

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Instituto de Ciências Exatas**  
**Programa de Pós-graduação em Física**

Gabriel Henrique Aguiar Schiess

**ESTUDO DE MODIFICAÇÕES DE RNA ATRAVÉS DE MODELOS  
MESOSCÓPICOS:  
Inosina e Pseudouridina**

Belo Horizonte  
2025

Gabriel Henrique Aguiar Schiess

**ESTUDO DE MODIFICAÇÕES DE RNA ATRAVÉS DE  
MODELOS MESOSCÓPICOS:  
Inosina e Pseudouridina**

Dissertação apresentada ao Programa de Pós-Graduação do Departamento de Física do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Física.

Orientador: Gerald Weber

Coorientadora: Pâmella Miranda de Moura

Belo Horizonte  
2025

Dados Internacionais de Catalogação na Publicação (CIP)

S332e Schiess, Gabriel Henrique Aguiar.  
Estudo de modificações de RNA através de modelos mesoscópicos: Inosina e Pseudouridina / Gabriel Henrique Aguiar Schiess. – 2025.  
92 f. : il.

Orientador: Gerald Weber.  
Coorientadora: Pâmella Miranda de Moura.  
Dissertação (mestrado) – Universidade Federal de Minas Gerais,  
Departamento de Física.  
Bibliografia: f. 64 -71.

1. Ácido ribonucleico. I. Título. II. Weber, Gerald. III. Moura, Pâmella Miranda de. IV. Universidade Federal de Minas Gerais, Departamento de Física.

CDU – 577.2:53 (043)



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

**FOLHA DE APROVAÇÃO**

A presente dissertação, intitulada "**Estudo de modificações de RNA através de modelos mesoscópicos: Inosina e Pseudouridina**" de autoria de **GABRIEL HENRIQUE AGUIAR SCHIESS** submetida à Comissão Examinadora, abaixo-assinada, foi aprovada para obtenção do grau de **MESTRE EM FÍSICA, área de concentração Física Aplicada e Interdisciplinar**, em 07 de fevereiro de 2025.

Belo Horizonte, 07 de fevereiro de 2025.

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

Pâmella Miranda de Moura  
Coorientadora do estudante  
Karolinska Institutet, Suécia

Prof. Ubirajara Agero Batista  
Departamento de Física/UFMG

Prof. Erik de Oliveira Martins  
Instituto Federal do Norte de Minas Gerais



Documento assinado eletronicamente por **Gerald Weber, Professor do Magistério Superior**, em 11/02/2025, às 14:44, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Erik de Oliveira Martins, Usuário Externo**, em 11/02/2025, às 20:17, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Pâmella Miranda de Moura, Usuário Externo**, em 12/02/2025, às 04:26, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ubirajara Agero Batista, Servidor(a)**, em 12/02/2025, às 08:54, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3949319** e o código CRC **F26B9240**.

---

Referência: Processo nº 23072.208410/2025-15

SEI nº 3949319

# AGRADECIMENTOS

Gostaria de agradecer a todas as pessoas incríveis que me acompanharam, incentivaram e inspiraram nesta trajetória.

Agradeço ao Gerald, por aceitar me orientar e compartilhar seus conhecimentos comigo. Obrigado por sua paciência, apoio, compreensão e por seu tempo. Agradeço por seus conselhos, sinceridade e, em especial, agradeço por suas excelentes piadas que sempre me arrancam risadas. Sua orientação foi fundamental para meu crescimento acadêmico e pessoal.

Agradeço à Pâmella, por aceitar me orientar e acompanhar desde a iniciação científica. Obrigado, por me ensinar e aconselhar nesse caminho. Agradeço por suas revisões, pelos puxões de orelha, por suas críticas construtivas e pela paciência. Seus conselhos contribuíram enormemente para o meu desenvolvimento, em especial, da minha forma de escrever e organizar.

Agradeço à Iza, por ser uma pessoa incrível, por me inspirar infinitamente, por estar sempre disposta a discutir ciência, por rir das minhas piadas e por ser uma cientista extraordinária. Obrigado por estar sempre ao meu lado, me apoiando e incentivando.

Agradeço à minha mãe, ao meu pai e ao meu irmão, pelo amor, apoio e companhia durante todo este percurso.

Agradeço ao meu grupo de pesquisa, por me acolherem, pelas trocas de artigos, pelas críticas essenciais para o desenvolvimento dos trabalhos e por serem grandes referências acadêmicas e profissionais.

Agradeço às pessoas maravilhosas que tenho o prazer de chamar de amigos, por sempre estarem ao meu lado, por todo seu apoio, pelas risadas e momentos que compartilhamos.

Agradeço à professora Kira Astakhova, pela nossa colaboração e por sua contribuição para esse trabalho.

Agradeço a todos os meus professores, que contribuíram imensamente para o meu aprendizado e crescimento como pesquisador.

Agradeço à secretaria do Programa de Pós-Graduação em Física pelo excelente trabalho e agilidade no auxílio na resolução de nossas demandas. Em especial, agradeço à Ana, à Marília e à Adriana.

Finalmente, agradeço às agências de fomento Agradeço ao CNPq, FAPEMIG e CAPES. Em especial, agradeço ao CNPq pela minha bolsa de mestrado, e à FAPEMIG pelo fomento meu grupo de pesquisa.

## RESUMO

Modificações de RNA são pequenas alterações químicas que ocorrem em suas bases nitrogenadas, levando a mudanças das propriedades físicas, químicas e estruturais. Atualmente, mais de 150 modificações de ocorrência natural em RNA são conhecidas. Essas modificações são muitas vezes essenciais para o funcionamento correto de diversos mecanismos celulares e para o desenvolvimento correto de organismos complexos. Além disso, os ácidos nucleicos modificados podem ser usados em uma ampla variedade de técnicas biotecnológicas, como RT-PCR, protocolos de entrega de medicamentos e vacinas de mRNA. A inosina (I) é resultado da deaminação de uma adenina (A) devido à atuação de três enzimas da família adenosina deaminase agindo em RNA (ADAR). A inosina em mRNA é, usualmente, reconhecida pelo ribossomo como uma guanosina (G), o que pode resultar na tradução não sinônima de proteínas, sendo um fator de variabilidade epigenética. A presença irregular de inosina na molécula de RNA está relacionada a várias doenças, em especial doenças cognitivas degenerativas, como o Alzheimer. A pseudouridina ( $\Psi$ ) é um isômero da uridina (U) e a modificação mais frequentemente encontrada na molécula de RNA, estando presente no mRNA, tRNA e em outros tipos de RNA. Sua ocorrência é resultado da atuação de diversas enzimas que atuam de forma altamente específica nas sequências de RNA. Uma das principais funções da pseudouridina é estabilizar estruturas secundárias de RNAs complexos, como o tRNA e o rRNA. A N1-metilpseudouridina ( $m^1\Psi$ ) é o resultado da metilação de pseudouridina e tem sido amplamente utilizada em vacinas de mRNA como um substituto da uridina. Neste trabalho, utilizamos o modelo de próximos vizinhos (NN) e o modelo mesoscópico Peyrard-Bishop (PB) para estudar a contribuição da inosina, pseudouridina e N1-metilpseudouridina para a estabilidade termodinâmica de duplexos de RNA. O modelo NN descreve essa estabilidade pelas contribuições energéticas de cada par de bases, enquanto que o modelo PB, um modelo de física-estatística, considera as ligações de hidrogênio e do empilhamento entre os pares de bases e seus vizinhos para a estabilidade termodinâmica do duplexo. Parâmetros referentes à essas interações foram obtidos para sequências de RNA modificado. Foram analisadas temperaturas de desnaturação de 71 sequências de RNA contendo inosina, 41 com pseudouridina e 9 com N1-metilpseudouridina. Nossos resultados sugerem que essas modificações são capazes de estabilizar o duplexo de RNA. Essa estabilidade proporcionada pelas modificações é, geralmente, dependente do contexto da sequência em que estão inseridas, podendo contribuir negativamente ou positivamente para a estabilidade do duplexo. A inosina mostra uma tendência de desestabilizar o duplexo. Em especial, o par IU apresenta resultados que indicam a presença de apenas uma ligação de hidrogênio. Em geral, os pares com pseudouridina atuam no sentido de estabilizar o duplexo. A N1-metilpseudouridina atua de forma similar à pseudouridina e apresenta uma interação mais forte com os vizinhos. Os parâmetros termodinâmicos obtidos para essas três modificações em RNA podem contribuir, com o cálculo de temperaturas de desnaturação e dos perfis de abertura, para o design de sondas em diagnósticos, e também no desenvolvimento de vacinas baseadas em RNA.

