left(\frac{2}{k_a} \right)^{\frac{N}{2}} \prod_{n=1}^N \int dr_n e^{-\beta r_n^2} = \left(\frac{2\pi}{\beta k_a} \right)^{\frac{N}{2}}. \quad (4.9)$$

A função de partição descrita na equação (4.5) pode agora ser fatorada, resultando em

$$Z_x = \left(\frac{2\pi}{\beta k_a} \right)^{\frac{N}{2}}, \quad (4.10)$$

e

$$Z_y = \prod_{n=1}^N \int dy_n \exp \left\{ -\beta \left[\frac{1}{2} \gamma (y_n - y_{n-1})^2 + V(y_n) \right] \right\}. \quad (4.11)$$

O resultado obtido nos mostra que a função de partição Z_x e Z_y são funcionalmente idênticas

a função de partição do modelo original proposto por Peyrard e Bishop [6]. A mudança de variável levou ao caso unidimensional do modelo Peyrard-Bishop com uma constante efetiva γ e uma diferente constante multiplicativa devido ao resultado de Z_x que depende diretamente do empilhamento descrito no Hamiltoniano. Portanto, a condição de *stacking* simétrico imposta originalmente pelos autores [6], não é necessária utilizando-se o potencial harmônico. A condição que relaciona as constantes de empilhamento previamente estabelecidas é descrita por γ , levando a relação

$$\frac{1}{\gamma} = \frac{1}{k_u} + \frac{1}{k_v}, \quad (4.12)$$

que corresponde a um conjunto de osciladores com molas em série. O resultado mostra então que apesar de termos constantes diferentes, não podemos pensar no sistema individualmente. A transformação canônica serviu não apenas para mostrar o movimento do centro de massa da dupla hélice (x_n) e o deslocamento relativo entre os monômeros (y_n), mas também que os nucleotídeos pareados de cada monômero se comportam como um conjunto de osciladores harmônicos em série. Esse caráter só fica evidente ao utilizarmos diferentes constantes de empilhamento, quaisquer que sejam os valores de k_u e k_v .

Notamos que a utilização de constantes assimétricas no Hamiltoniano para o potencial harmônico de empilhamento, levam a mesma equação do modelo original Peyrard-Bishop [6]. Apesar de isso parecer evidente depois de demonstrado, isso nunca foi descrito ou comentado pelos autores do modelo até onde seja de nosso conhecimento. Portanto, esse foi um primeiro resultado do nosso estudo na busca de adaptação do modelo Peyrard-Bishop no tratamento de estruturas assimétricas.

4.2.1 Verificação numérica

Como o Hamiltoniano assimétrico (4.1) acaba tendo a mesma forma analítica do modelo original, podemos utilizar a técnica de integral de transferência (IT) para descrever o *stacking* assimétrico ao longo de uma dupla hélice de DNA. Como vimos anteriormente, a diferença entre os dois problemas é dada pela constante elástica contida na função de partição mostrada pela equação (4.12). Ainda que analiticamente a diferença se baseie no *stacking* entre os pares de base contidos na molécula, numericamente foi necessário alterar o algoritmo

Intra-strand	Inter-strand
ApA \equiv TpT, ApC \equiv GpT,	AsA, TsT, AsC, GsT,
ApG \equiv CpT, CpA \equiv TpG,	AsG, CsT, CsA, TsG,
CpC \equiv GpG, GpA \equiv TpC,	CsC, GsG, GsA, TsC,
ApT, CpG, GpC, TpA	AsT, CsG, GsC, TsA

Tabela 4.1

A esquerda da tabela, temos os parâmetros intra-strand definidos pela notação XpY. A direita, os parâmetros inter-strand representados pela notação XsY.

desenvolvido pelo nosso grupo de pesquisa [59] de forma que seja possível declarar as constantes elásticas k_u e k_v separadamente. Isso torna possível a minimização dessas constantes e podemos determina-las através do método de equivalência, como no modelo original [32].

Nesse tipo de problema, uma distinção dos cálculos se faz necessária. No modelo para moléculas simétricas, as constantes elásticas dependiam das bases, que são próximos-vizinhos. Dada todas as combinações possíveis para as quatro bases orgânicas no *stacking* ao longo da hélice, chegaríamos ao número de 16 combinações totais, mas considerando a simetria da molécula, reduziríamos as constantes a 10 como mostrado na tabela 4.1. Nomeamos esses parâmetros como *intra-strand*, por estarem contido em ambos os monômeros da molécula e identificamos suas interações de empilhamento pela notação XpY. Essa notação nos diz que ao termos uma interação de stacking, como por exemplo GpT, significa que temos um par de base G-C seguido de um par de base A-T.

No caso de uma molécula assimétrica, os monômeros contidos possuem constantes elásticas diferentes. Ainda que utilizemos sempre a constante efetiva dada pela equação (4.12), não podemos explorar a simetria dessa molécula por motivos óbvios. Isso aumenta o número de parâmetros a ser minimizado, isto é, teremos que otimizar os 16 parâmetros de *stacking* como mostrado na segunda parte da tabela 4.1. Nomeamos esse parâmetros como *inter-strand*, pois se trata de constantes elásticas em monômeros diferentes interagindo entre si e utilizamos a notação XsY para indicar a interação na direção $5' \rightarrow 3'$ descrevendo sempre o monômero u do Hamiltoniano (4.1). Portanto, ao termos por exemplo uma interação do tipo GsT, estamos dizendo que na direção $5' \rightarrow 3'$ temos um *stacking* em u (k_u) das bases G e T e na direção contrária, $3' \rightarrow 5'$, um *stacking* das base C e A no monômero v (k_v).

Dados experimentais

O nosso próximo passo foi buscar sequências de DNA com diferentes concentrações na literatura. Encontramos o conjunto de sequências do grupo de Owczarzy *et al.* [56] com medidas de temperaturas de melting dado pelo experimento de absorção UV em função da temperatura em diferentes concentrações salinas, determinadas em dois caminhos: um ajuste das curvas de absorção e pelo método de van't Hoff apresentado na seção 2.4 do capítulo 2. Owczarzy *et al.* utilizaram um grupo de 92 sequências para as concentrações salinas de 0.069 mol/L, 0.119 mol/L, 0.220 mol/L, 0.621 mol/L e 1.020 mol/L. Aproveito aqui para relembrar também o conceito de temperatura de melting também apresentado na mesma seção do capítulo 2. A temperatura de melting é definida quando metade das sequências da solução contendo os duplexos (que nesse caso são DNAs) estão desnaturadas e a outra metade da solução não. Em seguida, utilizamos as temperaturas de melting dos autores com suas respectivas sequências para minimizar os parâmetros do nosso modelo pelo método de equivalência térmica discutido na seção 3.3.1. Buscamos agora discutir as propriedades físicas de duplexos assimétricos através dos parâmetros ajustados pela minimização.

Notação

Na descrição das pontes de hidrogênio pelo potencial de Morse, teremos os dois parâmetros ligados a energia de dissociação das bases D_{AT} e D_{CG} correspondendo as bases AT e CG, respectivamente. Os parâmetros de *stacking* entre os nucleotídeos vizinhos descritos pelo modelo PB, foram definidos pela notação *inter-strand* anteriormente. Por exemplo, no caso de um empilhamento entre uma adenina (A) e uma citosina (C) como vizinha, o parâmetro associado a essa interação é descrito como

$$\text{AsC:harmonic.ks.}$$

A notação do empilhamento se resume a constante efetiva, mas que rotula o monômero na direção $5' \rightarrow 3'$. Agora que estabelecida a notação dos nossos parâmetros, podemos seguir com o procedimento de minimização para os 18 parâmetros definidos pelo modelo no caso de fitas assimétricas.

Procedimento de minimização

O procedimento de minimização dos parâmetros do modelo foi feito em dois passos. Partimos de parâmetros iniciais genéricos com objetivo de evitar qualquer resultado tendencioso da minimização. Realizamos esses dois processos de minimização para as sequências de Owczarzy *et al.* em todas as concentrações (0.069 mol/L, 0.119 mol/L, 0.220 mol/L, 0.621 mol/L e 1.020 mol/L).

O primeiro passo da minimização consiste em partir de um conjunto inicial de valores para os parâmetros do modelo PB, definidos como $D_{AT} = 32$ meV, $D_{CG} = 73$ meV, e todos os parâmetros de *stacking* com valor de $k = 5.0$ eV/nm². Usamos a equação (3.22) variando todos os 18 parâmetros,

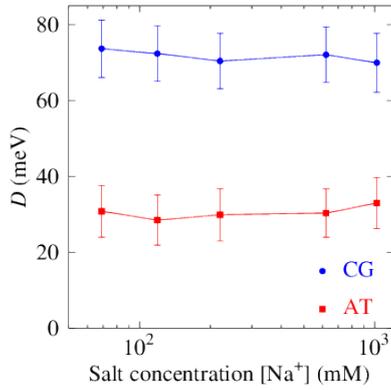
$$\chi^2 = \sum_{i=1}^l (T'_i - T_i)^2,$$

e repetimos o procedimento de minimização 200 vezes partindo de diferentes condições de contorno buscando um mínimo global. O resultado de cada parâmetro é obtido pela média sobre todos os resultados fornecidos pela minimização.

O segundo passo da minimização consiste em uma vez que aceito a minimização de χ^2 anteriormente, os valores médios obtidos serão usados como parâmetros iniciais. Entretanto, agora o conjunto experimental de temperaturas de melting é randomicamente alterado por pequenas quantidades correspondente aos seu erros experimentais. Tomamos o erro experimental de 1.0°C, erro aproximadamente ao de estruturas simétricas [32, 57] Este tipo de minimização nos permite explorar a influência do erro experimental em nossos parâmetros, levando a uma melhor precisão. Repetimos a minimização também em 200 vezes e o resultado final também é dado pela média das minimizações.

Resultados

Começamos discutindo os resultados obtido para as interações de hidrogênio em diferentes concentrações, ou seja, os parâmetros de Morse D do modelo PB que apresentamos na figura 4.3. Notamos que não há uma dependência significativa desses parâmetros com a concentração de sal. Os parâmetros de Morse obtidos estão consistentes com os valores de

**Figura 4.3**

Parâmetro de Morse D calculado por minimização para sequências A-T e C-G para o Hamiltoniano com constantes de empilhamento assimétricas.

minimizações já feitas anteriormente por nosso grupo [40, 57]. Ao compararmos o resultado com os parâmetros de Morse para o caso de constantes simétricas [40], não notamos uma diferença significativa entre os valores das constantes de dissociação. O resultado já é esperado, uma vez que não se altera o potencial de Morse na equação (4.1). Optamos por não variar os parâmetros de alcance λ_{AT} e λ_{CG} por dois motivos. Primeiro, buscamos mostrar com a verificação numérica que a mudança no Hamiltoniano 4.1 produz resultados consistentes a parte analítica proposta na seção anterior. Segundo, buscamos não aumentar a quantidade de parâmetros no processo de minimização.

Entretanto, o resultado da minimização para os parâmetros de empilhamento do DNA mostrados na figura 4.4, apresenta uma dependência com a concentração salina. O resultado pode ser discutido pela natureza dos ciclos orgânicos contido nas bases, pois sabemos que esses influenciam nas ligações π contida nas hélices. Ao analisar um empilhamento entre os vizinhos do tipo purina-purina (AsA, AsG, GsA e GsG), notamos que essas não são afetadas significativamente pelo aumento da concentração de sal. O mesmo pode ser concluído para o grupo de vizinhos do tipo pirimidina-pirimidina (CsC, CsT, TsC e TsT), o que é esperado devido somente a mudança de direção da posição das bases ao longo dos monômeros. O mesmo não pode ser dito no geral ao analisarmos interações dos grupos purina-pirimidina (AsT, GsC e GsT) e pirimidina-purina (TsA e TsG). Entretanto, encontramos exceções nos dois grupos. Notamos uma estabilidade com a concentração de sal para o empilhamento AsC no grupo purina-pirimidina e para os empilhamentos CsA e CsG no grupo pirimidina-purina.

Com relação aos valores encontrados para o empilhamento, os mesmos também são consistentes ao já calculados para DNA em nosso grupo de pesquisa [40]. Os valores obtidos são o dobro dos valores comparados ao cálculo com constantes simétricas [32]. O que nos leva a

pergunta: “Se usamos 16 parâmetros assimétricos continuamos obtendo algo que é equivalente aos 10 parâmetros simétricos?” A verificação numérica diz que sim. Apesar de ser um resultado simples, isto responde outra pergunta básica que fizemos quando começamos a estudar o modelo: “A aproximação de constantes iguais é realmente necessária?” Os resultados nos mostram que não. Entretanto, a utilização de constantes simétricas é uma boa aproximação para o modelo PB.

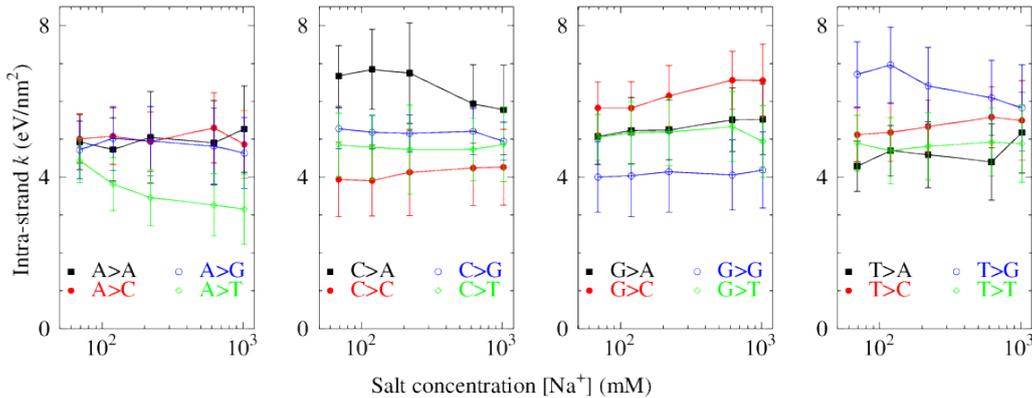


Figura 4.4

Parametrização da constante de empilhamento K para o potencial harmônico levando em conta todas as interações no *stacking* para fitas assimétricas.

Mostramos nesta seção que o cálculo utilizando constantes assimétricas para o potencial harmônico é possível no modelo Peyrard-Bishop. Nosso próximo passo é tentar utilizar essa assimetria no potencial sugerido por Dauxois *et al.* [5].

4.3 Stacking anarmônico

O modelo original introduzido em 1989 não reproduz bem a transição abrupta da desnaturação como observada experimentalmente para hélices longas como mostrado na figura 2.8. Em função desta observação, Dauxois *et al.* [5] introduziram um potencial não-harmônico de forma *ad hoc* buscando a uma melhor descrição do fenômeno de desnaturação do modelo Peyrard-Bishop [5]. Entretanto, Zhang *et al.* [44] mostraram que o resultado obtido pelo grupo de Dauxois é na verdade um artefato numérico. Ao se resolver o problema da divergência, o potencial proposto por Dauxois *et al.* não reproduz uma transição abrupta, como foi mostrado por Weber [49].

Partimos agora do pressuposto que o potencial de empilhamento adotado no modelo Peyrard-Bishop não seja harmônico. Vamos utilizar o potencial de Dauxois *et al.* [5] que é descrito como

$$W(y_n, y_{n-1}) = \frac{1}{2}k [1 + \rho e^{-\alpha(y_n + y_{n-1})}] (y_n - y_{n-1})^2, \quad (4.13)$$

onde ρ é uma constante que mede o grau de anarmonicidade e α tem unidade de inverso de comprimento medindo parte do alcance desse potencial. É importante notar que se $\rho \ll 1$, voltamos a ter novamente um potencial harmônico para descrever o empilhamento. Como se pode ver pela equação (4.13), o termo que representa a anarmonicidade do potencial também depende da posição dos nucleotídeos y_n e y_{n-1} . Para poder realizar a análise do empilhamento assimétrico reescrevemos os termos de potencial, substituindo a constante elástica k por uma função $k(y_n, y_{n-1})$ que leva sozinha o termo de anarmonicidade

$$k(y_n, y_{n-1}) = k [1 + \rho e^{-\alpha(y_n + y_{n-1})}], \quad (4.14)$$

isto é, temos um parâmetro elástico que varia com a posição em cada bloco de nucleotídeo e seus vizinhos ao longo da dupla hélice. Desta maneira, a equação (4.13) toma uma forma semelhante ao caso harmônico

$$W(y_n, y_{n-1}) = \frac{1}{2}k(y_n, y_{n-1})[y_n - y_{n-1}]^2. \quad (4.15)$$

Agora, ao invés de introduzir constantes assimétricas k_u e k_v como no caso harmônico, introduzimos a assimetria no coeficiente de anarmonicidade ρ_u e ρ_v ,

$$k_u(u_n, u_{n-1}) = k [1 + \rho_u e^{-\alpha(u_n + u_{n-1})}], \quad (4.16)$$

$$k_v(v_n, v_{n-1}) = k [1 + \rho_v e^{-\alpha(v_n + v_{n-1})}], \quad (4.17)$$

mantendo k igual para ambos os lados da fita. Com isso, nossa Hamiltoniana para monômeros assimétricos em um duplexo de DNA passa a ser escrita como

$$H = \sum_n \frac{1}{2m} (P_{u_n}^2 + P_{v_n}^2) + \frac{1}{2}k_u(u_n, u_{n-1})[u_n - u_{n-1}]^2 + \frac{1}{2}k_v(v_n, v_{n-1})[v_n - v_{n-1}]^2 + D \left(e^{-\frac{a}{\sqrt{2}}(u_n - v_n)} - 1 \right)^2. \quad (4.18)$$

Utilizaremos a mesma transformação canônica do modelo original, seguindo um raciocínio análogo ao feito para a Hamiltoniana (4.1) contendo potenciais harmônicos e *stacking* assimétrico. A Hamiltoniana em função de x_n e y_n pode ser escrita como

$$H = \sum_n \frac{1}{2m} (P_{x_n}^2 + P_{y_n}^2) + k_a(x_n, x_{n-1}, y_n, y_{n-1}) [(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2] \quad (4.19)$$

$$+ k_s(x_n, x_{n-1}, y_n, y_{n-1}) \{ [x_n - x_{n-1}] [y_n - y_{n-1}] \} + D(e^{-ay_n} - 1)^2.$$

Com um pouco de álgebra é possível ver que as funções k_a e k_s são escritas da forma,

$$k_a(x_n, x_{n-1}, y_n, y_{n-1}) = \frac{1}{4} [k_u(u_n, u_{n-1}) + k_v(v_n, v_{n-1})]$$

$$= \frac{k}{4} \left\{ 2 + e^{\frac{-\alpha}{\sqrt{2}}(x_n - x_{n-1})} \left[\rho_u e^{\frac{-\alpha}{\sqrt{2}}(y_n - y_{n-1})} + \rho_v e^{\frac{\alpha}{\sqrt{2}}(y_n - y_{n-1})} \right] \right\}, \quad (4.20)$$

$$k_s(x_n, x_{n-1}, y_n, y_{n-1}) = \frac{1}{2} [k_u(u_n, u_{n-1}) - k_v(v_n, v_{n-1})]$$

$$= \frac{k}{2} \left\{ e^{\frac{-\alpha}{\sqrt{2}}(x_n - x_{n-1})} \left[\rho_u e^{\frac{-\alpha}{\sqrt{2}}(y_n - y_{n-1})} - \rho_v e^{\frac{\alpha}{\sqrt{2}}(y_n - y_{n-1})} \right] \right\}. \quad (4.21)$$

Portanto a função de partição para esse Hamiltoniano fica da forma

$$Z = \int \prod_n dx_n dy_n \exp \left\{ -\beta \left[H_{\text{harm}} + \frac{-k}{4} \sum_n e^{\frac{-\alpha}{\sqrt{2}}(x_n - x_{n-1})} \left(\rho_u e^{\frac{-\alpha}{\sqrt{2}}(y_n - y_{n-1})} \times \right. \right. \right. \right.$$

$$\left. \left. \left. \times [(x_n - x_{n-1}) + (y_n - y_{n-1})]^2 + \rho_v e^{\frac{\alpha}{\sqrt{2}}(y_n - y_{n-1})} [(x_n - x_{n-1}) - (y_n - y_{n-1})]^2 \right] \right\} \quad (4.22)$$

Ou seja, a nossa função de partição possui um termo correspondente ao Hamiltoniano do modelo original H_{harm} e um termo adicional dado pela anarmonicidade e assimetria do potencial de empilhamento. Entretanto, a nossa função de partição não apresenta uma forma tão simples e separável para aplicarmos a técnica de integral de transferência como feita anteriormente (3.8 e 4.2). Isto se deve ao fato do nosso núcleo agora possuir quatro variáveis para cada integral na função de partição, sendo não separável e não permitindo que cheguemos a uma equação de primeira ordem, mas sim uma equação de segunda ordem na teoria de Fredholm para equações integrais.

A pergunta que nos fizemos a seguir foi a da possibilidade de se utilizar uma teoria que consiga trabalhar com um núcleo não separável como o que encontramos na equação (4.22). Isso seria interessante não somente para os potenciais descritos aqui, mas também na tentativa

de implementar outros potenciais. Encontramos nossa resposta utilizando blocos de funções de pulso bidimensionais (2D-BPFs) e apresentamos o formalismo que envolve essas funções a seguir.