Palavras-chave: RNA modificado; Modelos mesoscópicos; Temperatura de desnaturação;  
Bases modificadas.

## ABSTRACT

RNA modifications are small chemical changes that occur in their nitrogenous bases, leading to alterations in physical, chemical, and structural properties. Currently, more than 150 naturally occurring RNA modifications are known. These modifications are often essential for the proper functioning of various cellular mechanisms and for the correct development of complex organisms. Furthermore, modified nucleic acids can be used in a wide range of biotechnological techniques, such as RT-PCR, drug delivery protocols, and mRNA vaccines. Inosine (I) results from the deamination of adenine (A) due to the action of three enzymes from the adenosine deaminase family acting on RNA (ADAR). In mRNA, inosine is usually recognized by the ribosome as guanosine (G), which can result in non-synonymous protein translation, contributing to epigenetic variability. The irregular presence of inosine in the RNA molecule is associated with several diseases, especially neurodegenerative cognitive diseases such as Alzheimer's. Pseudouridine ( $\Psi$ ) is an isomer of uridine (U) and the most frequently found modification in RNA molecules, being present in mRNA, tRNA, and other RNA types. Its occurrence results from the action of various enzymes that act highly specifically on RNA sequences. One of the main functions of pseudouridine is to stabilize the secondary structures of complex RNAs, such as tRNA and rRNA. N1-methylpseudouridine ( $m^1\Psi$ ) results from the methylation of pseudouridine and has been widely used in mRNA vaccines as a substitute for uridine. In this work, we used the nearest-neighbor (NN) model and the Peyrard-Bishop (PB) mesoscopic model to study the contribution of inosine, pseudouridine, and N1-methylpseudouridine to the thermodynamic stability of RNA duplexes. The NN model describes this stability through the energetic contributions of each base pair. In contrast, the PB model, a statistical physics model, considers hydrogen bonds and stacking interactions between base pairs and their neighbors for the thermodynamic stability of the duplex. Parameters related to these interactions were obtained for modified RNA sequences. Melting temperatures of 71 RNA sequences containing inosine, 41 with pseudouridine, and 9 with N1-methylpseudouridine were analyzed. Our results suggest that these modifications can stabilize the RNA duplex. This stability provided by the modifications generally depends on the sequence context in which they are inserted, potentially contributing negatively or positively to duplex stability. Inosine shows a tendency to destabilize the duplex. In particular, the IU pair shows the presence of only one hydrogen bond. In general, pairs with pseudouridine act to stabilize the duplex. N1-methylpseudouridine behaves similarly to pseudouridine and shows stronger interaction with neighbors. The thermodynamic parameters obtained for these three RNA modifications may contribute, using the calculation of melting temperatures and opening profiles, to the design of probes in diagnostics and also the development of RNA-based vaccines.

Keywords: Modified RNA; Mesoscopic models; Melting temperature; Modified bases.

## LISTA DE FIGURAS

1.1	Estrutura das bases nitrogenadas . . . . .	15
1.2	Estrutura do duplexo de DNA . . . . .	16
1.3	Determinação da temperatura de desnaturação por absorção ultravioleta . . . . .	17
1.4	Gráfico de Van't Hoff . . . . .	18
1.5	Estrutura da pseudouridina e da N1-metilpseudouridina . . . . .	20
1.6	Pseudouridina em tRNA . . . . .	21
1.7	Ligações de hidrogênio de pares modificados com inosina . . . . .	22
3.1	Exemplo de aplicação do modelo NN . . . . .	26
3.2	Distribuição normal da função partição . . . . .	31
3.3	Dependência de $\omega$ com a temperatura . . . . .	31
3.4	Regressão linear para $T_m$ . . . . .	32
3.5	Transformações do método Nelder-Mead . . . . .	34
4.1	Potencial de Morse pares com Inosina . . . . .	42
4.2	Parâmetros de <i>stacking</i> para pares IC, IG, II e IU . . . . .	43
4.3	Perfil de abertura da região modificada do mRNA do gene RHOA . . . . .	47
4.4	Perfis de abertura de códons contendo inosina . . . . .	48
5.1	Potencial de Morse dos pares de $\Psi$ e $m^1\Psi$ . . . . .	53
5.2	Pontes de água . . . . .	54
5.3	Parâmetros de <i>stacking</i> para pares contendo $\Psi$ e $m^1\Psi$ . . . . .	56
5.4	Perfil de abertura $\Psi$ e $m^1\Psi$ . . . . .	58

## LISTA DE TABELAS

3.1	Parâmetros para o modelo NN em DNA . . . . .	27
3.2	Parâmetros para o modelo NN em RNA . . . . .	27
4.1	Evolução de $\chi^2$ para inosina . . . . .	38
4.2	Morse da inosina em DNA . . . . .	44
4.3	Sequências do gene RHOA . . . . .	46
4.4	Abertura média dos códons . . . . .	49
5.1	Evolução de $\chi^2$ para $\Psi$ . . . . .	51
5.3	Sequência para estudos de vacinas de mRNA . . . . .	59
A.1	Sequências de RNA contendo inosina . . . . .	72
A.2	Sequências de referência . . . . .	75
A.3	Regras de contexto . . . . .	78
A.4	Parâmetros de Morse para inosina . . . . .	79
A.5	Parâmetros harmônicos para inosina . . . . .	80
A.6	Parâmetros iniciais de contexto . . . . .	81
B.1	Sequências de RNA contendo pseudouridina e N1-metilpseudouridina . . . . .	82
B.2	Sequências de referência . . . . .	85
B.3	Regras de contexto . . . . .	88
B.4	Parâmetros de Morse para Pseudouridina e N1-metilpseudouridina . . . . .	89
B.5	Parâmetros harmônicos para Pseudouridina e N1-metilpseudouridina . . . . .	90
B.6	Parâmetros obtidos para $\Psi$ A sem contexto . . . . .	91
B.7	Parâmetros obtidos para $m^1\Psi$ A em concentração das fitas de $10 \mu\text{M}$ . . . . .	92

# SÚMARIO

<b>1</b>	<b>Introdução</b>	<b>14</b>
1.1	O que são DNA e RNA . . . . .	14
1.2	Desnaturação de ácidos nucleicos . . . . .	16
1.3	Bases modificadas . . . . .	19
1.3.1	Pseudouridina . . . . .	19
1.3.2	Inosina . . . . .	21
<b>2</b>	<b>Objetivos</b>	<b>24</b>
2.1	Objetivo geral . . . . .	24
2.2	Objetivos Específicos . . . . .	24
<b>3</b>	<b>Métodos</b>	<b>25</b>
3.1	Modelo de próximos vizinhos . . . . .	25
3.2	Modelo Peyrard-Bishop . . . . .	28
3.3	Obtendo temperaturas de <i>melting</i> com o modelo PB . . . . .	29
3.4	Algoritmo de minimização para obtenção de parâmetros . . . . .	32
3.5	Lidando com poucos dados . . . . .	35
<b>4</b>	<b>Inosina</b>	<b>37</b>
4.1	Introdução . . . . .	37
4.2	Procedimentos de minimização . . . . .	37
4.2.1	Grupo IC . . . . .	38
4.2.2	Grupo IG . . . . .	39
4.2.3	Grupo II . . . . .	39
4.2.4	Grupo IU . . . . .	39
4.2.5	Desvio padrão . . . . .	40
4.3	Resultados . . . . .	40
4.4	Inosina em DNA e RNA . . . . .	42
4.5	Perfis de abertura e contextos biológicos . . . . .	45