4.3.1 Aplicação blocos de funções de pulso (2D-BPFs) a núcleos de 4 variáveis genéricos

Os blocos de funções de pulsos (BPFs) são funções bem comportadas como veremos a seguir. A escolha por utilizar essas funções foi tanto pela tentativa de uma solução analítica da função de partição, quanto pelo fato de essas funções em duas dimensões (2D-BPFs) já terem sido utilizadas em equações integrais não lineares, como por exemplo, as equações de Volterra de primeiro e segundo tipo [60]. Buscamos aqui adaptar essas funções para as equações de Fredholm do nosso problema.

Seja um conjunto $m_1 m_2$ de 2D-BPFs $\Phi_{i_1 i_2}(x, y)$, onde $i_1 = 1, 2, \dots, m_1; i_2 = 1, 2, \dots, m_2$. Essas funções são definidas na região $x \in [0, T_1)$ e $y \in [0, T_2)$ da forma:

$$\Phi_{i_1 i_2}(x, y) = \begin{cases} 1, & (i_1 - 1)h_1 \leq x < i_1 h_1 \text{ e } (i_2 - 1)h_2 \leq y < i_2 h_2, \\ 0, & \text{caso contrário} \end{cases} \quad (4.23)$$

onde m_1 e m_2 são inteiros positivos de $h_1 = \frac{T_1}{m_1}$ e $h_2 = \frac{T_2}{m_2}$. Uma vez que as 2D-BPFs tomam somente um valor nestas subregiões, essas podem ser expressadas como blocos de funções de pulsos individuais em cada região, isto é,

$$\Phi_{i_1 i_2}(x, y) = \phi_{i_1}(x)\psi_{i_2}(y), \quad (4.24)$$

onde ϕ_{i_1} e ψ_{i_2} são blocos de funções de pulso unidimensionais (1D-BPFs) relatadas para variáveis x e y , respectivamente. As 2D-BPFs possuem algumas propriedades interessantes conhecidas como dissociação, ortogonalidade e completeza. As 2D-BPFs são dissociadas umas com as

outras quando

$$\Phi_{i_1 i_2}(x, y) \Phi_{j_1 j_2}(x, y) = \begin{cases} \Phi_{i_1 i_2}(x, y), & \text{para } i_1 = j_1 \text{ e } i_2 = j_2, \\ 0, & \text{caso contrário} \end{cases} \quad (4.25)$$

e ortogonais umas com as outras quando

$$\int_0^{T_1} \int_0^{T_2} \Phi_{i_1 i_2}(x, y) \Phi_{j_1 j_2}(x, y) dy dx = \begin{cases} h_1 h_2, & \text{para } i_1 = j_1 \text{ e } i_2 = j_2 \\ 0, & \text{caso contrário} \end{cases} \quad (4.26)$$

na região de $x \in [0, T_1)$ e $y \in [0, T_2)$, onde $i_1 = j_1 = 1, 2, \dots, m_1$ e $i_2 = j_2 = 1, 2, \dots, m_2$. A outra propriedade é a completeza, que para cada $f \in \mathcal{L}([0, T_1) \times [0, T_2))$ quando m_1 e m_2 vão para o infinito, a identidade de Parseval permanece

$$\int_0^{T_1} \int_0^{T_2} f^2(x, y) dy dx = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} f_{i_1 i_2}^2 \|\Phi_{i_1 i_2}(x, y)\|^2, \quad (4.27)$$

onde

$$f_{i_1 i_2} = \frac{1}{h_1 h_2} \int_0^{T_1} \int_0^{T_2} f(x, y) \Phi_{i_1 i_2}(x, y) dy dx. \quad (4.28)$$

O conjunto das 2D-BPFs pode ser escrito como um vetor $\varphi(x, y)$ de dimensão $m_1 m_2 \times 1$ como

$$\varphi(x, y) = \left[\Phi_{11}(x, y), \dots, \Phi_{1m_2}(x, y), \dots, \Phi_{m_1 1}(x, y), \dots, \Phi_{m_1 m_2}(x, y) \right]^T, \quad (4.29)$$

onde $\varphi(x, y)$ é um vetor coluna e lembramos que $(x, y) \in \mathcal{L}$. Das propriedades descritas para as 2D-BPFs, obtemos

$$\varphi(x, y) \varphi^T(x, y) = \begin{pmatrix} \Phi_{11}(x, y) & 0 & \cdots & 0 \\ 0 & \Phi_{12}(x, y) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_{m_1 m_2}(x, y) \end{pmatrix}, \quad (4.30)$$

$$\varphi^T(x, y) \varphi(x, y) = 1, \quad (4.31)$$

e

$$\varphi(x, y)\varphi^T(x, y)V = \tilde{V}\varphi(x, y), \quad (4.32)$$

onde V é um vetor de dimensão m_1m_2 e \tilde{V} é uma matriz diagonal contendo os elementos de V . Podemos concluir disso que para cada matriz A $(m_1m_2) \times (m_1m_2)$

$$\varphi^T(x, y)A\varphi(x, y) = \hat{A}^T\varphi(x, y), \quad (4.33)$$

onde \hat{A} é um vetor de dimensão m_1m_2 com elementos iguais aos da matriz diagonal A .

Uma outra propriedade interessante é escrever uma função em termos das 2D-BPFs. Supomos uma função $f(x, y)$ definida sobre $[0, T_1) \times [0, T_2)$ que pode ser expandida por 2D-BPFs como

$$f(x, y) \approx \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} f_{i_1i_2} \Phi_{i_1i_2}(x, y) = F^T \varphi(x, y) = \varphi^T(x, y)F, \quad (4.34)$$

onde F é um vetor $(m_1m_2 \times 1)$ dado por

$$F = \left[f_{11}, \dots, f_{1m_2}, \dots, f_{m_11}, \dots, f_{m_1m_2} \right]^T, \quad (4.35)$$

e $\varphi(x, y)$ definido em (4.29). Os coeficientes das 2D-BPFs são obtidos como

$$f_{i_1i_2} = \frac{1}{h_1h_2} \int_{(i_1-1)h_1}^{i_1h_1} \int_{(i_2-1)h_2}^{i_2h_2} f(x, y) \Phi_{i_1i_2}(x, y) dy dx, \quad (4.36)$$

tal que o erro médio quadrático entre $f(x, y)$ e a expansão é mínimo

$$\varepsilon = \frac{1}{T_1T_2} \int_0^{T_1} \int_0^{T_2} \left(f(x, y) - \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} f_{j_1j_2} \Phi_{j_1j_2}(x, y) \right)^2 dy dx. \quad (4.37)$$

Similarmente, uma função de quatro variáveis $K(x, y, s, t)$ em $[0, T_1) \times [0, T_2) \times [0, T_3) \times [0, T_4)$, pode ser aproximada com respeito as BPFs tal que

$$K(x, y, s, t) \simeq \varphi^T(x, y)\eta\Psi(s, t), \quad (4.38)$$

onde $\varphi(x, y)$ e $\Psi(s, t)$ são vetores de 2D-BPFs de dimensões m_1m_2 e m_3m_4 respectivamente, e η é uma matriz $(m_1m_2) \times (m_3m_4)$ contendo os coeficientes das 2D-BPFs. A equação (4.38)

é um ponto importante a ser discutido. Note que apesar do nosso núcleo não ser separável por coordenadas, a multiplicação dessas funções ao núcleo pode fazer com que esse passe a ser separável. Isso nos permite trabalhar com outros tipos de potenciais para descrever as interações ao longo da molécula dos que já são conhecidos na literatura do modelo. Assim, podemos trabalhar com equações não lineares e aproximar a função $f(x, y)$ como

$$[f(x, y)]^p = [F^T \varphi(x, y)]^p = \Lambda^T \varphi(x, y), \quad (4.39)$$

onde Λ é um vetor coluna cujo os elementos tem potência de p , e p é um número inteiro positivo. A integração do vetor φ definida em (4.29) pode ser aproximadamente obtida como

$$\begin{aligned} \int_0^x \int_0^y \varphi(\tau_1 \tau_2) d\tau_1 d\tau_2 &\simeq P\varphi(x, y) && \text{onde } x \in [0, T_1) \text{ e } y \in [0, T_2) \\ &= [E_{m_1 \times m_1} \otimes E_{m_2 \times m_2}] \varphi(x, y) \end{aligned} \quad (4.40)$$

onde P é a matriz operacional de integração para as 2D-BPFs e E a matriz operacional 1D-BPFs definida sobre $[0, T)$ com $h = T/m$ e $T = T_1 = T_2$ como segue

$$E = \frac{h}{2} \begin{pmatrix} 1 & 2 & 2 & \cdots & 2 \\ 0 & 1 & 2 & \cdots & 2 \\ 0 & 0 & 1 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (4.41)$$

Em (4.40), \otimes denota o produto de Kronecker definido como $A \otimes B = a_{ij}B$. Assim, a integral 2D para cada função $f(x, y)$ pode ser aproximada como

$$\begin{aligned} \int_0^x \int_0^y f(\tau_1, \tau_2) d\tau_1 d\tau_2 &\simeq \int_0^x \int_0^y F^T \varphi(\tau_1 \tau_2) d\tau_1 d\tau_2 \\ &\simeq F^T P \varphi(x, y). \end{aligned} \quad (4.42)$$

De posse dessa teoria, tentaremos aplicar esses conceitos na equação de Fredholm de segundo tipo a um núcleo similar ao modelo original.

4.3.2 Aplicação das 2D-BPFs à núcleos similares ao modelo Peyrard-Bishop original

Nesta seção vamos demonstrar que os blocos de funções de pulso podem ser empregadas ao modelo Peyrard-Bishop, utilizando como exemplo, uma sequência homogênea como no modelo original. Iremos aplicar as definições e propriedades acima a uma equação de Fredholm de segundo tipo, em um núcleo similar ao modelo original. Começamos tomando um núcleo genérico $K(x, y, s, t)$ a princípio, não separável para as variáveis x, y, s e t . O operador integral de transferência para um núcleo $K(x, y, s, t)$, possui a forma

$$\lambda f(x, y) = \int_0^{T_3} \int_0^{T_4} K(x, y, s, t) u(s, t) ds dt. \quad (4.43)$$

Tomaremos valores aproximados para as funções $f(x, y)$, $u(x, y)$ e para o núcleo $K(x, y, s, t)$:

$$\begin{aligned} f(x, y) &= F^T \varphi(x, y); \\ u(x, y) &= \varphi^T(x, y) \Lambda; \\ K(x, y, s, t) &= \varphi^T(x, y) \eta \varphi(s, t). \end{aligned} \quad (4.44)$$

Substituindo esses termos na equação (4.43) temos,

$$\begin{aligned} \lambda F^T \varphi(x, y) &= \int_0^{T_3} \int_0^{T_4} \varphi^T(x, y) \eta \varphi(s, t) \varphi^T(s, t) \Lambda ds dt, \\ &= \varphi^T(x, y) \eta \int_0^{T_3} \int_0^{T_4} \varphi(s, t) \varphi^T(s, t) ds dt \Lambda, \\ &= \varphi^T(x, y) \eta \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \int dx dy \Phi_{ij}(x, y) \Lambda, \\ &= h_1 h_2 \varphi^T(x, y) \eta \Lambda. \end{aligned} \quad (4.45)$$

Usando a propriedade que $F^T \varphi(x, y) = \varphi^T(x, y) F$ de acordo com a equação (4.34), os autovalores são escritos como

$$\lambda = h_1 h_2 \eta \Lambda F^{-1}. \quad (4.46)$$

Dado o formalismo de ensemble canônico no modelo Peyrard-Bishop, aplicando essa propriedade do operador IT dada pelo o formalismo das BPFs no núcleo da função canônica de partição,

obtemos

$$\begin{aligned}
Z &= \int \prod_n dx_n dy_n K(x_n, x_{n-1}, y_n, y_{n-1}), \tag{4.47} \\
&= \int dx_1 dy_1 \cdots dx_n dy_n K(x_1, x_0, y_1, y_0) K(x_2, x_1, y_2, y_1) \cdots K(x_n, x_{n-1}, y_n, y_{n-1}), \\
&= \int dx_1 dy_1 \cdots dx_n dy_n K(x_0, y_0, x_1, y_1) K(x_1, y_1, x_2, y_2) \cdots K(x_{n-1}, y_{n-1}, x_n, y_n), \\
&= \int dx_1 dy_1 \cdots dx_n dy_n \varphi^T(x_0, y_0) \eta_1 \varphi(x_1, y_1) \cdots \varphi^T(x_{n-1}, y_{n-1}) \eta_n \varphi(x_n, y_n).
\end{aligned}$$

Utilizamos aqui a condição de contorno onde o último par de base esta ligado com o primeiro par de base em uma cadeia contendo N pares de base como no modelo original [6], obtemos a função de partição da forma

$$\begin{aligned}
Z &= \int dx_0 dy_0 dx_1 dy_1 \cdots dx_N dy_N \varphi^T(x_0, y_0) \eta_1 \varphi(x_1, y_1) \varphi^T(x_1, y_1) \eta_2 \varphi(x_2, y_2) \cdots \times \tag{4.48} \\
&\quad \times \varphi^T(x_{N-1}, y_{N-1}) \eta_N \varphi(x_N, y_N) \delta(x_N - x_0) \delta(y_N - y_0), \\
&= \int dx_1 dy_1 \cdots dx_N dy_N \varphi^T(x_N, y_N) \eta_1 \varphi(x_1, y_1) \varphi^T(x_1, y_1) \eta_2 \varphi(x_2, y_2) \cdots \times \\
&\quad \times \varphi^T(x_{N-1}, y_{N-1}) \eta_N \varphi(x_N, y_N), \\
&= \int dx_N dy_N \varphi^T(x_N, y_N) \eta_1 \int dx_1 dy_1 \varphi(x_1, y_1) \varphi^T(x_1, y_1) \eta_2 \int dx_2 dy_2 \varphi(x_2, y_2) \varphi^T(x_2, y_2) \\
&\quad \times \cdots \times \eta_{N-1} \int dx_{N-1} dy_{N-1} \varphi(x_{N-1}, y_{N-1}) \varphi^T(x_{N-1}, y_{N-1}) \eta_N \varphi(x_N, y_N), \\
&= \int dx_N dy_N \varphi^T(x_N, y_N) \eta_1 \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \int dx_1 dy_1 \Phi_{ij}(x_1, y_1) \times \eta_2 \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \int dx_2 dy_2 \Phi_{ij}(x_2, y_2) \times \cdots \times \\
&\quad \eta_{N-1} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \int dx_{N-1} dy_{N-1} \Phi_{ij}(x_{N-1}, y_{N-1}) \eta_N \varphi(x_N, y_N), \\
&= (h_1 h_2)^{N-1} \int dx_N dy_N \varphi^T(x_N, y_N) \eta_1 \times \eta_2 \times \cdots \times \eta_{N-1} \times \eta_N \varphi(x_N, y_N).
\end{aligned}$$

Até agora aplicamos a técnica para um núcleo qualquer sem fazer quaisquer restrições ao núcleo estudado. Desde que definida as funções de blocos utilizadas, o formalismo de BPFs se aplica a qualquer situação. No caso de termos uma cadeia homogênea composta somente de bases CG como no artigo do modelo original 6, através de uma integração do tipo Gauss-Legendre, o nosso espaço bidimensional fica definido na região $\mathcal{L} = [0, 1) \times [0, 1)$ levando à $h_1 = h_2 = 1/m$.

Nossa função de partição então se resume a forma final

$$Z = \left(\frac{1}{m^2}\right)^N \int dx_N dy_N \varphi^T(x_N, y_N) \eta^N \varphi(x_N, y_N). \quad (4.49)$$

Concluimos que as 2D-BPFs podem ser aplicadas ao modelo Peyrard-Bishop como um método alternativo ao modelo Peyrard-Bishop ao original. A mesma técnica também pode ser aplicada a núcleos não separáveis como também, possibilita o uso de outros potenciais no modelo PB. Entretanto, a escolha das BPFs a se utilizar é a chave para a solução do problema, pois elas aproximam ou tornam o núcleo tratável. Além disso, as BPFs escolhidas precisam ser bem definidas para a implementação numérica. A escolha da BPF utilizada para a solução da equação (4.22) não foi possível ainda. A função precisa satisfazer as propriedades apresentadas para o nosso núcleo, e isso não tem sido uma tarefa fácil. Precisa-se também de um melhor entendimento da nossa parte para a solução da equação (4.22) numericamente, mas sem a BPF certa, não é possível pensar no método numérico a se aplicar.

4.4 Conclusões

Concluimos que a utilização de monômeros assimétricos para a molécula de DNA foi possível com a utilização de um potencial harmônico, não necessitando a premissa de considerar um *stacking* uniforme no modelo. Os valores dos parâmetros foram consistentes com cálculos já feitos utilizando o modelo PB e no caso das constantes de empilhamento, notamos uma dependência desses parâmetros com a concentração de sal. Esses resultados podem ser aplicados a qualquer duplexo. O resultado possibilita o estudo de diferentes estruturas com o modelo Peyrard-Bishop, como híbridos do tipo DNA/RNA [22], dupla hélices contendo *bulges* [1], *hairpins* [31], *loops* [61] e *LNAs* [62].

Concluimos também que com a utilização do potencial anarmônico de Dauxois *et al.* [5] no Hamiltoniano do modelo, torna o núcleo da técnica IT não separável, não tornando possível a solução por métodos padrões. Mostramos que o uso de 2D-BPFs se mostra um caminho promissor na solução desse tipo de problema e emprego de outros potenciais que os já adotados pelo modelo. A propriedade (4.38) parece não restringir a classe de núcleos que

se pode utilizar com essas funções. Isso enriquece ainda mais o modelo, o que pode permitir uma melhor descrição do fenômeno. Pretendemos no futuro fazer um teste numérico utilizando técnicas de solução de integrais convencionais para a função de partição, e vamos comparar os erros relativos ao método que empregamos. Talvez seja viável uma técnica de integração para esses tipos de potenciais em cadeias não muito grandes. Entretanto, isso é algo que precisa ser melhor entendido para a aplicação numérica.

Contudo, seguiremos a tese utilizando potenciais harmônicos. O mesmo são consistentes numericamente e contribuem da mesma forma que o potencial descrito por Dauxois *et al.* [5], como já mostrado por Zhang *et al.* [44]. Empregaremos esse resultado a seguir em estruturas assimétricas de RNAs contendo *bulges*.

5 Modelo mesoscópico assimétrico para o estudo de *single-bulges* em RNAs

5.1 Introdução

Neste capítulo, apresentaremos os resultados obtidos através do uso de um Hamiltoniano assimétrico no estudo de estruturas secundárias de RNA contendo *single-bulges*. Como vimos no capítulo 4, obtivemos as condições necessárias e a determinação do Hamiltoniano 4.1 no modelo Peyrard-Bishop para o estudo de estruturas assimétricas, e como mostrado, foi possível manter o formalismo numérico na solução da função de partição do problema (técnica de integral de transferência-IT) no caso de potenciais harmônicos. Aplicaremos o Hamiltoniano assimétrico 4.1 ao problema de estruturas de RNA contendo protuberâncias simples.

Começaremos na seção 5.2 definindo o que são protuberâncias e protuberâncias simples em duplexos, suas propriedades biológicas e a nomenclatura utilizada nesses tipos de moléculas. Aqui, relataremos também os experimentos raios-X e ressonância magnética nuclear (do inglês nuclear magnetic resonance-NMR) feitos nessas moléculas na determinação de suas estruturas e os resultados obtidos pelos experimentos com relação a identidade desses *bulges*.

Na seção 5.3, daremos foco a estruturas do grupo I de *single bulges* em RNA. Definiremos uma pseudo-base X que relata a posição do *bulge* no duplexo descrito pelo modelo. Em seguida, será apresentada a notação utilizada por nós na definição dos parâmetros do modelo associados aos vizinhos próximos ao *bulge* e também a minimização para esses parâmetros. Apresentamos os cálculos para os parâmetros de Morse que descrevem as pontes de hidrogênio para cada tipo de protuberância e sua identidade, como também os parâmetros de *stacking* associados aos primeiros e segundos vizinhos do mesmo.

Por fim, apresentamos os perfis de aberturas para sequências fazendo um paralelo com os experimentos de NMR já realizados na literatura e também as conclusões chegadas do trabalho aqui realizado.

5.2 Bulges em RNAs

As moléculas de RNA formam complexas estruturas terciárias que são intimamente relacionadas com suas funções biológicas dentro de organismos vivos. Em adição a duplas hélices regulares, diferentes conformações podem ser vistas nas estruturas de RNA como *mismatches*, *bulges*, *loops* e *hairpins*, algumas dessas estruturas já mostradas na figura 4.1 na página 43. O entendimento da formação dessas estruturas como também suas propriedades físicas associadas a esse tipo de configuração relacionadas as suas atividades biológicas, nos serve de motivação ao desenvolver o trabalho. No caso de *bulges*, objeto de nosso estudo, eles são formados em regiões na dupla hélice de RNA onde uma ou mais bases nitrogenadas não estão pareadas como em uma dupla hélice regular. Esse tipo de configuração gera resíduos de nucleotídeos dentro da estrutura de duplexos, isto é, as bases “sobram” dentro dessas duplas hélices, o que leva a uma quebra de simetria estrutural quando comparada a duplexos canônicos.