4.6	Conclusão . . . . .	49
<b>5</b>	<b>Pseudouridina</b>	<b>50</b>
5.1	Introdução . . . . .	50
5.2	Procedimentos de minimização . . . . .	50
5.3	Design das sequências de N1-metilpseudouridina . . . . .	51
5.3.1	Grupo $\Psi$ A sem contexto . . . . .	52
5.3.2	Desvio padrão . . . . .	52
5.4	Resultados . . . . .	53
5.5	Perfis de abertura e vacinas de mRNA . . . . .	57
5.6	Conclusão . . . . .	59
<b>6</b>	<b>Conclusão</b>	<b>61</b>
<b>7</b>	<b>Perspectivas</b>	<b>63</b>
	<b>REFERÊNCIAS</b>	<b>64</b>
	<b>APÊNDICE A - Dados utilizados e resultados obtidos para inosina em RNA</b>	<b>72</b>
	<b>APÊNDICE B - Dados utilizados e resultados obtidos para pseudouridina em RNA</b>	<b>82</b>

# 1 Introdução

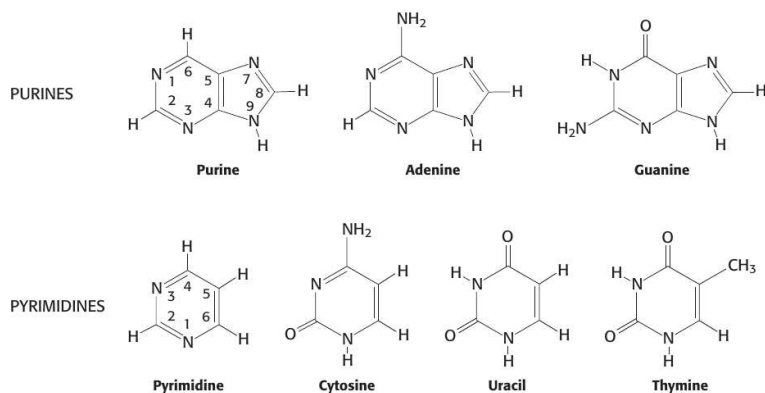
Nesta seção, apresentaremos a composição e estrutura do DNA e do RNA; os experimentos de desnaturação e sua importância; também as bases modificadas, em especial, a inosina e a pseudouridina que foram estudadas neste trabalho.

## 1.1 O que são DNA e RNA

Os estudos das moléculas de ácido desoxirribonucleico (*deoxyribonucleic acid*, DNA) e ácido ribonucleico (*ribonucleic acid*, RNA) têm levado, ao longo dos anos, à diversas descobertas e aplicações biológicas. O conhecimento científico nessa área tem avançado de forma cada vez mais rápida. Esse avanço se deve principalmente à implementação dos métodos de bioinformática. O avanço científico e tecnológico na área é responsável por diversos métodos amplamente utilizados em detecção de agentes patógenos, diagnósticos preventivos de doenças hereditárias ou associadas à expressão genética e, mais recentemente, desenvolvimento de vacinas [1]. O DNA e o RNA são biomoléculas que contêm informações genéticas dos seres vivos. Os ácidos nucleicos são responsáveis, principalmente, por codificar e mediar a produção das proteínas.

O DNA é a base da informação genética que é codificada através das bases nitrogenadas adenina (A), citosina (C), guanina (G) e timina (T), Figura 1.1. As bases são divididas em dois principais grupos, as purinas A e G e as pirimidinas C e T. A sequência das bases em uma cadeia de DNA determina a sequência e quais aminoácidos que serão incorporados à cadeia proteica. O RNA, por sua vez, é o mecanismo pelo qual a informação contida no DNA é expressa no meio celular. Existem diversos tipos de RNA presentes no citoplasma de célula, os principais são o RNA de transferência (*transfer RNA*, tRNA), o RNA mensageiro (*messenger RNA*, mRNA) e o RNA ribossomal (*ribosomal RNA*, rRNA). O RNA é composto pelas mesmas bases nitrogenadas que o DNA, exceto pela timina que é substituída pela uracila (U), ver Figura 1.1. A síntese proteica é determinada, majoritariamente, pela sequência de DNA correspondente a uma cadeia proteica. Para que esse processo ocorra, a sequência de DNA deve ser transcrita em uma sequência de RNA equivalente. Essa transcrição ocorre no núcleo celular através da mediação de uma enzima denominada RNA polimerase que é responsável por produzir, a partir das informações do DNA, todos os tipos de RNA.

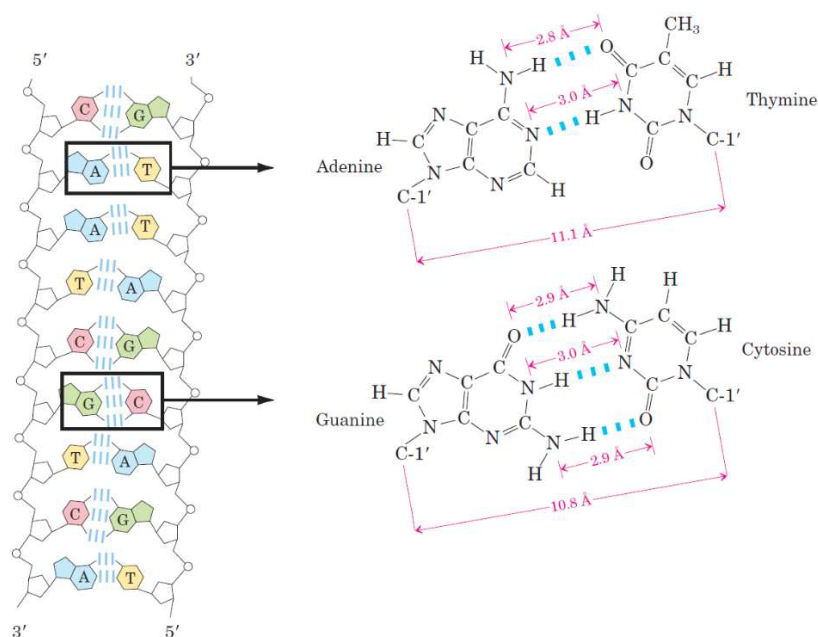
Estruturalmente, o RNA e o DNA são considerados polímeros cujo as menores unidades que os compõem, os monômeros, são uma combinação de bases nitrogenadas ligadas a um grupo fosfato e um açúcar do tipo ribose, para RNA, ou desoxirribose, para DNA. Esse monômero é denominado nucleotídeo. Os nucleotídeos se ligam um ao outro através dos grupos fosfato. Esse grupo conecta o carbono 3' de um açúcar ao 5' do próximo açúcar, permitindo a formação de fitas simples. Uma ponta da fita simples possui um carbono 5' livre e a outra ponta



**Figura 1.1**

Estrutura química das principais bases nitrogenadas. Figura retirada de Ref. [2].

possui um carbono 3' livre. Convencionalmente, a sequência é escrita no sentido do carbono 5' para o 3'. Duas fitas simples compatíveis podem hibridizar e formar uma dupla hélice, mais estável. A configuração da dupla hélice é tal que as fitas simples que a formam são antiparalelas, uma fita 5' – 3' e a outra 3' – 5'. As fitas simples são unidas por ligações de hidrogênio entre os pares de base que as formam, ver Figura 1.2. Os pares complementares, ou canônicos, AT (ou AU, em RNA) e GC, contribuem com duas e três ligações de hidrogênio, respectivamente. O pareamento canônico não é único, no entanto, é o mais comumente encontrado e também é denominado pares de Watson-Crick, em homenagem aos cientistas James Watson e Francis Crick [3], ganhadores do Nobel pelo desvendamento da estrutura de dupla hélice a partir das medidas de raio-X realizadas por Franklin e Wilkins [4].



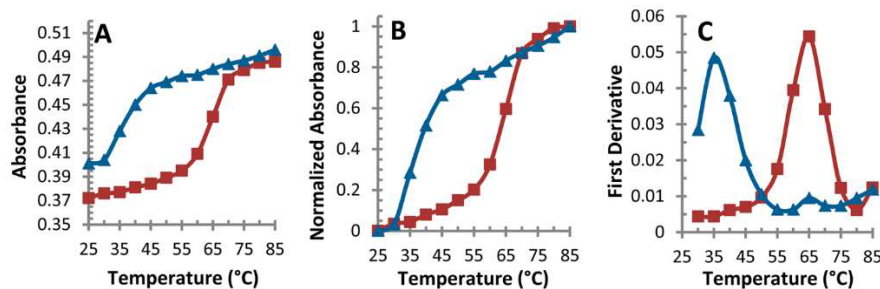
**Figura 1.2**

Estrutura de um duplexo de DNA representando as ligações de hidrogênio dos pares AT e GC. Figura retirada de Ref. [5].

## 1.2 Desnaturação de ácidos nucleicos

A desnaturação, ou *melting*, dos ácidos nucleicos é o fenômeno de dissociação de um duplexo em fita simples. A desnaturação pode ocorrer devido ao aumento da temperatura e outros fatores, como a variação da concentração salina da solução. Com o aumento da energia interna do sistema, as ligações de hidrogênio responsáveis por manterem a estrutura de fita dupla são rompidas e as fitas se separam. Conhecer as propriedades térmicas de uma sequência de DNA ou RNA é essencial para aplicações biotecnológicas, por exemplo, para produzir sondas de reação em cadeia da polimerase (*polymerase chain reaction*, PCR). A temperatura de *melting*  $T_m$  é definida como sendo a temperatura para a qual 50% da amostra está no estado de fita dupla e os outros 50% estão no estado de fita simples. Esse fenômeno pode ser estudado através de experimentos de absorção de ultravioleta (*ultraviolet*, UV), ver Figura 1.3. Os duplexos de DNA e RNA não absorvem frequências entre 260 e 280 nm, as fitas simples por sua vez absorvem essas frequências. Dessa forma, a curva de absorção em função da temperatura permite quantificar quanto da solução está em fita dupla ou em fita simples em uma dada temperatura. A absorbância é definida como sendo:

$$A = \log\left(\frac{I_0}{I}\right) \quad (1)$$



**Figura 1.3**

Determinação de temperatura de desnaturação por absorção ultravioleta. Curva de absorção em função da temperatura de duas sequências de DNA. A) Em vermelho, uma sequência complementar e, em azul, uma sequência contendo mismatches; B) a curva normalizada; C) a primeira derivada da curva de absorção. Figura retirada de Ref. [6]

A temperatura de desnaturação de uma solução de DNA ou RNA está diretamente associada à concentração das fitas em solução ( $C_t$ ), sendo diferente para fitas auto-complementares e não-auto-complementares. Uma fita é auto-complementar quando ela é sua própria sequência complementar, ou seja, quando lida no sentido  $3' - 5'$  complementa a sequência no sentido  $5' - 3'$ . Uma sequência é não-auto-complementar quando sua sequência complementar é diferente da sequência principal independente do sentido em que é lida. Por exemplo, a fita  $5' - \text{AUGCAU} - 3'$  é auto-complementar, pois  $3' - \text{UACGUA} - 5'$  é a sequência complementar e é exatamente a mesma sequência, mudando apenas o sentido em que é lida. A sequência  $5' - \text{AUUAGC} - 3'$ , por sua vez, não é auto-complementar, pois  $3' - \text{GCUAAU} - 5'$  é a complementar, mas não é a mesma sequência.

Inicialmente, vamos analisar as fitas não-auto-complementares. Sendo  $[X]$  e  $[Y]$  a concentração das fitas separadas e  $[XY]$  a concentração de duplexos formados. Definimos a fração  $f$  como a fração de duplexos dissociados e sendo  $C_t$  a concentração total de fitas  $X + Y$  na solução, temos a reação reversível:



$$f = \frac{[X]}{[XY]_{total}} = \frac{[Y]}{[XY]_{total}} \quad , \quad (3)$$

$$C_t = [X]_s + [Y]_s = [X] + [Y] + 2[XY] \quad (4)$$

Na temperatura de *melting*, temos  $f = \frac{1}{2}$ . A constante de equilíbrio da reação,  $K_{eq}$ , é dada por:

$$K_{eq} = \frac{[X][Y]}{[XY]} \quad (5)$$

O valor de  $K_{eq}$  na temperatura de *melting* é dado por:

$$[X]_m = [Y]_m = [XY]_m = \frac{C_t}{4} K_{eq} = \frac{[X][Y]}{[XY]} = \frac{(\frac{C_t}{4})(\frac{C_t}{4})}{\frac{C_t}{4}} = \frac{C_t}{4} \quad , \quad (6)$$

onde o subscrito  $m$  denota a concentração na temperatura de *melting*.

A energia livre de Gibbs para essa reação é dada por:

$$\Delta G = -RT_m \ln K_{eq} = \Delta H - T_m \Delta S \quad , \quad (7)$$

onde  $R$  é a constante universal dos gases. Essa equação estabelece uma relação entre a entropia  $S$ , a entalpia  $H$ , a temperatura de *melting*  $T_m$  e a constante de equilíbrio  $K_{eq}$ .

$$T_m = \frac{\Delta H}{\Delta S - R \ln(\frac{C_t}{4})} \quad (8)$$

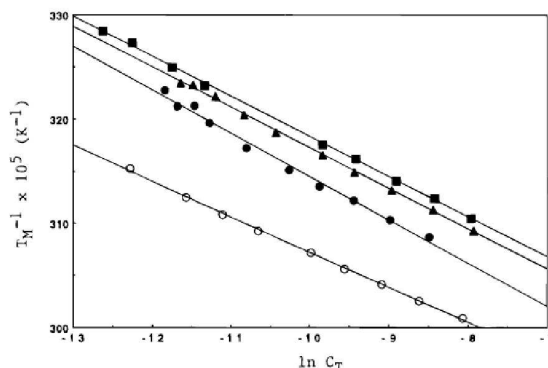
No caso auto-complementar, devido à simetria das fitas, temos  $K_{eq} = C_t$ :

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

Uma manipulação simples da equação permite reescrever na forma  $y = a + bx$ :

$$T_m^{-1} = \frac{\Delta S}{\Delta H} - \frac{R \ln(\frac{C_t}{4})}{\Delta H} \quad (10)$$

Dessa forma, é possível determinar, experimentalmente, valores de  $\Delta H$ ,  $\Delta S$  e  $\Delta G$  da reação de desnaturação de um duplexo. Na Figura 1.4, temos um gráfico  $T_m^{-1} \times \ln(\frac{C_t}{4})$ .



**Figura 1.4**

Gráfico de Van't Hoff da temperatura de *melting* em função da concentração total ( $T_m^{-1} \times \ln(\frac{C_t}{4})$ ). Cada reta representa uma sequência distinta. Figura retirada de Ref. [7].