Os *bulges* estão entre os tipos de configurações mais comuns no caso de pareamentos não-Watson-Crick. Eles estão envolvidos no dobramento de estruturas de RNAs fornecendo a flexibilidade necessária para alcançar as conformações ativas, gerando estruturas terciárias. O enovelamento dessas estruturas de RNAs é conhecido como *RNA folding*. *Bulges* também estão envolvidos em interações RNA-RNA, são ligantes de RNAs [63, 64] e sítios de reconhecimento semelhante a alças moleculares para máquinas ativas como proteínas [65, 66]. São as torções entre as interações de açúcar-fosfato com os vizinhos que esses *bulges* causam no duplexo, que expõem suas bases nitrogenadas as ligações externas como mostrado na figura 5.1b. Nucleotídeos pareados nesta configuração facilitam o acontecimento de processos biológicos no qual os *bulges* estão sujeitos.

Uma única base não pareada em uma das fitas do duplexo forma o menor *bulge* possível, conhecido na literatura como *single-bulge*. A figura 5.1 mostra esse tipo de protuberância para diferentes duplexos de RNA (figuras 5.1a e 5.1b) e em estruturas mais complexas, que nesse caso, do tipo *hairpin* (figuras 5.1c e 5.1d). Nesse trabalho, daremos foco no estudo duplexos contendo *single-bulges* em RNAs com diferentes arranjos de base em suas estruturas. Os *single-bulges* ocorrem naturalmente em RNA, frequentemente vistos em RNA ribossomal [1]. Entretanto, esses mesmos tipos de *bulges* podem ser sintetizados pelo agrupamento de dois

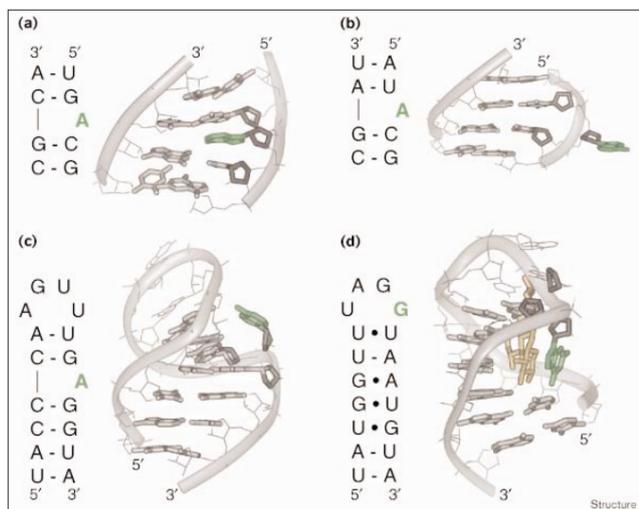


Figura 5.1

Exemplos de *bulges* em fitas duplas e *hairpins* de RNA. Note neste exemplo que a base relativa ao *bulge* pode estar inserida na fita dupla ou virada para fora (*base-flipping*). Figura da referência 65.

monômeros de diferentes comprimentos onde apenas uma base não é complementar. A formação de *single-bulges* depende da configuração de pares de base contida na sequência da dupla hélice. Essa configuração das sequências contidas dado os pares de base são determinantes na conformação e identidade das protuberâncias na região da dupla hélice [67].

As configurações de *single-bulges* são as mais diversas possíveis. Seguindo o pareamento Watson-Crick para RNA, podemos ter *single-bulges* para as quatro bases nitrogenadas conhecidas: *single-bulges adenosina*, *citossina*, *guanossina* e *uridina* correspondendo as bases não pareadas A, C, G e U, respectivamente. A identidade do *bulge* e seus vizinhos, influenciam na configuração que a estrutura assume. No artigo de Hermann e Patel [65] um *bulge adenosina*, conhecido como *base-flipping*, parece interagir muito fracamente com os seus vizinhos tendo então uma maior liberdade para rodar para fora do plano da dupla hélice como mostrado na figura 5.1b. Entretanto, a mesma base com diferentes vizinhos se mantém no plano da dupla hélice como mostrado na figura 5.1a e em trabalhos experimentais utilizando medidas em NMR [68, 69]. É conhecido na literatura de *bulges* que os vizinhos influenciam na forma e tamanho do *bulge* pelas interações que compõem essas estruturas [69].

Barthel e Zacharias mostraram através de um cálculo de dinâmica molecular que *bulges uridina* são mais propensos a ter a base fora do plano da dupla hélice que *bulges adenosina* [70]. Esse resultado é confirmado pelos experimentos de NMR feito por Popenda *et al.* [69]. O tamanho dos *bulges* nas estruturas secundárias de RNA influenciam na estabilidade da molécula como mostrado por Tinoco *et al.* [67]. Portanto, a quantificação dessas interações nos sulcos dos duplexos contendo *bulges* é um caminho que nos levará a um melhor entendi-

mento das estruturas secundárias e terciárias dessas molécula e conseqüentemente, seu papel nos processos biológicos em que estão envolvidos.

A dependência dos vizinhos ao redor do par de base na literatura de *bulges* define grupos de seqüências. Existem quatro grupos que descrevem os tipos de *single-bulges* encontrados para a base não pareada [1, 2]. O primeiro grupo que descreveremos aqui é o Grupo-I. Este grupo de *single-bulges* é composto por seqüências onde a base não pareada em um dos monômeros que compõem a dupla hélice é diferente das primeiras bases vizinhas ao seu redor. Isso faz com que a posição do *bulge* seja completamente conhecida. Por exemplo, suponhamos uma seqüência de comprimento 11 na direção 5' → 3' UGACGAUGCUA e sua fita complementar de comprimento 10 na direção 3' → 5' ACUGCACGAU. O melhor alinhamento das duas seqüências é



onde apenas o A central não tem um correspondente na seqüência complementar.

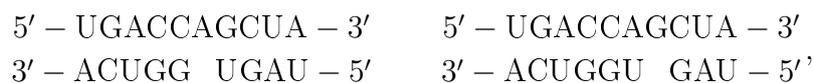
O Grupo-II de *single-bulges* é composto por seqüências onde a base não pareada é igual a pelo menos uma de suas primeiras vizinhas ao longo do monômero que está ligada. Por esta razão a localização do *bulge* não é bem definida. Como fizemos anteriormente para o primeiro grupo, vamos supor uma seqüência de comprimento 11 na direção 5' → 3' UGACAAAGCUA e sua fita complementar de comprimento 10 na direção 3' → 5' ACUGUUCGAU. Diferente do Grupo I, ao alinharmos as duas seqüências temos três possibilidades de complementaridade



Vemos então que não é possível determinar a localização exata do *bulge* apenas com as regras de complementariedade. Os experimentos utilizando seqüências com *bulges* no cálculo de temperatura de melting, não conseguem falar ao certo a localização das bases não pareadas. Eles simplesmente destacam a região em que os *single-bulges* podem estar. Não sabemos se *bulges* dessa natureza possuem somente uma configuração possível para a posição, isto é, pode-se ter a probabilidade de se formar o mesmo tipo de *bulge* em uma outra posição, levando a um outro tipo de folding do RNA também com uma função biológica ativa. O entendimento da locali-

zação do *bulge* desse grupo é fundamental e buscamos com esse trabalho dar o primeiro passo na determinação da base resíduo e se possível dizer as probabilidades para cada configuração. Entretanto, isso não será feito nesse trabalho. A tese é um primeiro passo para chegarmos a esse resultado.

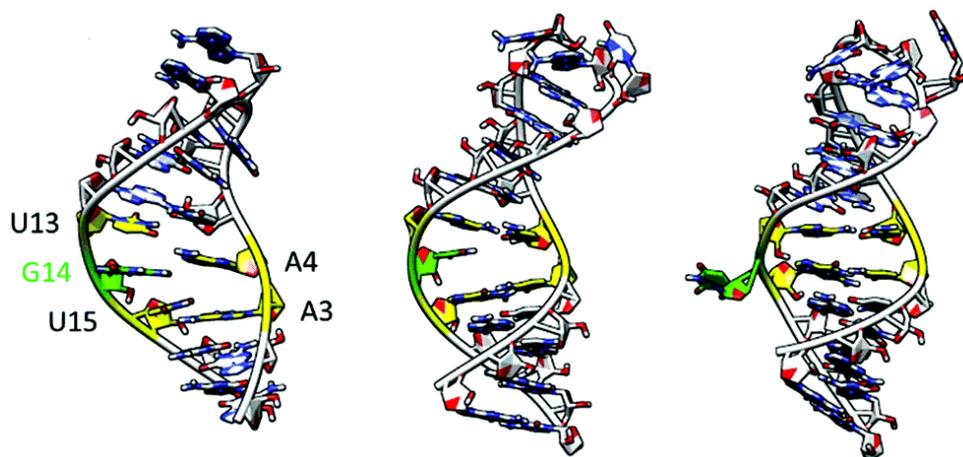
Os Grupos III e IV são *single-bulges* que contêm em suas sequências pares de base GU, ou seja, não respeitam o pareamento Watson-Crick canônico. O Grupo-III é formado por *single-bulges* similar ao Grupo-I, isto é, sem primeiros vizinhos iguais a base resíduo. Entretanto, estes possuem pelo menos um par de base GU ao longo de sua dupla fita e um desses pares de base GU é primeiro vizinho do *bulge*. Isso também gera uma ambiguidade na localização do *bulge*, pois uma ligação do tipo GU em RNA é tão provável quanto o pareamento canônico [20]. Assim, se tomarmos uma sequência de comprimento 11 na direção $5' \rightarrow 3'$ UGACCAGCUA e sua fita complementar de comprimento 10 na direção $3' \rightarrow 5'$ ACUGGUGAU, teremos duas possibilidades de configuração para a sequência



levando também a troca da identidade do *bulge*.

O Grupo-IV é análogo ao Grupo-II, ou seja, possui um vizinho da mesma natureza que a base resíduo. Nesse grupo, tem-se também pelo menos um par de base GU junto ao par de base não pareado. Esse grupo aumenta ainda mais o número de configurações possíveis que a sequência pode assumir. O par de base GU próximo a base não pareada ou a vizinha de mesma natureza próxima ao *bulge*, aumenta o número de configurações possíveis que a sequência pode assumir, podendo até afetar seus segundos vizinhos. O Grupo-IV se mostra um problema ainda maior em comparação aos outros grupos, pois o aumento do número de configurações, dificulta ainda mais a localização do *bulge* na sequência.

Como dito anteriormente, os grupos e seus vizinhos associados as sequências são importantes para o tamanho do *bulge* no sulco do duplexo. Do ponto de vista estrutural, *single-bulges* podem ser encontrados em diferentes configurações, como empilhados dentro da hélice (*intercalated*), fora do plano da molécula (*flipped out* ou *extrahelical*) ou inclinado interagindo com um dos seus vizinhos (*side-by-side*) em um dos sulcos da dupla hélice [65, 68, 71, 72].

**Figura 5.2**

Configurações possíveis para um *single-bulge* guanósina em uma molécula de RNA: a) *side by side*; b) *intercalated*; e c) *extrahelical*.

Podemos ver essas configurações na figura 5.2 para *single-bulges* guanósinas.

Qualquer uma dessas configurações compostas na molécula causam um decréscimo na temperatura de desnaturação da mesma [73], pois se tratam de configurações instáveis do ponto de vista da desnaturação, quando comparadas a estrutura canônica. A palavra instável aqui foi usada no sentido de que o bulge representa a ausência de interação na região da hélice do RNA, logo esse terá uma temperatura de *melting* menor comparada a uma estrutura canônica. Em particular, White e Draper [74, 75] estudaram os efeitos de intercalação de *single-bulges* em RNAs e mostraram que a inclusão de *bulges* tem um efeito dramático nas possibilidades de conformações da hélice e esses mesmos efeitos foram vistos em DNAs [76]. *Bulges* adenosinas flanqueados por pares de base mismatches GA foram cristalizados em HIV-1 RNA por Ennifar *et al.* [77] e estudado com difrações de raio-X revelando uma significativa assimetria no sulco do duplexo. Em alguns casos, *bulges* podem assumir duas diferentes configurações, por exemplo um *single-bulge* citosina pode assumir dois tipos diferentes de configurações extrahelical [78]. Para *bulges* uridinas, em um estudo de raio-X foi mostrado que o *bulge* pode se manter ou não em certas configurações em um menor sulco [79]. Similarmente, *bulges* em DNAs também desestabilizam a hélice [80–83].

Popenda *et al.* investigou as diferenças estruturais entre uma hélices regulares de RNAs comparadas as estruturas contendo um *single-bulge* adenosina do Grupo-II [68] e estruturas do Grupo-I para *single-bulges* adenosinas e uridinas [69] utilizando NMR. Os resultados apresentados por Popenda *et al.* [68] mostram que *single-bulges* adenosinas tende a ter uma conformação onde a base fica flanqueada no plano da molécula. O estudo de NMR realizado por Popenda *et al.* [68] mostra a posição correta para uma sequência do Grupo-II, confirmando

que a base resíduo é a adenina A5 para a sequência estudada analisando a assinatura do grupo imina da uracila [68]. Em seu trabalho no ano seguinte, Popena *et al.* [69] faz uma comparação entre sequências com identidades diferentes, utilizando sequências contendo *single-bulges adenosinas* e *uridinas*. Popena *et al.* [69] mostraram nesse trabalho que a sequência que continha o *single-bulge adenosina* se manteve no plano do duplexo e interagiu com um dos grupos CGs vizinhos (C14-G4:A5), o que leva a conclusão de uma base tripla na região caracterizando a conformação *side-by-side*. No caso da sequência contendo o *single-bulge uridina*, as medidas levaram a conclusão de que o *bulge* se encontra na configuração *flipped out*. Configuração que é característica para esse tipo de base como resíduo já argumentado por Barthel e Zacharias [70].

Experimentos de difração de raios x e espectroscopia NMR são os principais métodos para a avaliação de detalhes estruturais e interações em RNAs. Entretanto, discrepâncias aparecem na determinação estrutural de *bulges adenosinas* ao comparar os resultados de difração e NMR, como também a conformação que a base assume. Em estudos de NMR, o *bulge adenosina* é geralmente encontrado em alguma posição no plano da molécula (*intercalated* ou *side-by-side*), enquanto estruturas de difração de raios x, usualmente obtém-se *bulges* na conformação *extrahelical* em altas concentrações de sal [84]. Esperamos obter alguma informação adicional que ajude a resolver esse tipo de ambiguidade nos resultados utilizando um modelo mesoscópico.

A estabilidade termodinâmica de *bulges* na molécula são analisadas por medidas de eletroforese [85] e experimento de temperaturas de meting [1–4]. Uma importante conclusão desses estudos é que o modelo NN não parece ser o mais adequado para lidar com perturbações estruturais da dupla hélice [86]. Conclusões similares foram encontradas para estruturas de DNA contendo *bulges* [83, 87]. Cálculos com temperaturas de melting para *single-bulges* em *hairpins* de RNA [88–90], como também para *bulges* de ordem maiores [91], apresentam discrepâncias ainda mais consideráveis. Isso se deve a contagem de energia livre dessas estruturas que o modelo de primeiros vizinhos realiza. Os algoritmos baseados no modelo NN atribuem uma penalidade de 3.8 kcal/mol para cada *bulge* contido no duplexo [1]. Mas ainda assim esse modelo apresenta algumas limitações, como por exemplo, esse assume que as interações de empilhamento próximas ao *bulge* não sofrem perturbações na presença do *bulge*, o que não

condiz com a realidade. O modelo de primeiros vizinhos também não conta qualquer mudança na estabilidade do *bulge* devido a identidade da base resíduo não pareada. Como exemplo, um *bulge pirimidina* possui a mesma penalidade na conta da energia livre que um *bulge purina*.

Apesar de haver um grande número de trabalhos usando o modelo de primeiros vizinhos para predições de energia livre de duplexos de RNA contendo *bulges* [67, 86, 92–97], nenhuma relação foi encontrada entre a termodinâmica dos pares de base e a conformação do *bulge*. Métodos alternativos tem sido utilizados por Znosko *et al.* [1] como também por Blose *et al.* [2]. Por essas razões, o modelo de primeiros vizinhos tem se mostrado um modelo não adequado para análise termodinâmica desse tipo de estrutura. Acreditamos que a utilização do modelo Peyrard-Bishop pode fazer essa correlação entre a termodinâmica e suas propriedades estruturais, pois se trata de um modelo mais completo por descrever suas interações, quando comparado ao modelo de primeiros vizinhos. Notamos também que o *bulge* causa uma assimetria no duplexo algo também não levada em conta pelo modelo NN. O modelo mesoscópico PB se mostra uma ferramenta alternativa e útil no estudo desse tipo de estruturas.

Modelos mesoscópicos utilizam aproximações mais simplificadas para as interações moleculares nas estruturas comparados a modelos de dinâmica molecular, como modelos *coarse-grain*, onde esses são mais eficientes nas predições de estruturas secundárias [98]. Por outro lado, o grau de eficiência dado por modelos *coarse-grain* gera um alto custo computacional, o que o torna impraticável em algumas situações. Assim, os modelos mesoscópicos são alternativas interessantes, pois mesmo com suas limitações, conseguem resultados para estruturas mais complexas descrevendo boa parte das informações contidas nas estruturas [20, 99]. Assim, o modelo PB se torna útil para uma primeira avaliação e o descreveremos a seguir na próxima seção.

5.3 Modelo PB no estudo de *bulges*

No capítulo 3, vimos que o modelo Peyrard-Bishop é um importante modelo estatístico utilizado no estudo de duplexos e no capítulo 4 exploramos a condição de assimetria que

o modelo pode assumir. A função de partição resultante da assimetria do modelo, nos leva a

$$Z_y = \prod_{n=1}^N \int dy_n \exp \left\{ -\beta \left[\frac{1}{2} \gamma (y_n - y_{n-1})^2 + V(y_n) \right] \right\}, \quad (5.1)$$

onde temos que $\frac{1}{\gamma} = \frac{1}{k_u} + \frac{1}{k_v}$ faz a associação em série dos nucleotídeos que compõe o duplexo. Precisamos identificar o *bulge* no duplexo em nosso estudo.

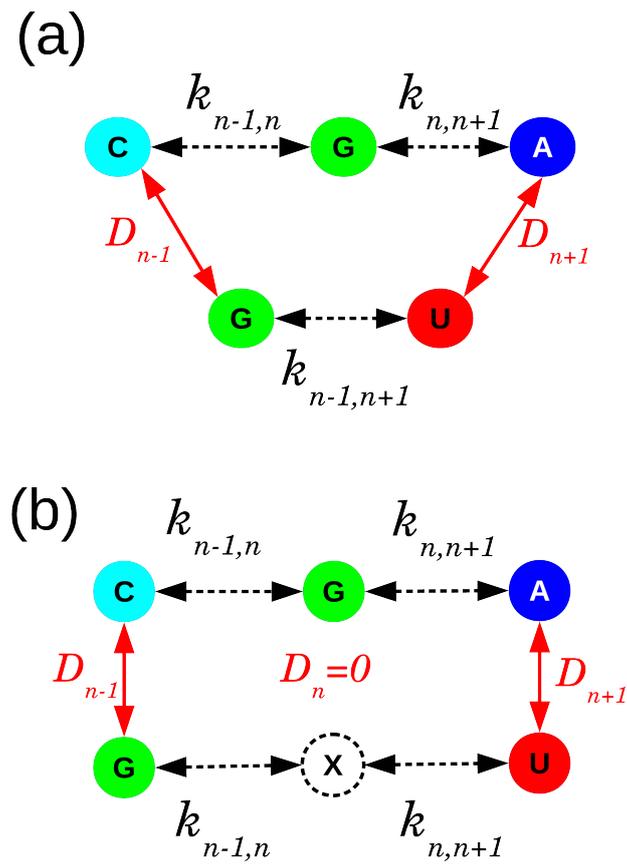
O modelo PB não leva em conta duplexos com monômeros de diferentes tamanhos, mas podemos superar essa dificuldade introduzindo uma pseudo-base na direção oposta ao *bulge*, como mostrado na figura 5.3. Usamos o carácter X, representando a pseudo-base, colocada na fita mais curta preenchendo visualmente a vaga, como no exemplo a seguir de uma projeção de um *bulge guanosina*



Iremos nos referir a base resíduo formada por *bulges* pela notação BX com B=A, C, G ou U. Em nosso exemplo da figura 5.3 temos um *bulge* do tipo GX.

O potencial de Morse dado da interação entre a pseudo-base e o *bulge* é fixo e tomado como nulo, isto é, $D = 0$ eV como mostrado esquematicamente na figura 5.3. Entretanto, o mesmo não pode ser feito para o *stacking* k entre o *single-bulge* e as bases vizinhas. Efetivamente, a inserção da pseudo-base X entre as duas bases na fita oposta ($3' \rightarrow 5'$), faz com que a interação de *stacking* na direção se divida em duas partes. Para entender isto, vamos usar a notação do Hamiltoniano da equação (4.1) e assumir por exemplo, o *bulge* localizado na n -ésima posição na fita u ($5' \rightarrow 3'$) e a base X na n -ésima posição da fita v ($3' \rightarrow 5'$). A interação das bases vizinhas em $n - 1$ e $n + 1$, será dada pelos termos $k_{v;n-1,n}$ e $k_{v;n,n+1}$ que representa a constante $k_{v;n-1,n+1}$ na situação do *bulge* como mostrado na figura 5.3.