### 1.3 Bases modificadas

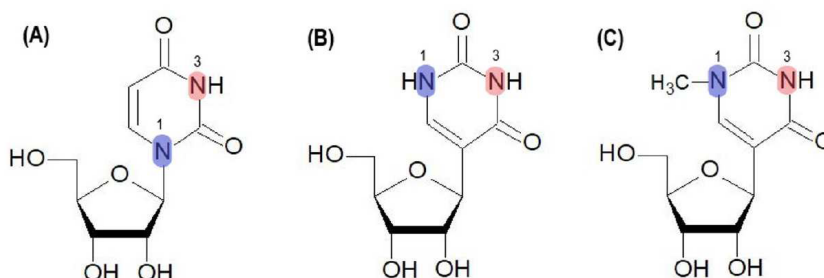
Existem na natureza outras bases nitrogenadas, além das canônicas A, C, G, T e U, capazes de estabilizar duplexos de DNA e RNA. Geralmente, essas bases são resultados de modificações na estrutura molecular das bases canônicas. O primeiro registro de uma base modificada foi um isômero da uridina detectada em RNA de leveduras *Saccharomyces cerevisiae* em 1956 e foi nomeada pseudouridina ( $\Psi$ ) [8, 9]. Atualmente, são conhecidas 17 modificações de ocorrência natural em DNA e mais de 150 possíveis modificações de ocorrência natural em RNA, que podem ser encontradas nos diversos tipos de RNA celular [10]. As bases alternativas de ocorrência natural possuem diversas funções em processos biológicos e bioquímicos, como na estabilização de estruturas do tRNA, rRNA, na eficiência da síntese proteica e na leitura alternativa do mRNA [11–14]. Em humanos, sua ocorrência é mediada por famílias de enzimas que atuam de forma altamente específica nos diversos tipos de RNA [13, 15–18]. As modificações podem ser usadas em diversas aplicações biotecnológicas, como aprimoramentos para *primers* de PCR [19], vacinas de RNA [1, 20] e no desenvolvimento de tratamento e diagnóstico de doenças [17, 18].

Neste trabalho, estudamos duas modificações de ocorrência natural em RNA e seu impacto na estrutura do RNA de dupla fita (*double strand RNA*, dsRNA): a hipoxantina, ou inosina (I), e a pseudouridina ( $\Psi$ ). As modificações também podem afetar de diversas formas a síntese proteica. A pseudouridilação, por exemplo, pode suprimir códons de terminação prematura [21]. A pseudouridina, atualmente, é utilizada em vacinas de mRNA para evitar reação imunológica [1, 22]. A inosina é resultado da deaminação da adenosina e sua ocorrência é devida a um grupo de enzimas denominado adenosina deaminase agindo em RNA (*adenosine deaminase acting in RNA*, ADAR), podendo ser interpretada pelo ribossomo como A, G e, em alguns casos, U [23], o que permite a ocorrência de tradução não-sinônima do mRNA. A ocorrência de ambas modificações é essencial para o desenvolvimento adequado das formas de vida, em especial, eucarióticas. Desequilíbrios na sua expressão estão associadas aos mais diversos tipos de doenças, como câncer e doenças cognitivas degenerativas, como o Alzheimer [24–28]. Conhecer o comportamento desses nucleotídeos em RNA pode permitir o desenvolvimento de terapias gênicas para o tratamento de doenças complexas.

#### 1.3.1 Pseudouridina

A pseudouridina ( $\Psi$ ) é a modificação mais abundante e foi a primeira a ser descoberta e caracterizada em 1956 [8, 9]. É o isômero C5-glicosídeo da uridina, no qual a uracila se liga à ribose através de uma ligação carbono-carbono ao invés de uma ligação nitrogênio-carbono [13, 29], ver Figura 1.5. Além disso, a rotação da base resulta na substituição do carbono C5 por um nitrogênio, sendo este um doador para ligações de hidrogênio, aumentando a capacidade de ligar com outras moléculas, incluindo proteínas. A modificação promove o enrijecimento

do esqueleto de fosfodiéster e um aumento na estabilidade térmica [29, 30]. A isomerização da uridina ocorre em RNA de forma natural sendo mediada por diversas enzimas que atuam de forma altamente específica variando os sítios e alvos da edição [13]. A pseudouridina ocorre nos diversos tipos de RNA e nas mais diversas formas de vida [13, 29], cumprindo diversos papéis, como a estabilização da estrutura do tRNA.



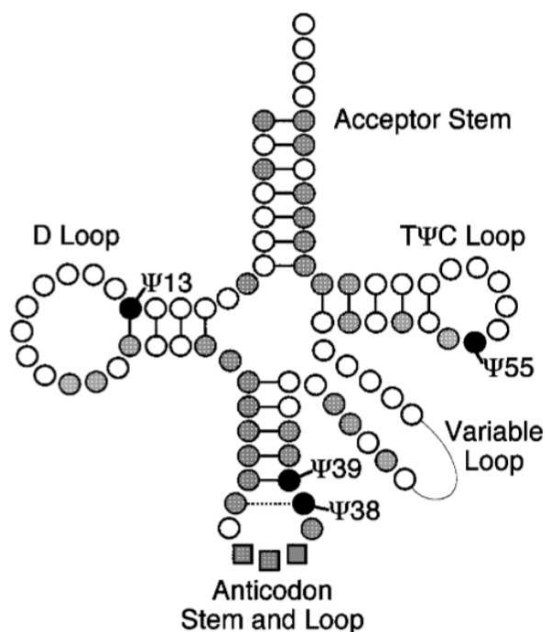
**Figura 1.5**

Estrutura química das bases: A) Uridina, B) pseudouridina e C) N1-metilpseudouridina. Figura retirada de Ref. [31].

Sua ocorrência em tRNA contribui para a estabilidade estrutural, por exemplo, o loop T $\Psi$ C ( $\Psi$ 55) e a haste D ( $\Psi$ 13). Nesta última, fornece uma forte interação entre as posições  $\Psi$ 39 e A31 [29], ver Figura 1.6. Tanto em tRNA quanto em mRNA, essa modificação é, geralmente, interpretada pelos mecanismos de decodificação celulares como uma uridina. Dessa forma, a modificação, em geral, não afeta a tradução celular, no entanto, a pseudouridina pode levar à supressão dos códons de parada [21, 29]. Supressão de enzimas responsáveis pela pseudouridilação em leveduras causa desde a redução do crescimento até a completa parada, em especial, em altas temperaturas [29]. Sendo assim, a pseudouridina é um mecanismo de variabilidade genética e também sua ocorrência está associada à eficiência da síntese proteica.

A pseudouridina também está presente no rRNA, onde é mais abundante, possuindo diversas funções desde a biogênese do ribossomo até a estabilização de interações RNA–RNA e RNA–proteína [30]. Apesar disso, devido à alta taxa de ocorrência, da complexidade da molécula e de seus variados mecanismos de funcionamento, pouco se sabe sobre sua função nesse tipo de RNA. A modificação é responsável, principalmente, por guiar a formação e estabilizar estruturas secundárias e terciárias do ribossomo [29, 30].

As bases nitrogenadas podem também sofrer metilação. Um grupo metil pode ser incorporado na base alterando a forma que essas bases interagem. A metilação é uma modificação que não altera as principais características da base [17], no entanto, pode ocasionar mudanças de algumas de suas propriedades físicas e estruturais, podendo, por exemplo, afetar positiva ou negativamente as interações com os pares vizinhos. O grupo N1-imino da pseudouridina também pode ser alvo de metilação, alterando algumas de suas propriedades, a nova base é denominada N1-metilpseudouridina (*N1-methylpseudouridine*,  $m^1\Psi$ ) [22, 31, 32].



**Figura 1.6**

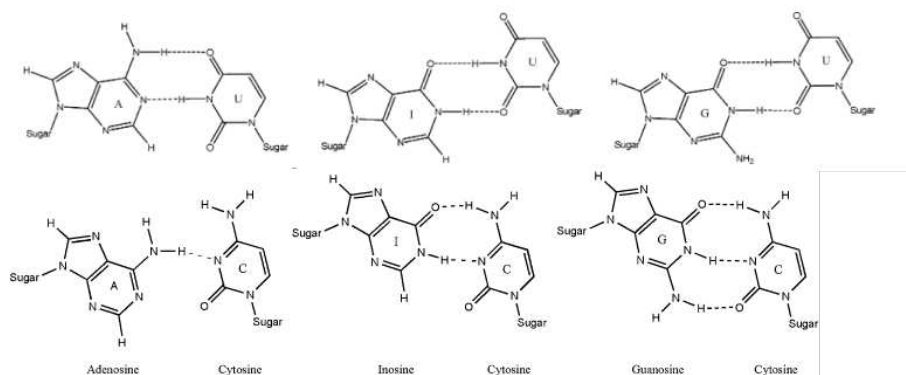
Estrutura do tRNA e posições de ocorrência de  $\Psi$ . Figura retirada de Ref. [29].