Em nosso método de minimização que descreverei a seguir na seção 5.5, iremos permitir que a base canônica BP assumam diferentes valores próximas ao *bulge* B. Para representar essa não ambiguidade, escrevemos essa como BP^B . Por exemplo, dado o par de base CG próximo a um *bulge adenosina* AX, as interações moléculares representadas por essa base serão descritas por CG^A , como adotado na referência 20. Portanto, o par de base CG^A é o par de base

**Figura 5.3**

Identificação de um *bulge* pelo modelo Peyrard-Bishop. Na figura (a) mostramos um exemplo de um *bulge* guanósina entre os pares de base CG e AU. Na figura (b) mostramos a introdução de uma pseudo-base X representando a falta de pareamento da base guanósina na direção oposta do duplexo, isto é, $D = 0$ eV.

com variáveis próxima ao *bulge* A, enquanto a base CG sem índice tem parâmetros constantes e não é flanqueada ao *bulge*. A notação aqui mostrada não carrega o índice que o *bulge* está localizado na direção $5' \rightarrow 3'$. Esse tipo de identificação sobrecarregaria a notação com um índice a mais. Todos os *bulges* contidos na sequências aqui utilizadas se encontram na direção $5' \rightarrow 3'$.

Como consequência desta notação, iremos lidar com três tipos de interações entre as bases em cada sequência:

Vizinhos mais próximos fixos - Fixed nearest-neighbors (FNN) são formados por pares de base AU ou CG longe de qualquer *bulge*. Quando nos referirmos a essas interações, estamos nos referindo ao pareamento depois dos segundos vizinhos contidos na sequência. Estes serão identificados da forma BPpBP. Por exemplo, AU_pCG é um par de base AU seguido por um par de base CG, não sendo vizinho ao *bulge*. Esses parâmetros não serão modificados durante o procedimento de minimização e assumiram os valores típicos para RNA canônicos.

Vizinhos mais próximos variáveis - Variable nearest-neighbors (VNN) são formados por dois pares de base AU e CG, onde um deles é flanqueado ao *bulge*. Estes parâmetros serão escritos como BP^BpBP ou BPpBP^B dependendo de qual base BP está próxima ao *bulge*. Os parâmetros VNNs estão contidos no conjunto de parâmetros que irá variar durante a minimização.

Vizinhos mais próximos ao *bulge* - Bulge nearest-neighbors (BNN) são parâmetros relacionados ao pareamento do *bulge* com a pseudo-base e um par de base vizinho podendo ser AU ou CG. Estes parâmetros serão escritos genericamente como BP^BpBX ou BXpBP^B. Em nosso exemplo no começo desta seção onde temos um *bulge* *guanósina* GX (figura 5.3), os parâmetros relacionados a esse *bulge* UA^GpGX a esquerda do *bulge* e GXpCG^G a direita do *bulge*.

Nossa estratégia na modelagem é permitir que os pares de base flanqueados ao *bulge* do tipo BP^B varie seu potencial de Morse D , enquanto os pares de base remanescentes CG e AU irão manter seu potencial constante. Portanto, todos pares de base do tipo BP^B irão

ter suas pontes de hidrogênio alteradas devido a presença do *bulge*. Iremos também variar as interações de *stacking* dos VNNs como também as dos BNNs, mas manteremos o *stacking* dos FNNs fixos com os valores canônicos.

5.4 Conjunto de sequências e suas respectivas temperaturas de melting

Nós localizamos um total de 80 sequências de *single-bulges* do Grupo I das referências 1–4, onde temos 18 *bulges adenosinas* (AX) como mostrado na tabela 5.1, 12 *citiosinas* (CX) contidos na tabela 5.2, 24 *guanossinas* (GX) escritas na tabela 5.3 e 26 *bulges uridinas* (UX) escritos na tabela 5.4. Note que Kent *et al.* [4] inicialmente descreve suas sequências como do Grupo III, entretanto a análise de energia livre realizada neste artigo confirma que as sequências pertencem de fato ao Grupo I.

As temperaturas de melting estavam na concentração de 100 μM . Recalculamos a concentração de duplexos para 200 μM a partir de suas respectivas variações de entropia e entalpias, isto é, ΔH e ΔS . [1–4]. Isto é feito como no caso de duplexos canônicos de RNA, devido a uma perda de energia conformacional na formação do duplexo a concentrações mais baixas [33]. A lista completa das sequências como também suas temperaturas de melting reajustadas para a nova concentração se encontram nas tabelas 5.1, 5.2, 5.3, e 5.4.

5.5 Otimização dos parâmetros

Utilizamos a técnica de equivalência térmica já introduzida no capítulo 3 para otimizar os parâmetros do potencial apresentado pela equação (3.22),

$$\chi^2 = \sum_{i=1}^N (T'_i - T_i)^2. \quad (5.2)$$

Como feito no caso de um DNA canônico [55], nós expandimos a função de partição 5.1 em termos não-diagonais da matriz de transferência. O índice de melting τ_i para cada sequência i é o termo de maior ordem da expansão e mede o quanto a sequência difere de um polímero de RNA

Sequências	T_i (°C)	T'_i (°C)	τ
GAGACAC/CUCXGUG	31.773	32.0594	2.09141
CGCACGC/GCGXGCG	40.5337	38.7617	2.21856
GAUAGAC/CUAXCUG	18.6312	18.8269	1.84038
GCGAGCG/CGCXCGC	40.5915	41.0442	2.26186
GUGCAUGAG/CACGXACUC	45.7818	45.8193	2.35245
GGCGACUCG/CCGCXGAGC	57.6588	57.5239	2.57449
GGCGAUUCG/CCGCXAAGC	47.2622	48.2162	2.39792
GGCGAUUCC/CCGCXAAGG	52.2388	51.4717	2.45967
GCACAGAGG/CGUGXCUC	51.9741	51.8996	2.46779
UGACACUCA/ACUGXGAGU	42.8814	43.6834	2.31193
CAGUAGAGC/GUCAXCUCG	42.0866	42.0999	2.28189
CAGUACAGC/GUCAXGUCG	41.7388	41.7992	2.27618
GAGCAGGUC/CUCGXCCAG	52.8855	53.65	2.501
GACUAUGUC/CUGAXACAG	36.7729	36.9177	2.18358
UGAGAGUCA/ACUCXCAGU	46.0243	45.1227	2.33923
GACCAUGUC/CUGGXACAG	47.6505	47.6544	2.38726
CCAUUACUACC/GGUAAXGAUGG	51.5794	51.5145	2.46049

Tabela 5.1

Sequências usadas para *bulges* adenosinas (AX).

Sequências	T_i (°C)	T'_i (°C)	τ
GUGCUUC/CACXAAG	25.9529	26.136	2.03765
GACGCUAGC/CUGCXAUCG	49.1605	49.3768	2.38328
UGAGCGUCA/ACUCXCAGU	47.1622	47.1179	2.34968
UCCUCGAAC/AGGAXCUUG	40.1201	39.9587	2.24321
GCAUCUGUG/CGUAXACAC	36.3495	36.7317	2.19522
GACUCUGUC/CUGAXACAG	44.8376	43.9963	2.30326
CAGACUAGC/GUCUXAUCG	33.2918	32.9619	2.13916
CAGGCAAGC/GUCCXUUCG	45.3926	45.5793	2.3268
CACACGCAC/GUGUXCGUG	50.746	50.3052	2.39708
CAGUCGAGC/GUCAXCUCG	44.5358	44.3463	2.30846
GACACAGUC/CUGUXUCAG	36.2821	36.7777	2.19591
CAUGCACUAC/GUACAXUGAUG	47.8305	47.7747	2.35945
CAUGACGCUAC/GUACUXCGAUG	55.1948	55.794	2.47871

Tabela 5.2

Sequências usadas para *bulges* citosinas (CX).

Sequências	T_i (°C)	T'_i (°C)	τ
GUGAUGAC/CAXUACUG	32.541	30.5781	1.77482
GUCGUAGC/CAGCAUXG	45.2115	41.8785	2.02213
CAUGAGAGC/GUAXUCUCG	43.8547	43.4581	2.0567
CAGUCGAGC/GUCAGXUCG	43.8637	45.2081	2.09499
CAAUAGUGC/GUUAUXACG	26.9797	30.681	1.77708
UGACGCUCA/ACUGXGAGU	44.1351	45.1317	2.09332
CAGAGCAGC/GUCUXGUCG	47.2519	47.8118	2.15198
GCUGACGAC/CGAXUGCUG	44.4223	47.0123	2.13448
UGAGUGUCA/ACUXACAGU	40.0807	38.3069	1.94397
GUGUGUGUG/CACAXACAC	39.0673	38.483	1.94782
CAUGAGUGC/GUACUXACG	38.3104	40.1794	1.98494
GACAUAGUC/CUGUAUXAG	37.1663	36.7537	1.90997
GACAGCCAG/CUGUXGGUC	46.5079	47.3222	2.14126
GACUGUGUC/CUGAXACAG	46.2153	42.183	2.02879
GACCGUAGC/CUGGXAUCG	48.2356	48.3729	2.16426
GACAGAGUC/CUGUXUCAG	39.7025	39.6599	1.97358
CAUCAGUGC/GUAGUXACG	36.5789	36.2498	1.89895
CAGUGCAGC/GUCAXGUCG	44.3036	46.3895	2.12085
GUCGAUCAC/CAGXUAGUG	38.4338	39.0326	1.95985
GAUGAGAGC/CUAXUCUCG	42.9811	44.6832	2.08351
CAGUCGAGUC/GUCAGXUCAG	50.5982	48.7838	2.17325
GCUCGACCUG/CGAGXUGGAC	55.0346	53.7379	2.28167
CAUGUGACUAC/GUACAXUGAUG	47.8073	47.3862	2.14266

Tabela 5.3

Sequências usadas para *bulges* guanosinas (GX).

Sequências	T_i (°C)	T'_i (°C)	τ
GCGUGCG/CGCXCGC	39.2824	42.3217	2.17117
GCGUCAG/CGCXGUC	39.9724	40.2488	2.1319
GCUCUAUG/CGAGXUAC	40.7755	39.6655	2.12085
GCUCUGUG/CGAGXCAC	41.2376	42.431	2.17325
GACAUGUGC/CUGUXCACG	49.6316	47.2478	2.26451
CAGGUAAGC/GUCCXUUCG	47.1278	46.7457	2.25499
GACGCUAGC/CUGCGXUCG	51.788	51.7208	2.34926
CAGUCGAGC/GUCXGCUCG	53.8397	55.9401	2.4292
CAGAUCAGC/GUCUXGUCG	45.7507	46.1826	2.24433
GACAUAGUC/CUGUXUCAG	38.8403	39.8892	2.12509
GUAUCGUCG/CAUXGCAGC	45.9897	45.4158	2.2298
CGACUGCAG/GCUGXCGUC	49.0449	49.819	2.31322
GCAUCUGUG/CGUXGACAC	44.5166	49.8076	2.31301
ACUGCAUCG/UGACGUXGC	53.6583	49.4762	2.30673
CAGACUAGC/GUCUGXUCG	46.5368	44.9177	2.22036
ACUGUGAGU/UGACXCUCA	47.6017	46.3589	2.24767
GCACUGAGG/CGUGXCUC	53.7339	53.9364	2.39123
CACAUGCAC/GUGUXCGUG	47.9972	46.8393	2.25677
UGACUCUCA/ACUGXGAGU	47.0096	46.8276	2.25655
GGCAUGACG/CCGUXCUGC	52.3475	53.1171	2.37571
CUGCUGUCG/GACGXCAGC	49.662	49.819	2.31322
GAGUCUCAC/CUCXGAGUG	55.9117	52.1917	2.35818
GAGCUGGUC/CUCGXCCAG	55.6933	56.3089	2.43619
GCACUGCAG/CGUGXCGUC	54.1529	52.7388	2.36854
GGUACUAAGC/CCAUGXUUCG	47.8664	50.0014	2.31668

Tabela 5.4

Sequências usadas para *bulges* uridinas (UX).

contendo somente bases CG. O índice de melting é correlatado para medidas de temperaturas de melting T_i usando uma regressão linear

$$T'_i = f_0 + f_1\tau_i, \quad (5.3)$$

onde f_0 e f_1 são coeficientes de regressão que são calculados em cada mudança de parâmetro. Neste trabalho, não foi possível fazer uma regressão dependente do comprimento como em DNAs [40] e RNAs [57] canônicos, devido a insuficiência de variações de comprimento das sequências para cada tipo de *bulge*.

A minimização da equação (5.2) é realizada separadamente para as quatro identidades do *bulge*, isto é, temperaturas de melting de sequências com *single-bulges adenosinas*, *citossinas*, *guanossinas* e *uridinas* foram separados e seus parâmetros minimizados separadamente. Durante a minimização nós variamos o potencial de Morse D de pares de base vizinhos ao *bulge*, pares de base do tipo BP^B, e interações de empilhamento k dos parâmetros dos subgrupos VNNs e BNNs. Para o subgrupo de parâmetros FNN, utilizamos os valores das interações já calculados anteriormente para RNAs [57]. Como o *bulge* é simulado por um sítio vago X na direção oposta da base resíduo, usamos o potencial de Morse $D = 0$ eV, representando a ausência de interação.

Realizamos duas rodadas de minimizações independentes para cada tipo de *bulge*:

Minimização Inicial: Os valores iniciais das interações de *stacking* k dos VNNs é um conjunto de valores randômicos com $\pm 20\%$ dos seus valores correspondentes FNNs. O mesmo é feito para os valores dos parâmetros para o potencial de Morse D dos pares de base flanqueados BP^B. Para os parâmetros de empilhamento BNNs, usamos $k = 2.5$ meV/nm² como valor inicial [40] e para os parâmetros do subgrupo FNN, usamos os valores já calculados para RNAs canônicos 57. Repetimos esse procedimento 200 vezes para cada tipo de *bulge*, que leva cerca de 4000 horas de tempo de processamento em processadores de 1.6 GHz. O valor resultante dos parâmetros é obtido pela média dos resultados providos pela minimização.

Minimização Final: Usamos os valores médios da minimização anterior como novos parâmetros iniciais nesta rodada de minimização. Entretanto, agora serão os dados experimentais que serão alterados randomicamente em pequenas quantidades correspondentes ao erro

Bulges-Model	$\langle \Delta T \rangle$ (°C)	χ^2 (°C ²)
AX-PB	0.473474	0.644346
AX-DPB	0.483394	0.659755
CX-PB	0.220868	0.386695
CX-DPB	0.214324	0.381078
GX-PB	1.11955	1.813
GX-DPB	1.09113	1.81743
UX-PB	1.33933	1.96863
UX-DPB	1.34049	1.95853

Tabela 5.5

Resultado dos parâmetros de minimização para $\langle \Delta T \rangle$ e χ^2 no conjunto utilizado.

experimental das medidas de temperatura [40]. Uma vez que as temperaturas de melting foram derivadas de diferentes fontes, assumimos um erro experimental de 1.3 °C, o mesmo utilizado nos cálculos de RNA canônicos [57]. Isto nos permite estimar a influencia do erro experimental em nossos parâmetros otimizados [40]. Novamente, todos os parâmetros são permitidos variar durante a minimização até alcançar o menor valor possível de χ^2 . Este procedimento foi repetido novamente 200 vezes e o resultado final apresentado aqui é a média de todas as minimizações.

5.6 Resultados e discussões

Começamos discutindo os resultados obtidos para as diferenças médias dos valores de temperaturas experimental e predita (ΔT) pelo modelo nas minimizações para cada identidade do *bulge*. A diferença de temperatura média foi de 0.44 °C para *bulges adenosinas* (AX), 0.32 °C para *bulges citosinas* (CX), 1.42 °C para *bulges guanosinas* (GX) and 1.44 °C para *bulges uridinas* (UX). O baixo valor de ΔT pode ser devido a pequena quantidade de sequências para AX e CX. Por outro lado, o alto valor de ΔT pode ser devido a incerteza sobre a verdadeira identidade dos *bulges* nas medidas de Kent *et al.* [4] devido a classificar os mesmos pelas energias livres. Ainda assim, os valores baixos de variação de temperatura demonstram que a minimização foi bem sucedida e a implementação para estruturas contendo *bulges* ocorreu com sucesso. Os resultados para a diferença de temperatura média ($\langle \Delta T \rangle$) e χ^2 se encontram na tabela 5.5.

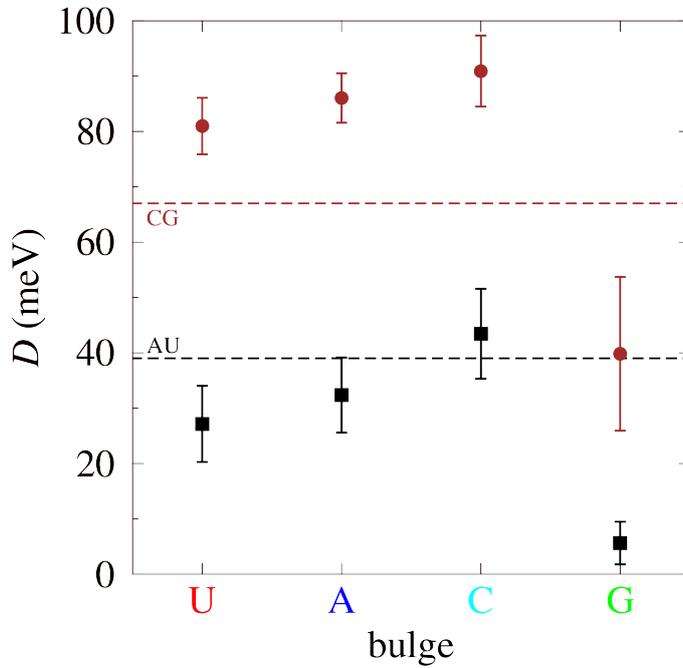
Note que ainda que não descrevamos os duplexos de RNA com o potencial de Dauxois *et al.* [5] no Hamiltoniano do modelo, os parâmetros calculados podem ser utilizados como valores iniciais para esse potencial. Como parte do estudo, utilizei os parâmetros calculados no

Hamiltoniano de Dauxois *et al.* [5] e obtivemos uma pequena variação de ΔT . Os resultados obtidos para *bulges adenosinas* AX quando comparado ao modelo PB com potencial harmônico aumentou de 0.44 °C para 0.45 °C utilizando $\rho = 2.0$ e $\alpha = 0.35 \text{ \AA}^{-1}$. O mesmo comportamento foi visto para as outras identidades de *bulges*. Isso reforça ainda mais o uso do potencial harmônico, pois não é encontrada uma diferença significativa nos cálculos de variação de temperatura. O mesmo comportamento é visto nos resultados para a variação de temperatura média e χ^2 para o modelo de Dauxois *et al.* [5] se encontram na tabela 5.5.

Na figura 5.4 mostramos os parâmetros médios D para os pares de base na vizinhança do *bulge* do tipo BP^B. Eles foram ordenados de forma a mostrar a diferença com relação ao caso canônico. O grupo de sequências contendo *bulges guanosinas* (GX) claramente apresenta uma interação mais fraca com relação as pontes de hidrogênio, sugerindo uma importante perturbação ou até mesmo uma ruptura dos pares de base vizinhos AU e CG. Para o grupo de sequências dos *bulges* restantes, isto é, AX, CX e UX, os *bulges* apresentam uma maior estabilidade estrutural (perturbam menos seus vizinhos) quando flanqueados por pares de base CG. Entretanto, essa estabilidade estrutural é menor quando os *bulges* estão flanqueados por pares de base AU, exceto no caso de *bulges citosinas*. Esse aumento e decréscimo de estabilidade estrutural apresentado pode não ser inteiramente devido as ligações de hidrogênio dos pares de base, considerando que o potencial de Morse é um potencial efetivo e esse pode levar em conta influências da posição da base com as bases vizinhas. Em outras palavras, um aumento no potencial de Morse pode ser devido a alguma interação entre as bases do *bulge*.

Enquanto os resultados para o potencial de Morse demonstram uma importante influência do *bulge* GX no duplexo, é necessário se olhar o empilhamento em torno desse *bulge*. O Hamiltoniano (4.1), mostra que há um balanço energético entre o potencial de Morse e o potencial de empilhamento. Portanto, uma perda de estabilidade devido a ponte de hidrogênio dada a influencia do *bulge*, poderia ser compensada pelo *stacking* na região. A figura 5.5 mostra a constante elástica k para os parâmetros que sofrem influência direta da ação do *bulge*, isto é, o grupo BNN. Os parâmetros foram ordenados de forma crescente pela diferença com relação aos mesmos parâmetros no caso canônico representados pelos quadrados pretos na figura 5.5

$$\delta = |k_{\text{unperturbed}} - k_{\text{perturbed}}|. \quad (5.4)$$

**Figura 5.4**

Valor médio da constante de dissociação D com respeito a pares de base AU^B (preto) e CG^B (marrom) próximos ao *bulge* B com $B=A, C, G$ or U . As linhas tracejadas correspondem aos valores da constante de dissociação para os pares de base não perturbados em uma estrutura canônica [57]. Os *bulges* são ordenados pela diferença quadrática total para o potencial de Morse não perturbado, sendo U o *single-bulge* que causou menor perturbação e G o *single-bulge* de maior perturbação. A identificação da identidade do *bulge* com as cores aqui utilizadas serão as mesmas para as demais figuras.

A figura 5.5b para os empilhamentos do subgrupo BNN mostra uma forte variação dos parâmetros para os *bulges guanosinas*, como no caso do potencial de Morse para as bases vizinhas ao *bulge*. Isso demonstra que os *bulges guanosinas* perturbam fortemente a região onde está localizado, influenciando sua vizinhança. Ainda que *bulges citosinas* também mostraram um grande desvio dos valores canônicos, não se vê um comportamento tão agudo como no caso de *bulges guanosinas*.

A figura 5.6 mostra os resultados encontrados para os parâmetros de *stacking* para o subgrupo VNN. Notamos que os parâmetros desse grupo mudaram muito pouco comparados aos parâmetros canônicos do subgrupo FNN. Lembramos que ao fazer as minimizações, não usamos o mesmo grupo de parâmetros iniciais k do caso canônico como conjunto inicial nessa otimização. Evitamos qualquer tendência numérica em nossos resultados. Logo, obtivemos resultados próximos aos valores canônicos somente pela convergência da minimização. Como os parâmetros do grupo VNN não estão envolvidos diretamente com os *bulges*, mesmo as maiores variações obtidas desse subgrupo comparados com os valores canônicos, são muito menores que os parâmetros observados no subgrupo BNN. Os parâmetros de maiores desvios no subgrupo VNN são para os *bulges citosinas* (CX) e *guanosinas* (GX), resultado similar ao encontrado para o subgrupo BNN. Entretanto, o parâmetro de *stacking* $AUpUA^U$ para os *bulges uridinas* (UX) apresentou um desvio maior que o valor canônico. Isto pode ser explicado pelo balanço energético entre os potenciais, pois o potencial de Morse para o par de base UA^U foi levemente

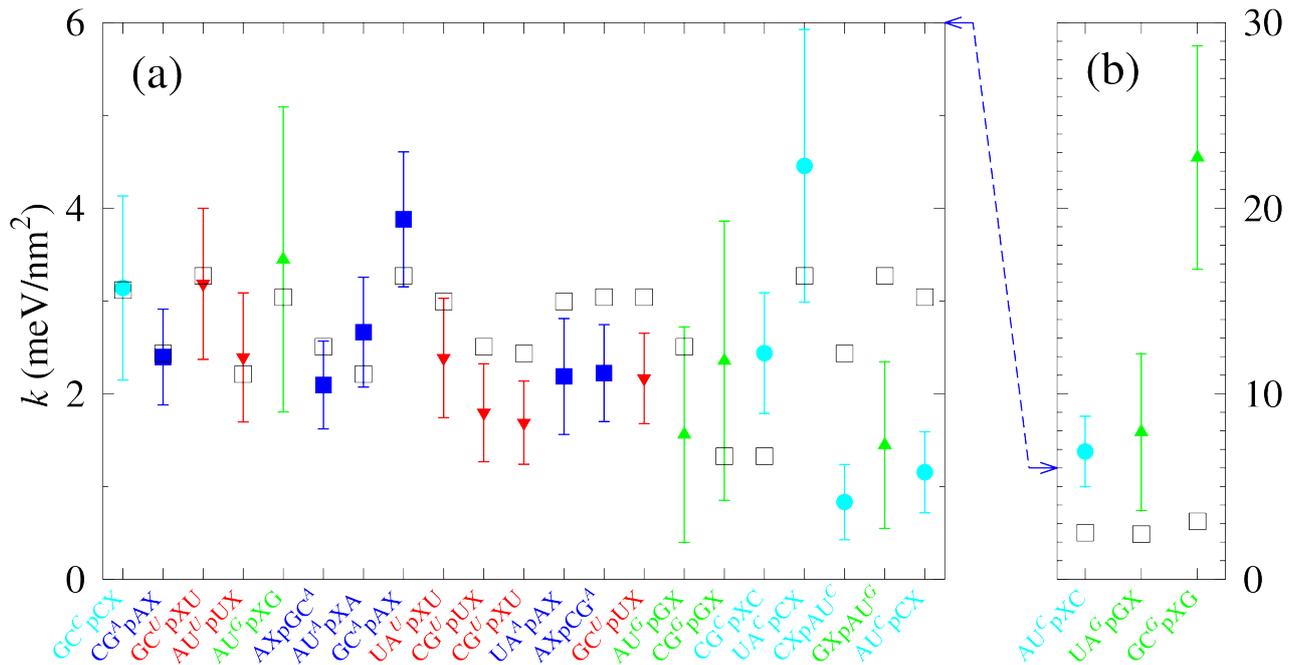


Figura 5.5

Empilhamento médio k dos parâmetros pertencentes ao grupo BNNs em ordem crescente dada pela diferença δ , equação (5.4). As constantes de empilhamento com maior diferença foram separadas na parte (b) com uma escala diferente no eixo vertical. A correspondência da escala entre as escalas da parte (a) e parte (b) é mostrada pela linha tracejada em azul. As cores dos parâmetros correspondem a identidade do *bulge* como na figura 5.4. Os empilhamentos canônicos são representados pelo quadrados pretos, valores que correspondem aos cálculos feitos em [57].

perturbado.

Os parâmetros de *stacking* e Morse calculados mostram uma relação importante do *bulge* contido em um dos monômeros do duplexo. Entretanto, não é correto inferir o efeito dos *bulges* somente olhando para os parâmetros calculados. Felizmente, o modelo Peyrard-Bishop nos permite calcular o perfil de abertura médio $\langle y_n \rangle$ pela função de partição (5.1). Podemos tirar vantagem desse cálculo pois é possível mostrar o efeito combinado de todos esses parâmetros, como também o efeito não-linear do Hamiltoniano característico em duplas hélices de RNAs. Entretanto, nesse tipo de cálculo precisamos respeitar as limitações do modelo, a temperatura de referência precisa ser muito mais baixa que a temperatura de melting do duplexo [59]. Em nossos cálculos, utilizamos a temperatura de 180 K. Portanto, os perfis de abertura demonstram um resultado mais qualitativo da estabilidade da hélice.

Na figura 5.7 mostramos alguns exemplos de perfis de abertura para diversos tipos de *bulges*. Por comparação, nós também acrescentamos na figura as chamadas *sequências-mãe* (parental sequences), isto é, a sequência equivalente com a base resíduo ligada a sua base real complementar que deu origem a sequência com o *bulge*. A sequência da figura 5.7(a) foi adaptada da figura 1a da referência 65, mostrando uma intercalação do *bulge* AX. Nesse cálculo, adicionamos 4 pares de base extras CGs nas extremidades com o objetivo de melhorar o perfil de abertura da molécula. É interessante notar que a abertura induzida por esse *bulge* é da mesma ordem que os desvios dos pares de base nas extremidades da molécula. Em outras palavras, esta é uma importante perturbação comparada a estruturas comuns de RNAs. Entretanto, a perturbação do *bulge* para essa sequência se estende aos dois pares de base no lado esquerdo do duplexo devido à presença de um par de base AU pertencendo ao subgrupo VNN.

Na figura 5.7(b) contendo as duas sequências de Popenda *et al.* [69], onde o *bulge* é flanqueado por bases CGs, temos um comportamento extremamente diferente. Para a sequência contendo o *bulge adenosina*, notamos que o perfil de abertura é mais baixo, porém da mesma ordem de grandeza do que o da figura 5.7(a). Esse perfil sugere que o *single-bulge adenosina* se encontra no plano da molécula. De fato, este foi o resultado obtido por Popenda *et al.* [69] para a sequência acima. Os resultados obtidos por nossos cálculos estão de acordo com os resultados obtidos em medidas de NMR para essa sequência. Um outro fator interessante também discutido por Popenda *et al.* [69] é o fato de pares de base CGs próximos ao *bulge* levarem a

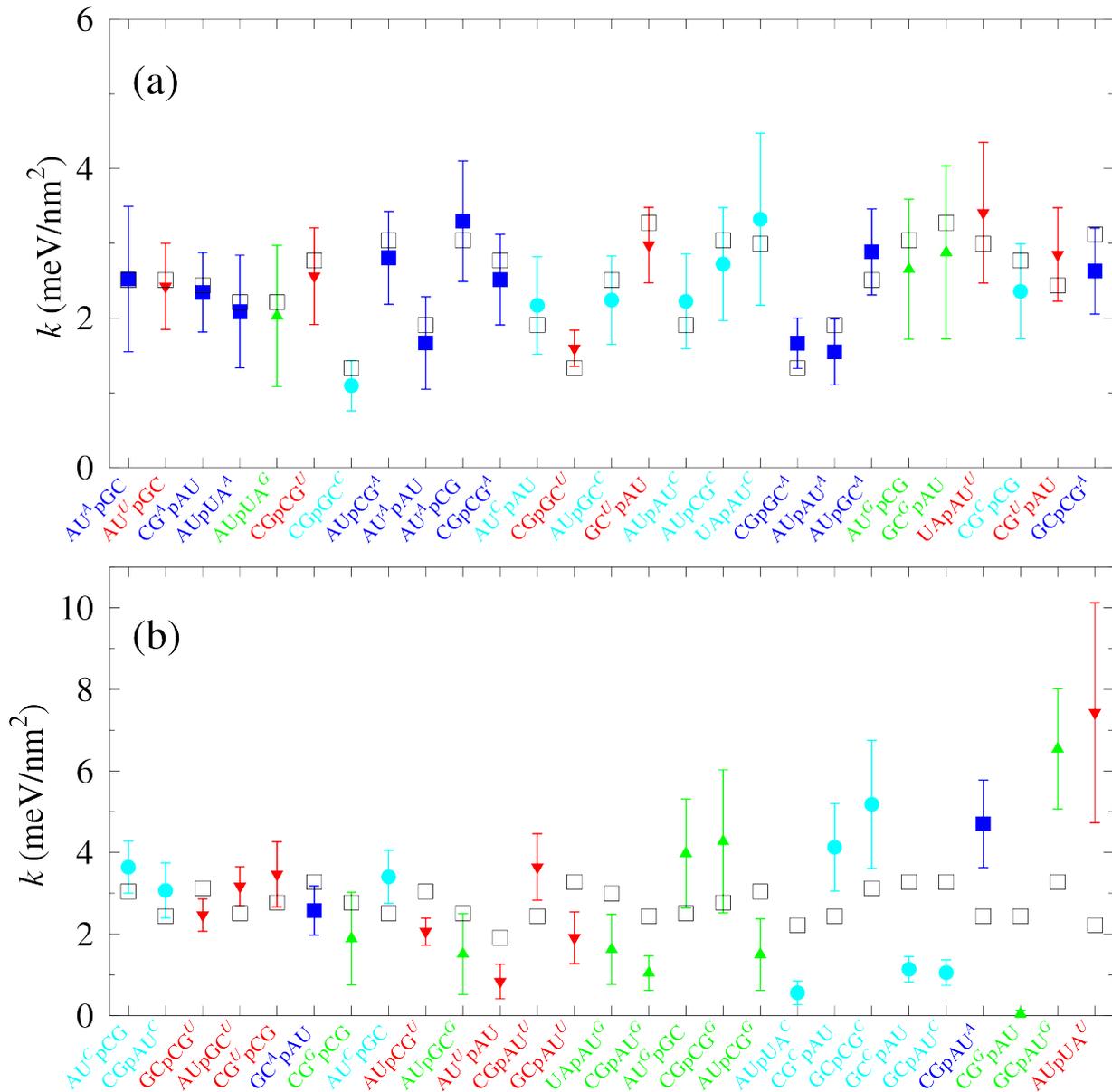


Figura 5.6

Valor médio do empilhamento k dos parâmetros do grupo VNN agrupados em ordem crescente pela diferença com relação aos parâmetros canônicos δ , equação 5.4. Note que as escalas verticais das partes (a) e (b) são diferentes. O esquema de cores definidos para a identidade do *bulge* foi mantido como antes 5.4. Os empilhamentos canônicos são representados pelo quadrados pretos, valores que correspondem aos cálculos feitos em [57].

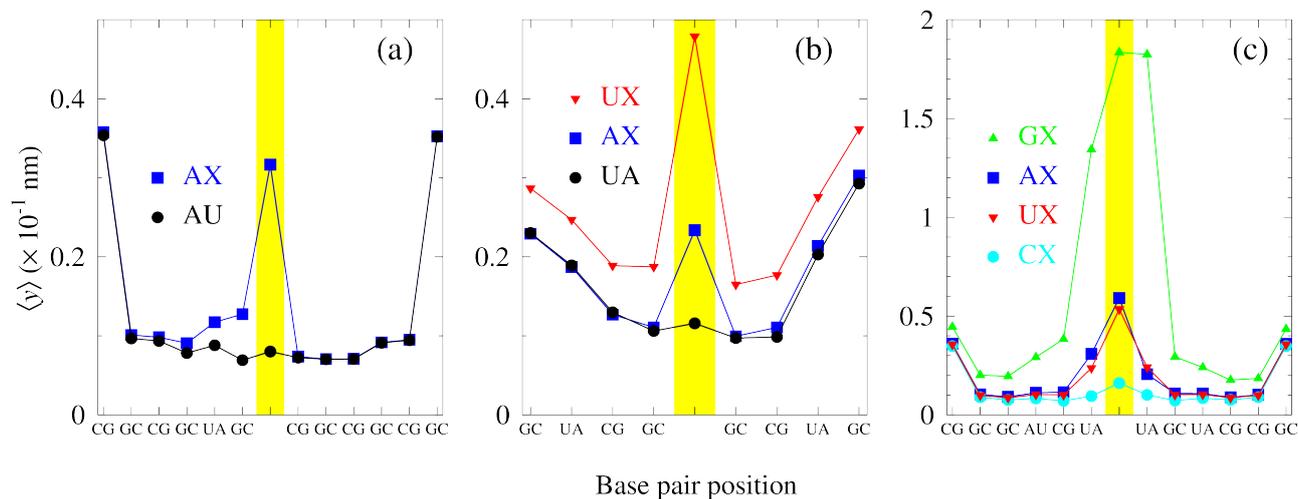


Figura 5.7

Valor médio dos perfis de abertura para diversas sequências do nosso conjunto experimental. A posição do *bulge* é destacada pela área amarela para *bulges* UX (triângulos vermelhos), AX (quadrados azuis), CX (círculos azuis ciano) e GX (triângulos verdes). A correspondente sequência-mãe é mostrada em preto com círculos pretos. Na figura (a) utilizamos as uma das sequências de [65]. Na figura (b) utilizamos as duas sequências de [69]. Na figura (c) é uma sequência arbitrária contendo todas as identidades dos *bulges*. Note que a escala vertical da figura (c) é diferente das figuras (a) e (b). Aqui mantivemos o mesmo esquema de cores para cada identidade do *bulge* como na figura 5.4.

estruturas mais estáveis, facilitando a localização experimental. Notamos essa estabilidade em nossos resultados na figura 5.7(b) ao notarmos que os pares de base flanqueados ao *bulge*, como também o grupo VNN, não sofrem grandes perturbações.

Ainda analisando os perfis de abertura da figura 5.7(b), notamos um alto desvio devido a presença do *bulge uridina*. Nossos resultados sugerem que os parâmetros de empilhamento do *bulge uridina* UX possuem uma maior flexibilidade. Os resultados calculados por nós sugerem que o *bulge* se encontra fora do plano do duplexo. Popena *et al.* [69] constataram por medidas de NMR que o *bulge uridina* da sequência acima se encontra fora do plano da molécula, demonstrando que *bulges uridinas* possuem uma certa preferência para tomar a configuração *flipped out*. Isso reforça ainda mais nossos resultados, pois conseguimos determinar este tipo de propriedade com o modelo PB. A estabilidade de *bulges* flanqueados por pares de base CGs também é observada em nossos cálculos já destacada por Popena *et al.* [69].

Na figura 5.7(c) mostramos uma sequência arbitrária para todos os quatro *bulges*. Notamos como anteriormente já visto pelos resultados do cálculo de parâmetros do modelo, *bulges guanosinas* claramente sofrem uma maior perturbação e essa perturbação se propaga ao longo de seus vizinhos. Até o momento, não é de nosso conhecimento medidas em NMR

para o entendimento dessas perturbações em *bulges guanosinas*. Esses resultados nos mostram que há alguma propriedade física interessante com relação a esses *bulges* que ainda não foi medida. Convidamos os experimentais que estudam *single-bulges* a fazerem essa verificação, para um melhor entendimento e confirmação do nosso resultado. Pelo perfil apresentado, *bulges guanosinas* parecem assumir a configuração *extrahelical* causando uma enorme assimetria no duplexo.

Dado os exemplos de perfis de abertura apresentados na figura 5.7, a correlação termodinâmica e a estrutura da molécula parece ser obtida com o modelo Peyrard-Bishop. Na análise das figuras 5.7(a) e (b) nossos resultados se mostram de acordo com sequências experimentais desses artigos e inferimos a configuração *extrahelical* do *bulge* UX pelo tamanho do perfil de abertura na figura 5.7(b) comparada ao *bulge* AX. Ainda que esse acordo de resultados seja encorajador para dizer sobre a estrutura do duplexo, sabemos que não há dados suficiente para oligonucleotídeos de RNA com *single-bulge* para estender essa análise.

É importante salientar aqui que estamos cientes de muitos poucos estudos estruturais de NMR que combinam as condições experimentais com as medidas de temperatura de melting das quais derivamos nossos resultados[68, 69]. Não consideramos os dados experimentais mais abundantes para RNAs ribossômicos, pois os *bulges* são muito mais sensíveis a proximidade de loops e a estruturas que possuem proteínas no meio, ou seja, estaríamos comparando condições experimentais extremamente diferentes.

5.7 Conclusões

A implementação de assimetria no modelo Peyrard-Bishop no estudo de *bulges* pertencentes ao Grupo I de RNAs foi realizada com sucesso. Conseguimos analisar os perfis de abertura e determinar os parâmetros das sequências para cada identidade do *bulge*. O acordo com os experimentos de NMR também foi mostrado. Mostramos também que os parâmetros calculados pelo modelo PB com potencial harmônico podem ser aplicados ao hamiltoniano de Dauxois *et al.* [5] diretamente. Isso não dificulta a utilização do potencial de Dauxois *et al.*, embora não haja uma melhora significativa com o uso desse potencial. Os resultados obtidos mostraram que *bulges guanosinas* apresentam desvios maiores nos cálculos dos parâmetros

como também no perfil de abertura. Seria interessante medidas de NMR nesses tipos de *bulges* para o entendimento desses desvios. Entretanto, conseguimos ainda que como primeiro passo uma relação entre a termodinâmica do duplexo com a estrutura, relação que não foi obtida pelo modelo mesoscópico de primeiros vizinhos. Nossos resultados também mostraram um acordo com os resultados experimentais para *bulges adenosinas* e *uridinas*. Como acontece em *bulges guanosinas*, nossos resultados não podem descrever propriedades dos *bulges citosinas* devido a falta de medidas de NMR nessas identidades. Buscamos estender nossos resultados para outros tipos de grupos de *bulges* em RNAs e DNAs, como no caso de *bulges* do Grupo II. Queremos determinar a posição da base não pareada nessas estruturas e utilizar os perfis de abertura nessa análise. Sabemos que a conformação que a estrutura irá assumir, reflete em suas propriedades energéticas e buscamos com o modelo PB discrimina-las.

6 Conclusões e perspectivas futuras

O trabalho aqui apresentado demonstra a capacidade do modelo Peyrard-Bishop no estudo de estruturas de duplexos não canônicos e assimétricos. O modelo se destaca pela capacidade de descrever as interações intramoleculares e intermoleculares contida nos duplexos: ligações de hidrogênio dos pares de base existentes e do *bulge* ($D = 0.0$), como também o *stacking* assimétrico entre os vizinhos (FNNs, VNNs e BNNs). O modelo Peyrard-Bishop nos permitiu estudar um grande número de sequências e determinar propriedades gerais para essas estruturas. O modelo possibilitou também fazer uso de dados experimentais na parametrização com o objetivo de tornar a descrição física do problema ainda mais realista e levar em conta os erros experimentais nos valores dos parâmetros.

No estudo de *bulges*, em particular no Grupo I, vimos que as condições de assimetria para o modelo com empilhamento harmônico, foram suficientes para a descrição do *bulge* e o estabelecimento dos valores dos parâmetros ligados a ele. Ao compararmos nossos cálculos com medidas de NMR, foi possível ver que o modelo consegue ainda que de forma aproximada, descrever as propriedades de conformação das sequências através dos perfis de abertura calculados. Notamos também uma perturbação maior por parte de *bulges guanosinas*, ainda não relatado na literatura. A conformação estrutural para essa identidade da protuberância precisa ser verificada com medidas de NMR. Percebe-se claramente uma configuração mais assimétrica que os outros tipos de *bulge*.

Concluimos que o objetivo da tese foi alcançado, pois conseguimos estabelecer uma relação entre as propriedades termodinâmicas com as propriedades estruturais das duplas hélices contendo assimetria por protuberâncias. O modelo Peyrard-Bishop se mostra uma ferramenta confiável na análise de *bulges* com relação as conformações adotadas pelas identidades das protuberâncias, e um estudo energético de energia livre pelo modelo pode ser usado na discussão de estabilidade desses duplexos pertencentes ao Grupo I. Os parâmetros de *single-bulges* do Grupo I calculados aqui foram incluídos em novas versões de nossos softwares Tfreg, e podem ser encontrados na página de nosso grupo para download: <https://sites.google.com/site/geraldweberufmg/home>.

Nossos softwares são livres e disponíveis para download em dispositivos Linux.

Como perspectivas futuras ainda com relação ao estudo de *bulges*, buscamos aplicar esse formalismo a outros grupos de *single-bulges*. Ainda que não apresentado aqui, já começamos a fazer os cálculos para o Grupo II de protuberâncias simples. Entretanto, a ambiguidade na posição da base resíduo faz com que tenhamos um número muito grande de combinações de conjuntos mesmo para poucas sequências:

- *bulges* AX: 14 sequências do Grupo II e 55.296 combinações;
- *bulges* CX: 11 sequências do Grupo II e 6.912 combinações;
- *bulges* GX: 20 sequências do Grupo II e mais de 35.000.000 de combinações;
- *bulges* UX: 07 sequências do Grupo II e 288 combinações.