Sua utilização em aplicações biotecnológicas vem se mostrando eficaz e permitindo o desenvolvimento de terapias e vacinas baseadas em mRNA que utilizam pseudouridina e  $m^1\Psi$  [1, 22], ver Figura 1.5. O uso dessas modificações representou um grande avanço no desenvolvimento das vacinas de mRNA, sendo utilizadas, em especial, nas vacinas contra a doença do coronavírus 2019 (*coronavirus disease 2019*, COVID-19) [1, 22, 32, 33]. Vacinas de mRNA que utilizam  $\Psi$  e  $m^1\Psi$  são mais eficientes que as não modificadas. Em especial, a modificação metilada é ainda mais eficiente. Isso se deve à menor imunogenicidade e à manutenção da síntese proteica correta [32]. Ainda não está clara a razão bioquímica que reduz a imunogenicidade destas modificações.

### 1.3.2 Inosina

A inosina é resultado da deaminação da adenosina. Em RNA, essa modificação é mediada por enzimas denominadas ADAR, e esse processo é conhecido como edição A para I, ou seja, adenina para inosina [14, 34, 35]. Essa é uma modificação de ocorrência natural e é essencial para o desenvolvimento da vida. Apesar disso, desequilíbrios na sua expressão estão associados à diversas doenças, como câncer e doenças cognitivas, como Alzheimer [25–27]. A inosina é similar a uma guanina, exceto pelo grupo 2-amino que a inosina não contém. Dessa forma, a inosina em RNA é muitas vezes interpretada pelos mecanismos celulares como uma guanina [36, 37]. O pareamento IC se difere do pareamento GC por ter uma ponte de hidrogênio a menos, devido à ausência do grupo 2-amino [38]. O pareamento IU possui duas ligações de hi-

drogênio assim como AU, sendo diferentes devido ao deslocamento das ligações de hidrogênio e na mudança dos aceptores e doadores [35, 39], ver Figura 1.7.



**Figura 1.7**

Ligações de hidrogênio dos pares AU, IU, GU, AC, IC e GC, respectivamente. Figura adaptada de Refs. [35, 38].

A inosina pode ocorrer nos mais diversos tipos de RNA celular, tanto em tRNA, mRNA e em outros níveis de RNA eucariote [14, 16, 40–42]. Quando ocorre no tRNA, a inosina está associada à eficiência do processo de tradução, ocorrendo, geralmente, nas posições 34, 37 e 57 [12, 16, 43]. A modificação na posição 57 ocorre apenas em *Archaea*, sendo posteriormente modificada para uma N1-metilinosina (*N1-methylinosine*, m<sup>1</sup>I) [16, 43, 44]. A modificação na posição 37 também é posteriormente metilada, ocorre apenas no tRNA<sup>ala</sup> eucariote, e foi associada com a baixa biomassa de plantas que crescem em situação de estresse [12, 16]. A ocorrência na posição 34 ocorre em 7 ou 8 tipos de tRNAs eucariote. Nessa posição, a modificação pode alterar o processo de tradução proteica, podendo se ligar com códons com U, C, e A na posição equivalente [16, 44].

A ocorrência em mRNA também é essencial para o correto funcionamento celular. Sua expressão varia fortemente entre os diversos tecidos, sendo mais abundante no cérebro [15]. Durante a síntese proteica, a inosina em mRNA pode resultar em traduções não-sinônimas, sendo um fator de variabilidade genética que, muitas vezes, resulta em proteínas disfuncionais, incompletas ou com desvio de função, sendo a causa de diversas doenças [23, 26, 27]. Apesar disso, a inosina é essencial e sua ausência em determinados mRNAs também pode causar problemas. A ADAR2 atuante no mRNA da subunidade GluA2 do receptor  $\alpha$ -amino-3-hidroxi-metil-5-4-isoxazolpropiónico (AMPA) induz a substituição de uma glutamina por uma arginina, reduzindo a permeabilidade por íons Ca<sup>2+</sup> do AMPA. A modificação é evolutivamente conservada em mamíferos, sendo que aproximadamente todos os mRNAs de GluA2 humanos possuem a edição [26, 45]. Baixos níveis de edição no mRNA da GluA2 em camundongos está associada com convulsões e mortes prematuras [26, 27, 46].

A inosina pode também ocorrer em outras formas de RNA, como microRNA (*microRNA*, miRNA) e RNA curto de interferência (*small interfering RNA*, siRNA), e em cada um

desses casos, a inosina desempenha diversos papéis [26, 27, 41, 42]. Ambos miRNAs e siRNA levam à degradação de transcritos de mRNA homólogos, que são mecanismos de repressão gênica e atuam através do complexo de silenciamento induzido por RNA (*RNA-induced silencing complex*, RISC) [14, 27]. Esse mecanismo depende da hibridização entre o miRNA, ou siRNA, com o mRNA alvo, dessa forma, a modificação A para I afeta sensivelmente esses processos. A inosina pode inibir interações de miRNA e siRNA ou alterar seus mRNA alvos, sendo potencialmente uma opção para terapias gênicas [14, 27].

A inosina pode ocorrer em DNA de diversas formas, no entanto, geralmente, ela será identificada e removida pelos mecanismos de reparo de DNA [27, 47]. A ocorrência da inosina em DNA e suas propriedades termodinâmicas já foram estudadas e parâmetros para os modelos que permitem a predição de  $T_m$  também já foram obtidos [48–51]. Aplicações biotecnológicas de inosina em DNA tiveram sucesso em *primers* degenerados de PCR [19, 52, 53] e sondas para triagem de bibliotecas de DNA genômico [54, 55]. Vários estudos já foram realizados para avaliar a estabilidade termodinâmica da inosina em RNA, no entanto, em contraste com estudos em DNA, poucos dados termodinâmicos estão disponíveis [35, 38, 40, 56–60]. Isso se deve ao fato de o duplexo de RNA ser uma molécula mais instável que o DNA, tornando difícil a realização de experimentos de *melting* de muitas sequências. Inosina em RNA pode vir a ser utilizada em *primers* degenerados de PCR e como uma opção para auxiliar, desenvolver ou melhorar terapias gênicas para tratamento de doenças complexas.

## **2 Objetivos**

### **2.1 Objetivo geral**

Obter parâmetros termodinâmicos para o modelo Peyrard-Bishop (PB) para inosina e pseudouridina em RNA.

### **2.2 Objetivos Específicos**

1. Parametrizar as bases modificadas, inosina e pseudouridina em RNA, para modelo Peyrard-Bishop.
2. Verificar as contribuições das modificações para a estrutura de dupla hélice.
3. Desenvolver e avaliar métodos para a parametrização a partir de poucos dados experimentais.
4. Interpretar fenômenos biológicos e propor aplicações biotecnológicas das bases modificadas em RNA à luz do modelo PB.

### 3 Métodos

Neste capítulo, iremos detalhar os métodos teóricos e computacionais utilizados para estudar o comportamento das modificações de RNA. Foram utilizados os modelos de próximos vizinhos (*nearest-neighbour*, NN) e Peyrard-Bishop (PB). O modelo NN é um modelo fenomenológico que permite o estudo da desnaturação com base em grandezas termodinâmicas obtidas dos experimentos de desnaturação. O modelo mesoscópico PB, baseado em física estatística, foi inicialmente proposto em 1989 por Peyrard e Bishop [61] para o estudo das propriedades físicas de duplexos de DNA (ou RNA) e do processo de desnaturação. O tratamento computacional foi feito utilizando o software TfReg [62, 63].

#### 3.1 Modelo de próximos vizinhos

O modelo de próximos vizinhos (NN) é um modelo fenomenológico amplamente utilizado no estudo de desnaturação de ácidos nucleicos [64–66]. É um modelo linear que consiste na soma direta das contribuições energéticas de cada dímero para a  $T_m$  do duplexo.