Temos como objetivo para esse grupo de *bulges*, dizer qual a configuração mais provável para as sequências experimentais desse grupo utilizando o modelo Peyrard-Bishop. O mesmo faremos para o Grupo III e IV. O uso do modelo PB como uma ferramenta capaz de descrever as conformações das protuberâncias junto aos experimentos para todos os grupos existentes, seria uma enorme contribuição para a ciência. Esperamos chegar lá em breve.

A Artigo publicado

Artigo publicado na *The Journal of Chemical Physics*.

An asymmetric mesoscopic model for single bulges in RNA

Erik de Oliveira Martins^{a)} and Gerald Weber^{b)}*Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil*

(Received 17 April 2017; accepted 28 September 2017; published online 17 October 2017)

Simple one-dimensional DNA or RNA mesoscopic models are of interest for their computational efficiency while retaining the key elements of the molecular interactions. However, they only deal with perfectly formed DNA or RNA double helices and consider the intra-strand interactions to be the same on both strands. This makes it difficult to describe highly asymmetric structures such as bulges and loops and, for instance, prevents the application of mesoscopic models to determine RNA secondary structures. Here we derived the conditions for the Peyrard-Bishop mesoscopic model to overcome these limitations and applied it to the calculation of single bulges, the smallest and simplest of these asymmetric structures. We found that these theoretical conditions can indeed be applied to any situation where stacking asymmetry needs to be considered. The full set of parameters for group I RNA bulges was determined from experimental melting temperatures using an optimization procedure, and we also calculated average opening profiles for several RNA sequences. We found that guanosine bulges show the strongest perturbation on their neighboring base pairs, considerably reducing the on-site interactions of their neighboring base pairs. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5006948>

I. INTRODUCTION

Single RNA bulges are unpaired bases that are responsible for important perturbations to the double helix. They occur naturally and are frequently seen in ribosomal RNA¹ but may equally be synthesized by matching strands of unequal length such that one or more bases are left unpaired. Bulges are thought to be responsible for recognition sites by acting as molecular handles.² Given the importance of bulges, it would be of interest to use mesoscopic models, such as those belonging to the class of Peyrard-Bishop (PB) models,³ to describe their thermodynamic properties. For instance, this would enable us to apply this type of model to bioinformatics applications such as secondary structure predictions. Mesoscopic models have an advantage over more complicated approaches such as coarse-grain models⁴ which are computationally too expensive to be extensively used for large scale bioinformatics applications. The simpler PB-type models would be efficient enough for this type of application and are of interest as they describe the molecular interactions which nearest-neighbor (NN) models, for instance, do not.⁵ However, PB models are still largely unable to deal with bulges and loops. Here, we will derive the conditions that need to be met by the model Hamiltonian to overcome some of these limitations. We then apply these conditions to the simplest and smallest perturbation which are single bulges flanked by Watson-Crick base pairs in RNA.

RNA bulges are usually classified in terms of the neighboring bases and location in the helix.⁶ Group I and II bulges are flanked by CG or AU base pairs, while group III and IV bulges may have at least one GU base pair as a neighbor. Group I and

III bulges have no repeated neighboring bases and therefore their location is well defined, see Fig. S1 of the [supplementary material](#). In contrast, groups II and IV have repeated neighbors and the exact location of the bulge in the sequence is ambiguous.

From a structural point of view, bulges are found stacked into the helix (intercalated), away from the helix (flipped out) or bent towards one of the grooves (side-by-side).^{2,7-9} This perturbation to the double helix causes a decrease of the denaturation temperatures.¹⁰ In particular, White and Draper^{11,12} studied the intercalation effects of single bulges in RNA and showed that the inclusion of bulges has a dramatic effect on the conformation possibilities of the helix, confirming similar observations for DNA bulges.¹³ Adenine bulges flanked by mismatched GA base pairs were crystallized in HIV-1 RNA by Ennifar *et al.*¹⁴ and studied with x-ray diffraction revealing a significant asymmetry in the deep groove. In some cases, bulges were found to assume two different conformations, for instance, single cytosine bulges were found to form two different types of extra-helical triples.¹⁵ For uridine bulges, it was found from x-ray diffraction that they flip out and protrude into the minor groove.¹⁶ Similar to RNA, DNA bulges were found to loop out¹⁷ and destabilize the helix.¹⁸⁻²⁰ Popena, Adamiak, and Gdaniec⁹ investigated structural differences between a regular RNA duplex and the one containing an adenosine group II bulge. They found that the distortion caused by the bulge propagates through the whole structure.

Barthel and Zacharias²¹ performed a molecular dynamics simulation for a specific sequence containing a single adenosine or uridine bulge and found that flipping out of the major groove was less favorable than flipping out of the minor groove. They also found that the extra-helical bulges were stabilized by electrostatic interactions while stacked bulges had van der

^{a)}Electronic mail: erikfisica@gmail.com^{b)}Electronic mail: gweberbh@gmail.com

Waals and nonpolar contributions. Molecular dynamics was also used to study single adenosine bulges under pressure.²² Flexibility and bending were simulated with coarse-grained models.²³

The thermodynamic stability of single RNA bulges were analyzed using gradient gel electrophoresis⁶ and melting temperature experiments.^{1,24–26} One important conclusion emerging from these studies is that the NN model is not well suited to handle the structural perturbation to the helix.²⁷ This follows similar conclusions for DNA bulges.^{20,28} Melting temperatures for single bulges in RNA hairpins^{29–31} as well as for longer bulges³² further highlight these shortcomings. Therefore, there is a motivation to use models that go beyond the nearest-neighbor model but which are still computationally efficient to handle a large number of sequences. Mesoscopic approaches, such as the Peyrard-Bishop (PB) model,³ are an interesting choice as they allow describing the stacking interaction separately from the hydrogen bonds.

The PB model is an important statistical model used to study mechanical and thermodynamic properties of DNA and RNA molecules.^{3,33} An advantage of the PB model is its intuitive way to describe interactions of hydrogen bonds and stacking interactions in double strands by a Hamiltonian. The PB model is under active development and is applied to a number of different physical situations involving oligonucleotides. Some recent examples for DNA are its use to study the bubble length distribution,³⁴ charge transport,^{35,36} overstretching transition,³⁷ and flexibility in circular DNA.³⁸ One interesting property of the model is that it is easily adaptable to describe further interactions such as the influence of solvents.^{39,40} Another crucial aspect of the PB model is that it can be parameterized from melting temperatures and the resulting parameters can be easily interpreted in terms of stacking interactions and hydrogen bonds.^{5,41} This parametrization is also remarkably sensitive, for instance, we were able to model the ends of a DNA sequence independently showing an important reduction of hydrogen bond strength for low salt concentrations.⁴² Note that parametrization can also be achieved via molecular dynamics.⁴³ For our purpose, the PB model has an important advantage over other PB-derived models, such as the Dauxois-Peyrard-Bishop (DPB) anharmonic model⁴⁴ or the harmonic model with added solvent potential,³⁹ as it requires fewer parameters making the optimization much easier. This is particularly important when there is a reduced number of melting temperatures.

However, to apply the PB model to bulges, we need to overcome some important limitations of the model. Perhaps the most important approximation is that the PB model considers the stacking interaction to be symmetrical, that is, it considers that the stacking is the same for both strands. For bulges this is certainly not true, whether they are intercalated or flipped out, the stacking on the bulge site will be very different for each strand. Here we modify the PB model to account for asymmetric stacking with some very surprising results that are of importance for normal double helices as well. The second difficulty posed by bulges is that one strand is at least one nucleotide shorter than the other, while the PB model considers only double strands of equal length. This limitation is easier to overcome by simply adding a pseudo-base pairing with the

bulge with no hydrogen bond in a similar approach as used for mismatches.⁴⁵

Having addressed the stacking asymmetry of the PB model, we were able to follow a parametrization procedure⁵ for single group I RNA bulges using the experimental melting temperature from Refs. 1 and 24–26. We have not included type II bulges in our analysis as there is an uncertainty of the location of the bulge position. Bulges neighboring GU (groups III and IV) were also presently not addressed as GU adopts a highly variable conformation, even when surrounded by canonical base pairs.³³

II. MATERIALS AND METHODS

A. The asymmetric stacking of the Peyrard-Bishop model

In the Peyrard-Bishop (PB) model, the Hamiltonian is written in terms of two displacements u_n and v_n , one for each strand, along the same direction perpendicular to the helix longitudinal axis,³

$$H = \sum_n \frac{1}{2} m (\dot{u}_n^2 + \dot{v}_n^2) + \frac{1}{2} k (u_n - u_{n-1})^2 + \frac{1}{2} k (v_n - v_{n-1})^2 + V(u_n - v_n), \quad (1)$$

where k is the stacking constant, V is a potential representing the hydrogen bonds, and m is the mass of the bases. It is implied that k is in fact $k_{n-1,n}$ and V is V_n , but we will omit these subscripts to ease the notation as they will not change the results. The effective potential representing the on-site interaction is given by a Morse potential⁴⁶

$$V(u_n - v_n) = D(e^{-\frac{1}{\lambda\sqrt{2}}(u_n - v_n)} - 1)^2, \quad (2)$$

where D is the main parameter representing the strength of the potential and λ controls the potential width.

Crucially, Eq. (1) considers the stacking interaction as symmetric, using a uniform harmonic coupling k . A uniform stacking prevents us from applying the model to situations where there is an important strand asymmetry. Here, we introduce asymmetric stacking k_u and k_v , corresponding to the displacements u_n and v_n . This approach still allows us to use the same change of variables as for the strand-symmetric model³

$$x_n = \frac{1}{\sqrt{2}}(u_n + v_n), \quad y_n = \frac{1}{\sqrt{2}}(u_n - v_n), \quad (3)$$

such that

$$u_n = \frac{1}{\sqrt{2}}(x_n + y_n), \quad v_n = \frac{1}{\sqrt{2}}(x_n - y_n), \quad (4)$$

and the factorization in regard to the momentum also still applies,

$$Z = (2\pi m k_B T)^N Z_{xy}. \quad (5)$$

However, unlike the symmetric PB model, we cannot immediately decouple the x and y terms of the partition function,

instead we end up with cross terms,

$$Z_{xy} = \prod_{n=1}^N \int dx_n dy_n \exp \left[-\beta \left\{ \frac{1}{2} k_a [(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2] + k_s (x_n - x_{n-1})(y_n - y_{n-1}) + V(y_n) \right\} \right], \quad (6)$$

where

$$k_a = \frac{1}{2}(k_u + k_v), \quad k_s = \frac{1}{2}(k_u - k_v). \quad (7)$$

Nevertheless, it is still possible to factorize the partition function with some algebraic effort. We rewrite the function partition Z_{xy} in the form

$$Z_{xy} = \prod_{n=1}^N \int dy_n \exp \left\{ -\beta \left[\frac{1}{2} \gamma (y_n - y_{n-1})^2 + V(y_n) \right] \right\} \times \int dx_n \exp \left\{ -\beta \left[\sqrt{\frac{k_a}{2}} (x_n - x_{n-1}) + \frac{k_s}{\sqrt{2k_a}} (y_n - y_{n-1}) \right]^2 \right\}, \quad (8)$$

where we used the definition

$$\gamma = \frac{k_a^2 - k_s^2}{k_a}. \quad (9)$$

We introduce another change of variables

$$r_n = \sqrt{\frac{k_a}{2}} (x_n - x_{n-1}) + \frac{k_s}{\sqrt{2k_a}} (y_n - y_{n-1}) \quad (10)$$

with

$$dr_n = \sqrt{\frac{k_a}{2}} dx_n, \quad (11)$$

which reduces the second group of integrals in Eq. (8) to

$$\left(\frac{2}{k_a} \right)^{\frac{N}{2}} \prod_{n=1}^N \int dr_n e^{-\beta r_n^2} = \left(\frac{2\pi}{\beta k_a} \right)^{\frac{N}{2}}. \quad (12)$$

The partition function Eq. (8) can now be factorized,

$$Z_x = \left(\frac{2\pi}{\beta k_a} \right)^{\frac{N}{2}} \quad (13)$$

and

$$Z_y = \prod_{n=1}^N \int dy_n \exp \left\{ -\beta \left[\frac{1}{2} \gamma (y_n - y_{n-1})^2 + V(y_n) \right] \right\}. \quad (14)$$

The important result here is that the factorized partition functions Z_x and Z_y are functionally identical to the partition function of the symmetric model.³ The only difference is that the original symmetric k is now replaced by an effective stacking constant γ in Z_y ,

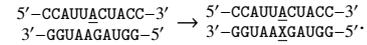
$$\frac{1}{\gamma} = \frac{1}{2} \left(\frac{1}{k_u} + \frac{1}{k_v} \right). \quad (15)$$

This result is valid for any interstrand potential V and for harmonic stacking potentials. However, it cannot be applied to anharmonic stacking as proposed by Dauxois, Peyrard, and Bishop⁴⁴ where this type of decoupling is not possible. In this case, the Z_{xy} partition function would have to be solved

numerically as a two-dimensional Fredholm equation which is out of the scope of this work. One should be aware that the anharmonic or DPB model displays a sharp increase in the average opening mainly due to a numerical artifact,^{39,47} and parameter optimizations for this model have not produced better results than the harmonic PB model.⁵

B. Bulge implementation and notation

The PB model cannot account for strands of different sizes but we bypass this difficulty by introducing a pseudo-base at the opposite position of the bulge as shown schematically in Fig. 1. We use the character X, representing the pseudo-base, placed on the shorter opposite strand visually filling in the vacancy, as in the following example of an adenine bulge:



We will refer to the pseudo-base-pair formed by bulges using the generic notation BX with B = A, C, G, or U. In our example, that would be an AX bulge-pseudo-base.

The Morse potential of the bulge-pseudo-base will be set to zero as there is no interstrand interaction as shown schematically in Fig. 1. However, we cannot do the same for the stacking interaction k between the pseudo-base X and its neighbors. Effectively, as we are inserting the pseudo-base X between two bases, the stacking interaction will be split into two parts. To understand this, let us use the strand notation of the Hamiltonian Eq. (1) and assume, for instance, that the bulge is located at the n th position on the u -strand and the X base at the n th site on the v -strand. The stacking interaction of the neighboring bases at $n-1$ and $n+1$, $k_{v;n-1,n}$ and $k_{v;n,n+1}$, will represent what would otherwise be $k_{v;n-1,n+1}$ bridging the bulge, see Fig. 1.

In our optimization scheme, described in Sec. II D, we will allow for the canonical base pairs BP to assume different

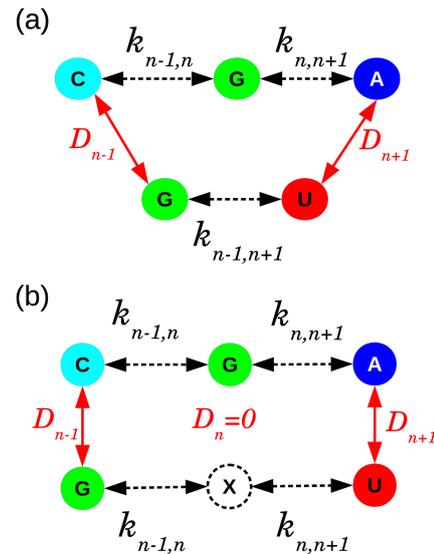


FIG. 1. Bulge implementation. (a) shows an example of a G bulge flanked by CG and AU base pairs. (b) shows the introduction of a pseudo-base X with a zero Morse potential opposite to the G base.

parameters when they are neighbors of a bulge B. To represent this unambiguously, we write this as BP^{B} . For instance, a CG base pair next to an AX bulge will be written as CG^{A} , similarly as adopted in Ref. 33. Therefore, CG^{A} is a base pair with variable parameters next to a bulge A, while CG without the superscript has constant parameters and is not flanking a bulge. We do not distinguish on which side of the CG^{A} is located as doing so would require an additional index and would double the number of parameters to be minimized.

A consequence of this notation is that we will deal with three types of nearest-neighbours.

- Fixed nearest-neighbors (FNNs)* are formed by two AU or CG base pairs away from any bulge. These will be indicated in the form BPpBP . For instance, AUpCG is an AU base pair followed by a CG base pair, neither being flanked by a bulge. FNNs will not change their parameters during the minimization procedure.
- Variable nearest-neighbors (VNNs)* are formed by two AU or CG base pairs, one of which is flanking a bulge. This is written either as $\text{BP}^{\text{B}}\text{pBP}$ or as BPpBP^{B} depending on which base pair BP is next to a bulge B. VNNs will vary their parameters during the minimization.
- Bulge nearest-neighbors (BNNs)* are formed by a bulge-pseudo-base of type BX and either an AU or a CG base pair. This will be written generically as $\text{BP}^{\text{B}}\text{pBX}$ or BXpBP^{B} . In our example, at the beginning of this section, that would be an AX bulge with $\text{UA}^{\text{A}}\text{pAX}$ nearest-neighbor at the left-hand side and AXpCG^{A} at the right-hand side.

Our modeling strategy is to allow the bulge-flanking base-pairs, of type BP^{B} , to vary their Morse potential D , while the remaining CG and AU base pairs will keep their Morse potential constant. For example, CG^{A} indicates that this particular CG base pair will have its hydrogen bonds altered due to the presence of an adenine bulge. We will also vary the stacking interactions of VNNs as well as of BNNs, and only FNNs will remain fixed.

C. Melting temperature data set

We collected a total of 80 sequences with group I single bulges from Refs. 1 and 24–26, of which 18 are adenine (AX), 12 are cytosine (CX), 24 are guanosine (GX), and 26 are uridine bulges (UX). Note that Kent *et al.*²⁶ initially described their bulges as being of group III; however, their free energy analysis confirmed that they are in fact group I bulges.

The melting temperatures were recalculated to a strand concentration of 200 μM from their respective total enthalpy and entropy variations, ΔH and ΔS .^{1,24–26} This is done for consistency with our previous calculations for CG and AU RNA base pairs.⁴¹ The complete list of sequences and their adjusted melting temperatures are shown in Tables S1–S4 of the [supplementary material](#).

D. Optimization of the model parameters

We used the technique of thermal equivalence to optimize the parameters of the potentials presented.⁴⁸ The bulge parameters were optimized by minimizing the square differences

between experimental melting temperature T_i and predicted melting temperature T'_i of all sequences,⁵

$$\chi^2 = \sum_{i=1}^N (T'_i - T_i)^2. \quad (16)$$

We will also refer to the average melting temperature deviation ΔT

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T'_i - T_i|. \quad (17)$$

The predicted melting temperature T'_i is calculated from its melting index τ_i .⁴⁸ We expand the classical partition function, Eq. (14), into non-diagonal terms of the transfer integral matrix.⁴⁹ The adimensional melting index τ_i for each sequence i is the order of largest term of this expansion and is a measure of how much the sequence differs from a poly-CG reference sequence. This index is then correlated to the measured temperatures T_i using linear regression⁵⁰

$$T'_i = f_0 + f_1 \tau_i, \quad (18)$$

where f_0 and f_1 are the regression coefficients that are recalculated every time the model parameters change.

It was not possible to use a length-dependent regression as previously used for DNA⁵ and RNA⁴¹ due to insufficient variations of sequence lengths in the melting temperature data set. However, it was possible to perform separate regressions for each type of bulge. Further details concerning the model implementation are described in Refs. 48 and 49, and a free software implementation is reported in Ref. 51.

E. Initial model parameters and minimization procedure

The minimization of Eq. (16) is carried out separately for the four types of bulges, that is, melting temperatures of adenine, cytosine, guanosine, and uracil bulges were separated and their parameters optimized independently. During minimization, we varied the Morse potential D of all bulge-flanking base pairs, of type BP^{B} , and the stacking interaction k of VNNs and BNNs. For the canonical base pairs and FNNs, we used the parameters previously calculated for RNA.⁴¹ As the bulge is simulated by a vacant site X at the opposite strand, we use a zero Morse potential, $D = 0$, representing the absent interstrand interaction as shown in Fig. 1.

We performed two rounds of minimizations, independently for each type of bulge.

1. Initial minimization

The initial stacking interaction k of VNNs is set to a random value within $\pm 20\%$ of their corresponding FNNs. The same is done with the Morse potential D of bulge-flanking base pairs BP^{B} . For BNNs, we use $k = 2.5 \text{ meV/nm}^2$ as an initial value,⁵ and for the FNNs, we used the RNA parameters calculated in Ref. 41. We repeated this procedure 200 times for each bulge type, which takes around 4000 h of processing time on 1.6 GHz processors. The resulting value of each parameter is obtained by averaging over all results provided by the minimization.

2. Final minimization

We use the averaged results from the previous round as new initial parameters. But now the experimental data set is randomly altered by small amounts corresponding to the experimental error of the temperature measurements.⁵ Since the melting temperatures were derived from different sources, we assumed an experimental error of 1.3 °C which is the same used for the calculation of canonical RNA.⁴¹ This allowed us to estimate the influence of experimental errors on our optimized parameters.⁵ Again, all parameters are allowed to vary during this minimization to reach the smallest possible value of χ^2 . This procedure was repeated again 200 times, and the final results presented here are averaged values over all minimizations.

III. RESULTS AND DISCUSSION

The first result of this work is to show that the stacking asymmetry is of little relevance for the partition function of the PB Hamiltonian. Equation (14) shows that the Z_y asymmetric partition function is formally identical to its symmetric counterpart.³ The Z_x partition function Eq. (13) will be numerically different; however, as it is always factored out in subsequent calculations, this is of no practical consequence. The stacking constant k can now be interpreted as a harmonic average of two asymmetric constants k_u and k_v , as shown in Eq. (15). This is physically similar to an equivalent elastic constant of two springs in series, which is indeed the case as both harmonic potentials are written along the same direction. This result has consequences beyond its application to bulges as it helps explain why the stacking asymmetry so far has not been a problem for the application of the PB model. From a practical point of view, it also means that we can apply the PB without modification to any situation that involves asymmetric stacking, and we only need to be aware that the resulting stacking interaction is in fact an average of two intra-strand stacking factors.