A energia livre de Gibbs de uma sequência  $i$  em uma temperatura  $T$  é dada por:

$$\Delta G_i^{tot}(T) = \Delta H_i^{tot} - T\Delta S_i^{tot} \quad (11)$$

Os valores de  $\Delta H_i^{tot}$  e  $T\Delta S_i^{tot}$  são dados pela contribuição individual de cada dímero através das relações:

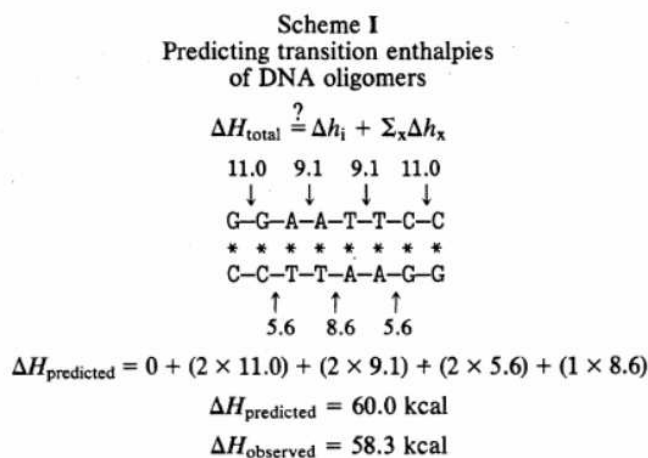
$$\Delta S_i^{tot} = \sum_{\alpha} n_{i\alpha} \Delta S_{\alpha} + \sum_c n^c \delta S^c \quad , \quad (12)$$

$$\Delta H_i^{tot} = \sum_{\alpha} n_{i\alpha} \Delta H_{\alpha} + \sum_c n^c \delta H^c \quad , \quad (13)$$

onde  $n_{i\alpha}$  é a quantidade de dímeros do tipo  $\alpha$  e  $\Delta S_{\alpha}$  e  $\Delta H_{\alpha}$  correspondem, respectivamente, às contribuições de entropia e entalpia de cada dímero.  $\delta H^c$  e  $\delta S^c$  são parâmetros de correção que dependem de diversos fatores, como simetrias, contribuições das bases terminais, entre outros, que devem ser considerados em cada caso. A energia livre de Gibbs a  $37^{\circ}C$  ( $\Delta G_{37}$ ) é utilizada para estudo de desnaturação de duplexos em temperaturas biológicas.  $\Delta G_{37}$  é dado pela equação 11 quando  $T$  é igual a  $37^{\circ}C$ .

Levando em conta todas as possíveis combinações de dímeros em DNA canônico, obtemos  $4^2 = 16$  parâmetros do tipo  $\alpha$ , que pode ser reduzido por relações de simetria. Por exemplo, o dímero AUpAU é equivalente ao dímero UApUA e o dímero AUpGC é equivalente ao dímero CGpUA. Os valores de  $\Delta S_{\alpha}$  e  $\Delta H_{\alpha}$  podem ser determinados a partir dos resultados

experimentais e da equação 10 por métodos algébricos [66] ou de otimização [65].



**Figura 3.1**

Aplicação do modelo NN para o cálculo de  $\Delta H$  de hibridização da sequência de DNA, 5'-GGAATTCC-3'. Figura retirada de Ref.[67].

Um exemplo do uso do modelo NN é apresentado na Figura 3.1. Consideramos a sequência de DNA, 5'-GGAATTCC-3', onde cada base faz par com seu correspondente canônico e é uma sequência auto-complementar [67]. Cada vizinho contribui energeticamente para a estabilidade do duplexo com um valor de  $\Delta H_{\alpha}$ . O valor da contribuição de cada vizinho é indicada pela seta, ver Figura 3.1. Os valores de  $\Delta H_{\alpha}$  indicados resultam num valor total  $\Delta H^{tot}$  de 60 kcal, enquanto o valor obtido experimentalmente é de 58.3 kcal. O mesmo procedimento é utilizado para obter  $\Delta S^{tot}$ , assim, a equação 10 pode ser utilizada para obter uma predição para temperatura de melting.

Os parâmetros para o modelo NN, incluindo os fatores de correção, podem ser obtidos através de diferentes métodos. Os valores para esses parâmetros variam dependendo de alguns fatores, como o conjunto de dados, as correções que são consideradas e os protocolos de otimização utilizados para realizar o cálculo desses valores. [65, 66, 68]. As Tabelas 3.1 e 3.2 apresentam parâmetros de NN e os fatores de correção de iniciação ( $\Delta H^{init}$ ), fator terminal ( $\Delta H^{term}$ ) e correção de simetria ( $\Delta H^{symm}$ ), utilizados para predição de  $T_m$  de sequências de RNA e DNA.

**Tabela 3.1**

Conjunto de parâmetros  $\Delta H$ ,  $\Delta S$  e fatores de correção para o modelo NN para DNA. Tabela adaptada da Ref. [65]. Os parâmetros P-SL96 e P-SL98 são das Ref. [7, 66].

Parâmetro	Valores						
	P-SL96	AOP-SL96	P-SL98	AOP-SL98	AOP-OW04	AOP-OW04DSC	AOP-CMB
$\Delta H_{A_P A}$	-8.4	-7.2(1)	-7.9	-8.4(1)	-8.6(2)	-8.2(4)	-7.20(4)
$\Delta H_{A_P C}$	-8.6	-10.4(1)	-8.4	-10.2(1)	-1.8(2)	-8.3(3)	-7.26(4)
$\Delta H_{A_P G}$	-6.1	-6.8(2)	-7.8	-5.3(1)	-7.0(3)	-6.4(3)	-6.46(5)
$\Delta H_{A_P T}$	-6.5	-9.5(2)	-7.2	-0.4(2)	-2.2(3)	-3.5(6)	-5.69(7)
$\Delta H_{C_P A}$	-7.4	-3.2(1)	-8.5	-12.9(2)	-15.0(2)	-8.5(3)	-8.58(6)
$\Delta H_{C_P C}$	-6.7	-7.49(8)	-8.0	-8.9(2)	-11.0(2)	-7.6(5)	-8.77(6)
$\Delta H_{C_P G}$	-10	-5.5(1)	-11	-9.3(1)	-14.4(4)	-12.1(6)	-9.70(6)
$\Delta H_{G_P A}$	-7.7	-2.77(9)	-8.2	-10.6(1)	-9.5(3)	-11.7(3)	-7.10(6)
$\Delta H_{G_P C}$	-11	-9.1(1)	-9.8	-14.5(2)	-6.8(4)	-12.5(3)	-9.59(8)
$\Delta H_{T_P A}$	-6.3	1.8(1)	-7.2	-6.1(2)	-13.8(3)	-12.9(5)	-4.08(7)
$\Delta S_{A_P A}$	-24	-20.3(4)	-22	-23.8(3)	-24.3(6)	-23(1)	-20.2(1)
$\Delta S_{A_P C}$	-23	-28.6(4)	-22	-28.3(5)	-3.2(7)	-22.2(8)	-19.3(1)
$\Delta S_{A_P G}$	-16	-17.9(5)	-21	-13.0(4)	-18.5(8)	-16.8(7)	-17.1(2)
$\Delta S_{A_P T}$	-19	-26.6(5)	-20	0.9(5)	-5.6(9)	-10(2)	-15.9(2)
$\Delta S_{C_P A}$	-19	-8.3(4)	-23	-36.6(6)	-42.0(7)	-23.0(8)	-23.2(2)
$\Delta S_{C_P C}$	-16	-19.7(3)	-20	-22.1(5)	-29.3(7)	-19(1)	-22.7(2)
$\Delta S_{C_P G}$	-26	-12.9(3)	-27	-22.9(4)	-39(1)	-31(2)	-24.9(2)
$\Delta S_{G_P A}$	-20	-6.9(3)	-22	-29.4(5)	-26(1)	-32.0(9)	-18.9(2)
$\Delta S_{G_P C}$	-28	-23.9(3)	-24	-38.9(7)	-16(1)	-32.4(8)	-24.4(2)
$\Delta S_{T_P A}$	-19	4.6(5)	-21	-17.7(7)	-40.3(8)	-37(2)	-11.3(2)
$\Delta S_{AT}^{init}$	-9.0	-6.0(2)	—	—	—	—	—
$\Delta S_{GC}^{init}$	-5.9	-0.9(1)	—	—	—	—	—
$\Delta H_{5'AT}^{term}$	0.40	-0.40(3)	—	—	—	—	—
$\Delta S^{symm}$	-1.4	0.67(6)	-1.4	-0.30(4)	—	—	-0.59(3)
$\Delta H_{CG}^{term}$	—	—	0.10	0.7(1)	4.9(4)	10.6(7)	0.49(4)
$\Delta S_{CG}^{term}$	—	—	-2.8	-1.5(3)	12(1)	29(2)	-0.7(1)
$\Delta H_{AT}^{term}$	—	—	2.3	1.3(2)	6.5(3)	12.6(8)	2.53(7)
$\Delta S_{AT}^{term}$	—	—	4.1	0.2(5)	17(1)	35(3)	5.7(2)

**Tabela 3.2**

Conjunto de parâmetros  $\Delta G_{37}$ ,  $\Delta H$  e  $\Delta S$  para o modelo NN para RNA. Tabela adaptada da Ref. [68].