The analysis of the asymmetric stacking Hamiltonian means that we can fully apply the existing framework of the PB model and perform parameter optimization in the same way as for the symmetric model.⁵ Essentially, one only needs to be aware that the resulting stacking interaction is in fact an effective parameter describing the interaction of both strands, see Eq. (15). The PB model was developed for strands of equal length, yet bulges arise due to the fact that one strand is shorter than the other. We compensate for this by introducing a pseudo-base X that does not interact with the bulge, see Fig. 1. This non-interaction is represented by setting the Morse potential between the bulge and pseudo-base to zero. For the intra-strand stacking representing the shorter strand which bridges the bulge, this will be split into two components. The resulting stacking interaction, at the bulge region, will therefore be an effective parameter representing both strands as well as the bulge region as discussed in Sec. II B.

Following the procedure used in our work for canonical RNA,⁴¹ minimization was carried out in two main rounds: The first one using random initial parameters and the second one by varying the melting temperatures (Sec. II C) to evaluate the influence of the experimental uncertainty on our results,

see Sec. II E. The error bars in the following figures represent the uncertainty evaluation of the second minimization round. The final average difference between measured and predicted temperatures, ΔT Eq. (17), was 0.44 °C for AX, 0.32 °C for CX, 1.42 °C for GX, and 1.44 °C for UX, see also Table S5 of the [supplementary material](#) for χ^2 . The low ΔT may be due to the small amount of sequences available for AX and CX. On the other hand, the larger ΔT could be due to the uncertainty regarding the true identity of the bulges measured by Kent *et al.*²⁶ as noted in Sec. II C. Either way, these values indicate that the minimization progressed well despite the difficulties represented by the regression scheme of Eq. (18). Note that although our results were optimized for the PB model, they can be used for the DPB model with very little change of ΔT . For instance, ΔT increases from 0.44 °C to 0.45 °C for AX using $\rho = 2.0$ and $\alpha = 0.35 \text{ \AA}^{-1}$, see also Table S5 of the [supplementary material](#).

In Fig. 2, we show the averaged Morse parameters D for the base-pairs in the vicinity of a bulge, type BP^B. They are ordered to increasing difference to the canonical unperturbed base-pairs. G bulges clearly stand out with very low Morse potentials which would suggest an important perturbation or even a disruption of the neighboring AU and CG base pairs. For U, A, and C bulges, there appears to be a stabilization of the CG base-pairs while AU base-pairs appear to be less stable, except for C bulges. This increased on-site stabilization may not be entirely due to the base-pair hydrogen bonding, considering that the Morse potential is an effective potential; it may also account for influences of the bulged base towards its neighboring bases. In other words, some of the increased Morse potentials may be due to some interstrand interaction of the bulge.

While the results for the Morse potential indicate an important influence of the GX bulge, one cannot understand the influence of the bulge solely from the on-site interaction. The model Hamiltonian, Eq. (1), shows that there is an energetic balance between the on-site Morse potential and the

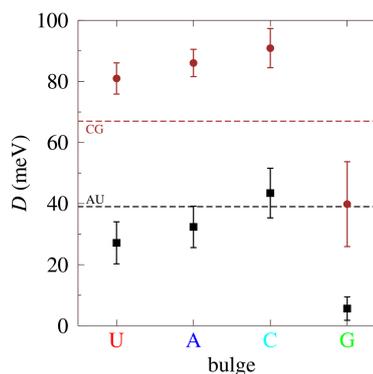


FIG. 2. Average Morse potential D of AU^B (black boxes) and CG^B (brown bullets) base pairs neighboring a bulge B, with B = A, C, G, or U. The dashed lines are the corresponding Morse potential of unperturbed base pairs AU and CG.⁴¹ Bulges are shown ordered by total squared difference to the unperturbed Morse potential, with U being the bulge that caused the smallest perturbation to base pairs and G being the bulge that caused the largest perturbation to base pairs. The color coding of the horizontal labels is the same used for the remaining figures.

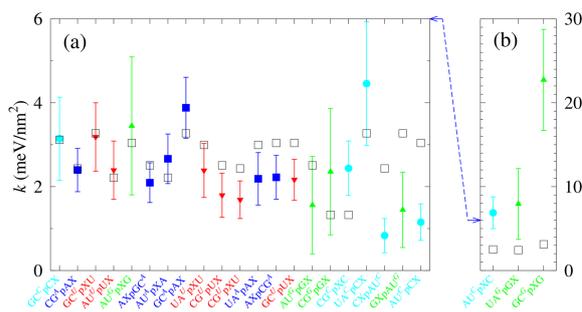


FIG. 3. Average stacking constants k of BNNs ordered to increasing difference δ , Eq. (19). The stacking constants with the largest difference were separated in (b) with a different vertical scale as in (a). The correspondence of scale between (a) and (b) is shown by the dashed blue line. Color coding follows Fig. 2. Black squares are the stacking constants of the corresponding unperturbed FNNs from Ref. 41.

harmonic nearest-neighbor potential, that is, a loss of stabilization due to hydrogen bonding could be compensated by the stacking interaction. Figure 3 shows the harmonic constants k for bulge perturbed BNNs ordered to increasing difference δ to their equivalent unperturbed FNNs (shown as black squares),

$$\delta = |k_{\text{unperturbed}} - k_{\text{perturbed}}|. \quad (19)$$

Figure 3(b) shows that also for the stacking interaction, there is a strong variation for GX bulges, yet here also CX bulges show important deviation from the equivalent FNNs. Figure 4 shows the stacking interaction for VNNs, that is, canonical NNs next to a bulge. For most VNNs, there is very little change compared for their equivalent FNNs. Note that for VNNs, we did not use as initial k value that of their FNNs; therefore, in the case where the VNNs converged to the same average value as the FNNs, this was not due to a biased minimization. As these VNNs do not directly involve a bulge, even the largest variations are much smaller than those observed for the BNNs of Fig. 3(b). The majority of VNNs with large deviations are for CX and GX bulges similar to what was observed for BNNs, yet intriguingly the largest variation of all is for the UX related AUpUA^{U} which corresponds to the least perturbed Morse potential.

A. Opening profiles

The results of interstrand and stacking interaction shown in Sec. III point to an intricate interplay between these two components of the PB Hamiltonian. Therefore, it is not

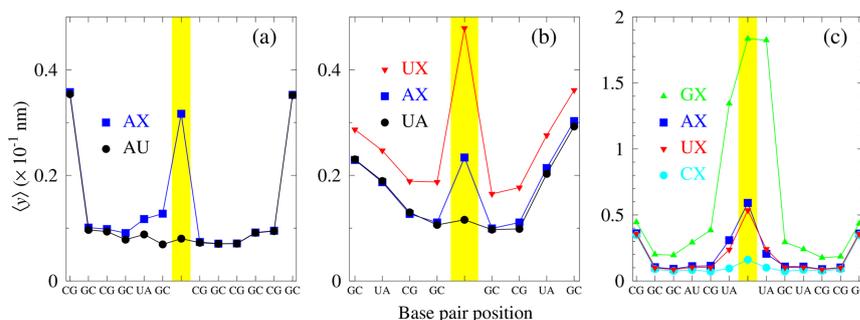


FIG. 5. Average opening profiles for several example sequences. The position corresponding to the bulge is highlighted by a yellow area for UX (red down triangles), AX (blue boxes), CX (cyan bullets), and GX (green up triangles). The corresponding parental sequences are shown as black bullets. Sequences are adapted from (a) Ref. 2 and (b) Ref. 52, while (c) is an arbitrary sequence. Note that the vertical scale of (c) differs from (a) and (b). Color coding follows Fig. 2.

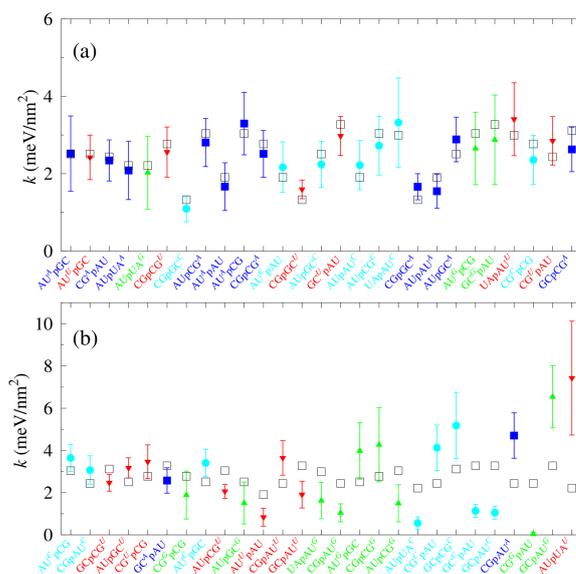


FIG. 4. Average stacking constants k of VNNs ordered to increasing difference δ , Eq. (19). Note that the vertical scale of (a) and (b) differs. Color coding follows Fig. 2. Black squares are the stacking constants of the corresponding FNNs from Ref. 41.

straightforward to infer the effect of the bulge simply by inspecting those parameters. Fortunately, the PB model allows the calculation of the average opening profiles by computing the expected value for y_n from the partition function of Eq. (14). This has the advantage of showing the combined effect of all these parameters as well as the non-linear effect of the model Hamiltonian which may extend over the whole sequence. However, we need to caution that when analyzing these profiles, one needs to be aware that, due to a limitation of the model, the reference temperature in the partition functions needs to be set much lower than the actual melting temperatures; here we used 180 K.⁵¹ Therefore the opening profiles should only be seen as a qualitative representation of the helix stability.⁵¹

In Fig. 5, we show some examples of opening profiles for several types of bulges. For comparison, we also show what we call the corresponding parental sequences, that is, the equivalent sequence with the bulge position filled in by a real complementary base. The sequence of Fig. 5(a) was adapted from Fig. 1(a) of Ref. 2, which shows an intercalated AX bulge, by adding extra CG base pairs at the ends. It is

interesting to note that the strand opening induced by the bulge is of the same order as that of its terminals. In other words, it is an important perturbation yet comparable to other common structures in the RNA sequence. However, the bulge perturbation extends over two base-pair positions, towards the left-hand side, due to the presence of an UA base pair. This is markedly different from the situation presented in Fig. 5(b), showing two sequences from the work of Popena *et al.*,⁵² where the bulge is surrounded by two GC base-pairs to each side. The sequences of Fig. 5(b) is one of the very few NMR studies we are aware of for short RNA in solution with group I bulges. Another sequence studied with NMR by van den Hoogen *et al.*⁷ is almost identical to the one of Fig. 5(b) and is shown in Fig. S2 of the [supplementary material](#). Figure 5(c) shows an arbitrary sequence with all four bulges, where the GX bulge clearly stands out for its very strong perturbation, especially to its neighboring base-pairs.

Given the examples of opening profiles presented in Fig. 5, it would be tempting to try a correlation of the average openings with structural data. For instance, considering the sequence of Fig. 5(b), the NMR data reports a side-by-side formation for AX and a looped-out conformation for UX.⁵² This seems to be consistent with the opening profile, that is, a small opening for AX and a much larger opening for UX. While this agreement is encouraging, there is not enough structural data for short sequences with single bulges in solution to extend this analysis. We are aware of very few NMR structural studies^{7,52} that match the experimental conditions to those of the melting temperature measurements from which we derive our results. We have not considered the more abundant structural data on ribosomal RNAs as it is evident that bulges are very sensitive to the proximity of loops and proteins which are present in these cases, that is, we would be comparing exceedingly different experimental conditions.

IV. CONCLUSIONS

We derived the conditions for which the Peyrard-Bishop mesoscopic model can describe highly asymmetric RNA strands. These conditions hold for any model with a harmonic stacking potential which includes the solvent potential Hamiltonian.³⁹ This was applied successfully to model the melting temperatures of group I RNA bulges, and we showed that the resulting parameters can also be used for other PB-derived models such as the DPB model.⁴⁴ We showed that guanosine bulges display the largest difference in model parameters when compared with Watson-Crick base pairs resulting in strong perturbations to their neighboring base pairs. These results make it possible to extend this type of approach to other types of bulges in RNA as well as in DNA.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for Fig. S1 illustrating group I and group II bulges. Figure S2 shows additional examples of average opening profiles with U-bulges. Tables S1–S4 show all sequences and experimental and predicted temperatures used in this work. Table S5 shows fitting quality parameters for the PB and the PBD models.

ACKNOWLEDGMENTS

The authors acknowledge the agencies that funded this research: CNPq, Fapemig, and Capes.

APPENDIX : SOFTWARE AND DATA AVAILABILITY

The bulge parameters calculated in this work were included in version 3.1 of our TtReg software⁵¹ and can be found at <http://tinyurl.com/tfregufmg>.

- ¹B. M. Znosko, S. B. Silvestri, H. Volkman, B. Boswell, and M. J. Serra, "Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges," *Biochemistry* **41**, 10406–10417 (2002).
- ²T. Hermann and D. J. Patel, "RNA bulges as architectural and recognition motifs," *Structure* **8**, R47–R54 (2000).
- ³M. Peyrard and A. R. Bishop, "Statistical mechanics of a nonlinear model for DNA denaturation," *Phys. Rev. Lett.* **62**, 2755–2757 (1989).
- ⁴P. Šulc, F. Romano, T. E. Ouldridge, J. P. Doye, and A. A. Louis, "A nucleotide-level coarse-grained model of RNA," *J. Chem. Phys.* **140**, 235102 (2014).
- ⁵G. Weber, J. W. Essex, and C. Neylon, "Probing the microscopic flexibility of DNA from melting temperatures," *Nat. Phys.* **5**, 769–773 (2009).
- ⁶J. Zhu and R. M. Wartell, "The effect of base sequence on the stability of RNA and DNA single base bulges," *Biochemistry* **38**, 15986–15993 (1999).
- ⁷Y. T. van den Hoogen, A. A. van Beuzekom, E. de Vroom, G. A. van der Marel, J. H. van Boom, and C. Altona, "Bulge-out structures in the single-stranded trimer AUA and in the duplex (CUGGUGCGG). (CCGCCAG). A model-building and NMR study," *Nucleic Acids Res.* **16**, 5013–5030 (1988).
- ⁸W. A. Hastings, Y. G. Yingling, G. S. Chirikjian, and B. A. Shapiro, "Structural and dynamical classification of RNA single-base bulges for nanostructure design," *J. Comput. Theor. Nanosci.* **3**, 63–77 (2006).
- ⁹L. Popena, R. W. Adamiak, and Z. Gdaniec, "Bulged adenosine influence on the RNA duplex conformation in solution," *Biochemistry* **47**, 5059–5067 (2008).
- ¹⁰J. W. Nelson, F. H. Martin, and I. Tinoco, "DNA and RNA oligomer thermodynamics: The effect of mismatched bases on double-helix stability," *Biopolymers* **20**, 2509–2531 (1981).
- ¹¹S. A. White and D. E. Draper, "Single base bulges in small RNA hairpins enhance ethidium binding and promote an allosteric transition," *Nucleic Acids Res.* **15**, 4049–4064 (1987).
- ¹²S. A. White and D. E. Draper, "Effects of single-base bulges on intercalator binding to small RNA and DNA hairpins and a ribosomal RNA fragment," *Biochemistry* **28**, 1892–1897 (1989).
- ¹³J. W. Nelson and I. Tinoco, Jr., "Ethidium ion binds more strongly to a DNA double helix with a bulged cytosine than to a regular double helix," *Biochemistry* **24**, 6416–6421 (1985).
- ¹⁴E. Ennifar, M. Yusupov, P. Walter, R. Marquet, B. Ehresmann, C. Ehresmann, and P. Dumas, "The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges," *Structure* **7**, 1439–1449 (1999).
- ¹⁵Y. Xiong and M. Sundaralingam, "Two crystal forms of helix II of xenopus laevis 5S rRNA with a cytosine bulge," *RNA* **6**, 1316–1324 (2000).
- ¹⁶Y. Xiong, J. Deng, C. Sudarsanakumar, and M. Sundaralingam, "Crystal structure of an RNA duplex r(gugucgac)₂ with uridine bulges," *J. Mol. Biol.* **313**, 573–582 (2001).
- ¹⁷L. Joshua-Tor, D. Rabinovich, H. Hope, F. Frolov, E. Appella, and J. L. Sussman, "The three-dimensional structure of a DNA duplex containing looped-out bases," *Nature* **334**, 82–84 (1988).
- ¹⁸D. A. LeBlanc and K. M. Morden, "Thermodynamic characterization of deoxyribooligonucleotide duplexes containing bulges," *Biochemistry* **30**, 4042–4047 (1991).
- ¹⁹H. Beltz, J. Azoulay, S. Bernacchi, J.-P. Clamme, D. Ficheux, B. Roques, J.-L. Darlix, and Y. Mély, "Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7," *J. Mol. Biol.* **328**, 95–108 (2003).
- ²⁰F. Tanaka, A. Kameda, M. Yamamoto, and A. Ohuchi, "Thermodynamic parameters based on a nearest-neighbor model for DNA sequences with a single-bulge loop," *Biochemistry* **43**, 7143–7150 (2004).

- ²¹A. Barthel and M. Zacharias, "Conformational transitions in RNA single uridine and adenosine bulge structures: A molecular dynamics free energy simulation study," *Biophys. J.* **90**, 2450–2462 (2006).
- ²²P. Kumar, J. Lehmann, and A. Libchaber, "Kinetics of bulge bases in small RNAs and the effect of pressure on it," *PLoS One* **7**, e42052 (2012).
- ²³J. S. Schreck, T. E. Ouldrige, F. Romano, A. A. Louis, and J. P. Doye, "Characterizing the bending and flexibility induced by bulges in DNA duplexes," *J. Chem. Phys.* **142**, 165101 (2015).
- ²⁴J. M. Blose, M. L. Manni, K. A. Klapek, Y. Stranger-Jones, A. C. Zyra, V. Sim, C. A. Griffith, J. D. Long, and M. J. Serra, "Non-nearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops," *Biochemistry* **46**, 15123–15135 (2007).
- ²⁵M. D. McCann, G. F. Lim, M. L. Manni, J. Estes, K. A. Klapek, G. D. Frattini, R. J. Knarr, J. L. Gratton, and M. J. Serra, "Non-nearest-neighbor dependence of the stability for RNA group II single-nucleotide bulge loops," *RNA* **17**, 108–119 (2011).
- ²⁶J. L. Kent, M. D. McCann, D. Phillips, B. L. Panaro, G. F. Lim, and M. J. Serra, "Non-nearest-neighbor dependence of stability for group III RNA single nucleotide bulge loops," *RNA* **20**, 825–834 (2014).
- ²⁷C. E. Longfellow, R. Kierzek, and D. H. Turner, "Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides," *Biochemistry* **29**, 278–285 (1990).
- ²⁸C. A. S. A. Minetti, D. P. Remeta, R. Dickstein, and K. J. Breslauer, "Energetic signatures of single base bulges: Thermodynamic consequences and biological implications," *Nucleic Acids Res.* **38**, 97–116 (2010).
- ²⁹D. R. Groebe and O. C. Uhlenbeck, "Thermal stability of RNA hairpins containing a four-membered loop and a bulge nucleotide," *Biochemistry* **28**, 742–747 (1989).
- ³⁰G. F. Lim, G. E. Merz, M. D. McCann, J. M. Gruskiewicz, and M. J. Serra, "Stability of single-nucleotide bulge loops embedded in a GAAA RNA hairpin stem," *RNA* **18**, 807–814 (2012).
- ³¹A. L. Dishler, E. L. McMichael, and M. J. Serra, "Determination of the secondary structure of group II bulge loops using the fluorescent probe 2-aminopurine," *RNA* **21**, 975–984 (2015).
- ³²J. C. Tomcho, M. R. Tillman, and B. M. Znosko, "Improved model for predicting the free energy contribution of dinucleotide bulges to RNA duplex stability," *Biochemistry* **54**, 5290–5296 (2015).
- ³³T. D. Amarante and G. Weber, "Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model," *J. Chem. Inf. Model.* **56**, 101–109 (2016).
- ³⁴G. Kalosakas and S. Ares, "Dependence on temperature and guanine-cytosine content of bubble length distributions in DNA," *J. Chem. Phys.* **130**, 235104 (2009).
- ³⁵G. Kalosakas, "Charge transport in DNA: Dependence of diffusion coefficient on temperature and electron-phonon coupling constant," *Phys. Rev. E* **84**, 051905 (2011).
- ³⁶N. N. Oiwa, C. E. Cordeiro, and D. W. Heermann, "The electronic behavior of zinc-finger protein binding sites in the context of the DNA extended ladder model," *Front. Phys.* **4**, 13 (2016).
- ³⁷A. E. B. Pupo, F. Falo, and A. Fiasconaro, "DNA overstretching transition induced by melting in a dynamical mesoscopic model," *J. Chem. Phys.* **139**, 095101 (2013).
- ³⁸M. Zoli, "J-factors of short DNA molecules," *J. Chem. Phys.* **144**, 214104 (2016).
- ³⁹G. Weber, "Sharp DNA denaturation due to solvent interaction," *Europhys. Lett.* **73**, 806–811 (2006).
- ⁴⁰A. Singh and N. Singh, "Phase diagram of mechanically stretched DNA: The salt effect," *Phys. A* **392**, 2052–2059 (2013).
- ⁴¹G. Weber, "Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures," *Nucleic Acids Res.* **41**, e30 (2013).
- ⁴²I. Ferreira, T. D. Amarante, and G. Weber, "DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration," *J. Chem. Phys.* **143**, 175101 (2015).
- ⁴³B. Alexandrov, V. Gelev, Y. Monisova, L. Alexandrov, A. Bishop, K. Rasmussen, and A. Usheva, "A nonlinear dynamic model of DNA with a sequence-dependent stacking term," *Nucleic Acids Res.* **37**, 2405 (2009).
- ⁴⁴T. Dauxois, M. Peyrard, and A. R. Bishop, "Entropy-driven DNA denaturation," *Phys. Rev. E* **47**, R44–R47 (1993).
- ⁴⁵N. Singh and Y. Singh, "Effect of defects on thermal denaturation of DNA oligomers," *Phys. Rev. E* **64**, 042901 (2001).
- ⁴⁶P. M. Morse, "Diatomic molecules according to the wave mechanics. II. Vibrational levels," *Phys. Rev.* **34**, 57 (1929).
- ⁴⁷Y.-L. Zhang, W.-M. Zheng, J.-X. Liu, and Y. Z. Chen, "Theory of DNA melting based on the Peyrard-Bishop model," *Phys. Rev. E* **56**, 7100–7115 (1997).
- ⁴⁸G. Weber, N. Haslam, N. Whiteford, A. Prügel-Bennett, J. W. Essex, and C. Neylon, "Thermal equivalence of DNA duplexes without melting temperature calculation," *Nat. Phys.* **2**, 55–59 (2006).
- ⁴⁹G. Weber, N. Haslam, J. W. Essex, and C. Neylon, "Thermal equivalence of DNA duplexes for probe design," *J. Phys.: Condens. Matter* **21**, 034106 (2009).
- ⁵⁰R. V. Maximiano and G. Weber, "Deoxyinosine mismatch parameters calculated with a mesoscopic model result in uniform hydrogen bonding and strongly variable stacking interactions," *Chem. Phys. Lett.* **631–632**, 87–91 (2015).
- ⁵¹G. Weber, "TfReg: Calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models," *Bioinformatics* **29**, 1345–1347 (2013).
- ⁵²L. Popenda, L. Bielecki, Z. Gdaniec, and R. W. Adamiak, "Structure and dynamics of adenosine bulged RNA duplex reveals formation of the dinucleotide platform in the C: GA triple," *Arkivoc* **3**, 130–144 (2009).

Referências Bibliográficas

- [1] Brent M Znosko, Sara B Silvestri, Heather Volkman, Bob Boswell, e Martin J Serra. Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges. *Biochemistry*, 41(33):10406–10417, 2002.
- [2] Joshua M Blose, Michelle L Manni, Kelly A Klapac, Yukiko Stranger-Jones, Allison C Zyra, Vasiliy Sim, Chad A Griffith, Jason D Long, e Martin J Serra. Non-nearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops. *Biochemistry*, 46(51):15123–15135, 2007.
- [3] Michael D McCann, Geoffery FS Lim, Michelle L Manni, Julie Estes, Kelly A Klapac, Gregory D Frattini, Robert J Knarr, Jessica L Gratton, e Martin J Serra. Non-nearest-neighbor dependence of the stability for RNA group II single-nucleotide bulge loops. *RNA*, 17(1):108–119, 2011.
- [4] Jessica L Kent, Michael D McCann, Daniel Phillips, Brandon L Panaro, Geoffrey FS Lim, e Martin J Serra. Non-nearest-neighbor dependence of stability for group III RNA single nucleotide bulge loops. *RNA*, 20(6):825–834, 2014.
- [5] T. Dauxois, M. Peyrard, e A. R. Bishop. Entropy-driven DNA denaturation. *Phys. Rev. E*, 47(1):R44–R47, 1993.
- [6] M. Peyrard e A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62(23):2755–2757, 1989.
- [7] James D Watson *et al.* Molecular biology of the gene. *Molecular biology of the gene.*, (2nd edn), 1970.
- [8] P Strathern e Watson Crick. o dna em 90 minutos, 2001.
- [9] MHF Wilkins, RG Gosling, e WE Seeds. Physical studies of nucleic acid: Nucleic acid: an extensible molecule? *Nature*, 167(4254):759–760, 1951.

- [10] Rosalind E Franklin e Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.
- [11] Xiang-Jun Lu e Wilma K Olson. 3dna: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research*, 31(17): 5108–5121, 2003.
- [12] Gerald S. Manning. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly Reviews of Biophysics*, 11(2): 179–246, 1978. doi: 10.1017/S0033583500002031.
- [13] URL <http://ndbserver.rutgers.edu/service/ndb/atlas/summary?searchTarget=5IZP>.
- [14] Michel Peyrard. Nonlinear dynamics and statistical physics of dna. *Nonlinearity*, 17(2): R1, 2004.
- [15] Chris R Calladine e Horace Drew. *Understanding DNA: the molecule and how it works*. Academic press, 1997.
- [16] URL <http://www.ead.hemocentro.fmrp.usp.br/joomla/index.php/publicacoes/folhetins/469-dna-o-sentido-da-vida>.
- [17] Eric Westhof. Isostericity and tautomerism of base pairs in nucleic acids. *FEBS letters*, 588(15):2464–2469, 2014.
- [18] URL <https://www.rcsb.org/pdb/home/home.do>.
- [19] Neocles B Leontis e Eric Westhof. Conserved geometrical base-pairing patterns in rna. *Quarterly reviews of biophysics*, 31(4):399–455, 1998.
- [20] Tauanne D. Amarante e Gerald Weber. Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model. *J. Chem. Inf. Model.*, 56(1):101–109, 2016. doi: 10.1021/acs.jcim.5b00571. URL <http://dx.doi.org/10.1021/acs.jcim.5b00571>.
- [21] Kathleen P Howard. Thermodynamics of dna duplex formation: A biophysical chemistry laboratory experiment. *J. Chem. Educ*, 77(11):1469, 2000.

- [22] Naoki Sugimoto, Shu-ichi Nakano, Misa Katoh, Akiko Matsumura, Hiroyuki Nakamuta, Tatsuo Ohmichi, Mari Yoneyama, e Muneo Sasaki. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochem.*, 34(35):11211–11216, 1995.
- [23] Lauren A Levine, Matthew Junker, Myranda Stark, e Dustin Greenleaf. A DNA melting exercise for a large laboratory class. *Journal of Chemical Education*, 92:1928–1931, 2015. doi: 10.1021/acs.jchemed.5b00049.
- [24] Izabela Ferreira, Tauanne D. Amarante, e Gerald Weber. DNA terminal base pairs have weaker hydrogen bonds especially for AT under low salt concentration. *J. Chem. Phys.*, 143:175101, 2015. doi: 10.1063/1.4934783.
- [25] M Ageno, E Dore, e C Frontali. The alkaline denaturation of dna. *Biophysical journal*, 9(11):1281–1311, 1969.
- [26] Ross B Inman e Maria Schnös. Partial denaturation of thymine-and 5-bromouracil-containing λ dna in alkali. *Journal of molecular biology*, 49(1):931–969, 1970.
- [27] Sherrie Schreiber-Gosche e Robert A Edwards. Thermodynamics of oligonucleotide duplex melting. *Journal of Chemical Education*, 86(5):644, 2009.
- [28] Grégoire Altan-Bonnet, Albert Libchaber, e Oleg Krichevsky. Bubble dynamics in double-stranded DNA. *Physical Review Letters*, 90(13):138101, 2003.
- [29] A Monrichok, G Gruner, e G Zocchi. Trapping intermediates in the melting transition of DNA oligomers. *EPL (Europhysics Letters)*, 62(3):452, 2003.
- [30] Yan Zeng, Awrasa Monrichok, e Giovanni Zocchi. Bubble nucleation and cooperativity in dna melting. *Journal of molecular biology*, 339(1):67–75, 2004.
- [31] Garima Mishra, Debaprasad Giri, MS Li, e Sanjay Kumar. Role of loop entropy in the force induced melting of dna hairpin. *The Journal of chemical physics*, 135(3):07B611, 2011.
- [32] Gerald Weber, Niall Haslam, Jonathan W. Essex, e 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.

- [33] Tianbing Xia, John SantaLucia, Jr., Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, e 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.
- [34] 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>.
- [35] John SantaLucia, Jr., H T Allawi, e P A Seneviratne. Improved nearest-neighbour parameters for predicting DNA duplex stability. *Biochem.*, 35:3555–3562, 1996.
- [36] Ilyas Yildirim e Douglas H Turner. RNA challenges for computational chemists. *Biochem.*, 44(40):13225–13234, 2005.
- [37] BS Alexandrov, LT Wille, KØ Rasmussen, AR Bishop, e KB Blagoev. Bubble statistics and dynamics in double-stranded dna. *Physical review E*, 74(5):050901, 2006.
- [38] Nikos Theodorakopoulos. Dna denaturation bubbles at criticality. *Physical Review E*, 77(3):031919, 2008.
- [39] Jayme De Luca, Elso Drigo Filho, Antonio Ponso, e José Roberto Ruggiero. Energy localization in the peyrard-bishop dna model. *Physical Review E*, 70(2):026213, 2004.
- [40] Gerald Weber, Jonathan W. Essex, e Cameron Neylon. Probing the microscopic flexibility of DNA from melting temperatures. *Nat. Phys.*, 5:769–773, 2009. doi: 10.1038/nphys1371.
- [41] Philip M Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Physical Review*, 34(1):57, 1929.
- [42] Titus S. van Erp, Santiago Cuesta-Lopez, Johannes-Geert Hagmann, e Michel Peyrard. van Erp et al. Reply:. *Phys. Rev. Lett.*, 96(23):239802, 2006. doi: 10.1103/PhysRevLett.96.239802. URL <http://link.aps.org/abstract/PRL/v96/e239802>.

- [43] Marc Joyeux e Sahin Buyukdagli. Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous DNA thermal denaturation. *Phys. Rev. E*, 72:051902, 2005.
- [44] Yong-Li Zhang, Wei-Mou Zheng, Ji-Xing Liu, e Y. Z. Chen. Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E*, 56(6):7100–7115, 1997.
- [45] JA Krumhansl e JR Schrieffer. Dynamics and statistical mechanics of a one-dimensional model hamiltonian for structural phase transitions. *Physical Review B*, 11(9):3535, 1975.
- [46] Maria Barbi, Simona Cocco, e Michel Peyrard. Helicoidal model for dna opening. *Physics Letters A*, 253(5):358–369, 1999.
- [47] Simona Cocco e Rémi Monasson. Statistical mechanics of torque induced denaturation of dna. *Physical Review Letters*, 83(24):5178, 1999.
- [48] Marc Joyeux e Sahin Buyukdagli. Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous dna thermal denaturation. *Physical Review E*, 72(5):051902, 2005.
- [49] Gerald Weber. Sharp DNA denaturation due to solvent interaction. *Europhys. Lett.*, 73(5):806–811, 2006. doi: 10.1209/epl/i2005-10466-6.
- [50] Sahin Buyukdagli e Marc Joyeux. Theoretical investigation of finite size effects at dna melting. *Physical Review E*, 76(2):021917, 2007.
- [51] Marc Joyeux e Ana-Maria Florescu. Dynamical versus statistical mesoscopic models for dna denaturation. *Journal of Physics: Condensed Matter*, 21(3):034101, 2008.
- [52] Michel Peyrard, Santiago Cuesta-Lopez, e Dimitar Angelov. Experimental and theoretical studies of sequence effects on the fluctuation and melting of short dna molecules. *Journal of Physics: Condensed Matter*, 21(3):034103, 2008.
- [53] S Behnia, M Panahi, A Akhshani, e A Mobaraki. Mean lyapunov exponent approach for the helicoidal peyrard–bishop model. *Physics Letters A*, 375(41):3574–3578, 2011.
- [54] Slobodan Zdravković. Helicoidal peyrard–bishop model of dna dynamics. *Journal of Non-linear Mathematical Physics*, 18(supp02):463–484, 2011.

- [55] Gerald Weber, Niall Haslam, Nava Whiteford, Adam Prügel-Bennett, Jonathan W. Essex, e Cameron Neylon. Thermal equivalence of DNA duplexes without melting temperature calculation. *Nat. Phys.*, 2:55–59, 2006. doi: 10.1038/nphys189.
- [56] Richard Owczarzy, Yong You, Bernardo G. Moreira, Jeffrey A. Manthey, Lingyan Huang, Mark A. Behlke, e Joseph A. Walder. Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochem.*, 43:3537–3554, 2004.
- [57] 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>.
- [58] Viachaslau Bernat e Matthew D Disney. Rna structures as mediators of neurological diseases and as drug targets. *Neuron*, 87(1):28–46, 2015.
- [59] 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>.
- [60] K Maleknejad, S Sohrabi, e B Baranji. Application of 2d-bpfs to nonlinear integral equations. *Communications in Nonlinear Science and Numerical Simulation*, 15(3):527–535, 2010.
- [61] David W Staple e Samuel E Butcher. Solution structure of the hiv1 frameshift inducing stem-loop rna. *Nucleic acids research*, 31(15):4326–4331, 2003.
- [62] Harleen Kaur, Jesper Wengel, e Souvik Maiti. Thermodynamics of dna- rna heteroduplex formation: effects of locked nucleic acid nucleotides incorporated into the dna strand. *Biochemistry*, 47(4):1218–1227, 2008.
- [63] Douglas H Turner. Bulges in nucleic acids. *Current Opinion in Structural Biology*, 2(3): 334–337, 1992.
- [64] Robin R Gutell, JJ Cannone, Z Shang, Y Du, e MJ Serra. A story: unpaired adenosine bases in ribosomal rnas. *Journal of molecular biology*, 304(3):335–354, 2000.

- [65] Thomas Hermann e Dinshaw J Patel. RNA bulges as architectural and recognition motifs. *Structure*, 8(3):R47–R54, 2000.
- [66] Debra A Peattie, Steve Douthwaite, Roger A Garrett, e Harry F Noller. A "bulged" double helix in a rna-protein contact site. *Proceedings of the National Academy of Sciences*, 78(12):7331–7335, 1981.
- [67] Ignacio Tinoco, Olke C Uhlenbeck, e Mark D Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230(5293):362–367, 1971.
- [68] Lukasz Popenda, Ryszard W Adamiak, e Zofia Gdaniec. Bulged adenosine influence on the RNA duplex conformation in solution. *Biochemistry*, 47(18):5059–5067, 2008.
- [69] Lukasz Popenda, Lukasz Bielecki, Zofia Gdaniec, e Ryszard W Adamiak. Structure and dynamics of adenosine bulged rna duplex reveals formation of the dinucleotide platform in the c: Ga triple. *ARKIVOC: Online Journal of Organic Chemistry*, 2009.
- [70] André Barthel e Martin Zacharias. Conformational transitions in RNA single uridine and adenosine bulge structures: a molecular dynamics free energy simulation study. *Biophysical Journal*, 90(7):2450–2462, 2006.
- [71] Yvonne Th van den Hoogen, Aart A van Beuzekom, Erik de Vroom, Gijs A van der Marel, Jacques H van Boom, e Cornelis Altona. Bulge-out structures in the single-stranded trimer AUA and in the duplex (CUGGUGCGG).(CCGCCAG). a model-building and NMR study. *Nucleic Acids Res.*, 16(11):5013–5030, 1988.
- [72] Whitney A Hastings, Yaroslava G Yingling, Gregory S Chirikjian, e Bruce A Shapiro. Structural and dynamical classification of RNA single-base bulges for nanostructure design. *Journal of Computational and Theoretical Nanoscience*, 3(1):63–77, 2006. doi: 10.1166/jctn.2006.005.
- [73] Jeffrey W Nelson, Francis H Martin, e Ignacio Tinoco. DNA and RNA oligomer thermodynamics: The effect of mismatched bases on double-helix stability. *Biopolymers*, 20(12):2509–2531, 1981.

- [74] Susan A White e David E Draper. Single base bulges in small RNA hairpins enhance ethidium binding and promote an allosteric transition. *Nucleic Acids Research*, 15(10): 4049–4064, 1987.
- [75] Susan A White e David E Draper. Effects of single-base bulges on intercalator binding to small RNA and DNA hairpins and a ribosomal RNA fragment. *Biochemistry*, 28(4): 1892–1897, 1989.
- [76] Jeffrey W Nelson e Ignacio Tinoco Jr. Ethidium ion binds more strongly to a DNA double helix with a bulged cytosine than to a regular double helix. *Biochemistry*, 24(23):6416–6421, 1985.
- [77] E Ennifar, M Yusupov, P Walter, R Marquet, B Ehresmann, C Ehresmann, e P Dumas. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, 7(11):1439–1449, 1999.
- [78] YONG Xiong e MUTTAIYA Sundaralingam. Two crystal forms of helix II of xenopus laevis 5S rRNA with a cytosine bulge. *RNA*, 6(9):1316–1324, 2000.
- [79] Yong Xiong, Junpeng Deng, Chellappanpillai Sudarsanakumar, e Muttaiya Sundaralingam. Crystal structure of an RNA duplex r(gugucgcac)₂ with uridine bulges. *Journal of Molecular Biology*, 313(3):573–582, 2001.
- [80] Leemor Joshua-Tor, Dov Rabinovich, Håkon Hope, Felix Frolow, Ettore Appella, e Joel L Sussman. The three-dimensional structure of a DNA duplex containing looped-out bases. *Nature*, 334(6177):82–84, 1988.
- [81] Darryl A LeBlanc e Kathleen M Morden. Thermodynamic characterization of deoxyribo-oligonucleotide duplexes containing bulges. *Biochemistry*, 30(16):4042–4047, 1991.
- [82] Hervé Beltz, Joel Azoulay, Serena Bernacchi, Jean-Pierre Clamme, Damien Ficheux, Bernard Roques, Jean-Luc Darlix, e Yves Mély. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *Journal of Molecular Biology*, 328(1):95–108, 2003.

- [83] Fumiaki Tanaka, Atsushi Kameda, Masahito Yamamoto, e Azuma Ohuchi. Thermodynamic parameters based on a nearest-neighbor model for DNA sequences with a single-bulge loop. *Biochemistry*, 43(22):7143–7150, 2004.
- [84] Helen J Merianos, JIMIN WANG, e PETER B MOORE. The structure of a ribosomal protein s8/spc operon mrna complex. *Rna*, 10(6):954–964, 2004.
- [85] Jian Zhu e Roger M Wartell. The effect of base sequence on the stability of RNA and DNA single base bulges. *Biochemistry*, 38(48):15986–15993, 1999.
- [86] Carl E Longfellow, Ryszard Kierzek, e Douglas H Turner. Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry*, 29(1):278–285, 1990.
- [87] Conceição A. S. A. Minetti, David P. Remeta, Rian Dickstein, e Kenneth J. Breslauer. Energetic signatures of single base bulges: thermodynamic consequences and biological implications. *Nucleic Acids Res.*, 38(1):97–116, 2010.
- [88] Duncan R Groebe e Olke C Uhlenbeck. Thermal stability of RNA hairpins containing a four-membered loop and a bulge nucleotide. *Biochemistry*, 28(2):742–747, 1989.
- [89] Geoffrey FS Lim, Gregory E Merz, Michael D McCann, Jocelyn M Gruskiewicz, e Martin J Serra. Stability of single-nucleotide bulge loops embedded in a GAAA RNA hairpin stem. *RNA*, 18(4):807–814, 2012.
- [90] Abigael L Dishler, Elizabeth L McMichael, e Martin J Serra. Determination of the secondary structure of group II bulge loops using the fluorescent probe 2-aminopurine. *RNA*, 21(5):975–984, 2015.
- [91] Jeremy C Tomcho, Magdalena R Tillman, e Brent M Znosko. Improved model for predicting the free energy contribution of dinucleotide bulges to RNA duplex stability. *Biochemistry*, 54(34):5290–5296, 2015.
- [92] IGNACIO TINOCO JUN, Philip N Borer, Barbara Dengler, Mark D Levine, Olke C Uhlenbeck, Donald M Crothers, e Jay Gralla. Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology*, 246(150):40, 1973.

- [93] Michael Zuker e Patrick Stiegler. Optimal computer folding of large rna sequences using thermodynamics and auxiliary information. *Nucleic acids research*, 9(1):133–148, 1981.
- [94] Douglas H Turner, Naoki Sugimoto, e Susan M Freier. Rna structure prediction. *Annual review of biophysics and biophysical chemistry*, 17(1):167–192, 1988.
- [95] Michael Zuker. On finding all suboptimal foldings of an rna molecule. *Science*, 244(4900):48–52, 1989.
- [96] Martin J Serra e Douglas H Turner. [11] predicting thermodynamic properties of rna. In *Methods in enzymology*, volume 259, pages 242–261. Elsevier, 1995.
- [97] David H Mathews, Jeffrey Sabina, Michael Zuker, e Douglas H Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of rna secondary structure1. *Journal of molecular biology*, 288(5):911–940, 1999.
- [98] Petr Šulc, Flavio Romano, Thomas E Ouldridge, Jonathan PK Doye, e Ard A Louis. A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.*, 140(23):235102, 2014. doi: 10.1063/1.4881424.
- [99] Rodolfo Vieira Maximiano e Gerald Weber. Deoxyinosine mismatch parameters calculated with a mesoscopic model result in uniform hydrogen bonding and strongly variable stacking interactions. *Chem. Phys. Lett.*, 631–632:87–91, 2015. doi: 10.1016/j.cplett.2015.04.045.