Parâmetros	$\Delta G_{37}^{\circ}$ (kcal/mol)	$\Delta H^{\circ}$ (kcal/mol)	$\Delta S^{\circ}$ (eu)	
5'AA3'	3'UU5'	-0.93 (0.03)	-6.82 (0.79)	-19.0 (2.5)
5'AU3'	3'UA5'	-1.10 (0.08)	-9.38 (1.68)	-26.7 (5.2)
5'UA3'	3'AU5'	-1.33 (0.09)	-7.69 (2.02)	-20.5 (6.3)
5'CU3'	3'GA5'	-2.08 (0.06)	-10.48 (1.24)	-27.1 (3.8)
5'CA3'	3'GU5'	-2.11 (0.07)	-10.44 (1.28)	-26.9 (3.9)
5'GU3'	3'CA5'	-2.24 (0.06)	-11.40 (1.23)	-29.5 (3.9)
5'GA3'	3'CU5'	-2.35 (0.06)	-12.44 (1.20)	-32.5 (3.7)
5'CG3'	3'GC5'	-2.36 (0.09)	-10.64 (1.65)	-26.7 (5.0)
5'GG3'	3'CC5'	-3.26 (0.07)	-13.39 (1.24)	-32.7 (3.8)
5'GC3'	3'CG5'	-3.42 (0.08)	-14.88 (1.58)	-36.9 (4.9)
Iniciação		4.09 (0.22)	3.61 (4.12)	-1.5 (12.7)
Por terminal AU		0.45 (0.04)	3.72 (0.83)	10.5 (2.6)
Correção de simetria (auto-complementar)		0.43	0	-1.4
Correção de simetria (não auto-complementar)		0	0	0

### 3.2 Modelo Peyrard-Bishop

O modelo Peyrard-Bishop (PB) considera a fita de DNA (ou RNA) como uma estrutura homogênea e bidimensional e cada base nitrogenada como uma mesma partícula de massa  $m$ . O modelo descreve as interações de empilhamento e das ligações de hidrogênio de forma separada. Uma vez conhecidos os parâmetros de interação entre as bases e os pares vizinhos, o modelo permite prever temperaturas de desnaturação para sequências de DNA ou RNA. Inicialmente, definimos as variáveis  $\mu_n$  e  $\nu_n$ , sendo os graus de liberdade, e representam o deslocamento da  $n$ -ésima base em relação a sua posição de equilíbrio.

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

A ligação de hidrogênio é modelada por um potencial de Morse  $V$  [69] dado por

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

onde  $D$  é a energia de dissociação do par de base e  $\lambda_a$  é o alcance de atuação do potencial. A interação de empilhamento  $k$ , ou *stacking*, entre pares de bases vizinhos é modelada como um acoplamento harmônico, semelhante ao comportamento de uma mola, descrito pela equação:

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

Assim, o hamiltoniano do sistema é:

$$H = \sum_N \frac{m}{2}[\dot{\mu}_n^2 + \dot{\nu}_n^2] + \frac{k}{2}(\mu_n - \mu_{n-1})^2 + \frac{k}{2}(\nu_n - \nu_{n-1})^2 + D[e^{-\frac{(\mu_n - \nu_n)}{\lambda_a}} - 1]^2 \quad (17)$$

Podemos reescrever o hamiltoniano do sistema em termos das coordenadas cartesianas  $y_n$  e  $x_n$  utilizando a transformação:

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

O hamiltoniano agora é:

$$H = \sum_N \frac{1}{2m}[p_n + q_n] + \frac{k}{2}(y_n, y_{n-1})^2 + \frac{k}{2}(x_n, x_{n-1})^2 + D[e^{-\frac{y_n}{\lambda_a}} - 1]^2, \quad (19)$$

onde  $p_n$  e  $q_n$  são os momentos das coordenadas  $x_n$  e  $y_n$ , os termos harmônicos representam as interações entre os pares  $(n, n - 1)$  e o potencial de Morse descreve a interação entre as duas bases que formam um par. A partir do hamiltoniano podemos escrever a função partição  $Z$ :

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

Os termos de momento possuem solução:

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

$$Z_p Z_q = (2\pi m K_\beta T)^N \quad , \quad (22)$$

onde  $N$  é o número de pares de bases da fita e  $K_\beta$  é a constante de Boltzman. No eixo  $x$ , a única contribuição é o termo harmônico. A função  $Z_x$  é:

$$Z_x = \left(\frac{2\pi}{\beta k}\right)^{N/2} \quad (23)$$

Em  $y$ , temos:

$$Z_y = \int dy_1 \int dy_2 \dots \int dy_n \prod_N e^{-\beta[w(y_n, y_{n-1}) + V(y_n)]} \quad , \quad (24)$$

onde  $V(y_n) = (D[e^{\frac{y_n}{\lambda}} - 1])$ . Conhecendo os parâmetros de interação de cada par de bases é possível obter informações físicas de grande relevância para o estudo dos oligonucleotídeos, como a abertura média entre cada par de bases  $\langle y_n \rangle$ .

$$\langle y_n \rangle = \frac{1}{Z} \int \prod_n y_n e^{-\beta H} = \frac{1}{Z_y} = \int y_n dy_1 \int dy_2 \dots \int dy_n \prod_N e^{-\beta[w(y_n, y_{n-1}) + V(y_n)]} \quad (25)$$

### 3.3 Obtendo temperaturas de *melting* com o modelo PB

O modelo PB permite o cálculo de temperaturas de *melting*  $T_m$ . Isso pode ser feito estabelecendo o que é chamado equivalência térmica entre as sequências. Utiliza-se um índice  $\tau$  proporcional a  $T_m$  para agrupar sequências por temperaturas de *melting* similares. Se duas sequências possuem índices equivalentes, então as sequências são equivalentes e suas  $T_m$  também são equivalentes. Esse índice é dado pelo expoente da função partição. O índice pode ser ajustado utilizando resultados de experimentos de desnaturação. Isso permite calcular novas temperaturas de *melting* através de uma regressão linear dos índices. Podemos utilizar o método da integral de transferência para calcular a função partição  $Z$  do modelo PB. O maior termo da expansão é o índice  $\tau_i$  de uma dada sequência  $i$ . Considerando condições periódicas de contorno, temos:

$$Z = Tr[C^{(1,2)}C^{(2,3)}\dots C^{(N,1)}] \quad , \quad (26)$$

onde  $C^{(n,n+1)}$  representa a interação entre vizinhos e  $C^{(N,1)}$  é a condição de contorno periódica, ou seja, o primeiro e o último par ligados. As interações entre os pares de base AT (AU para o RNA) e GC são considerados fraca ( $w$ ) e forte ( $s$ ), respectivamente, o que reduz as interações de vizinhos para quatro tipos possíveis,  $C^{(s,s)}$ ,  $C^{(w,w)}$ ,  $C^{(s,w)}$  e  $C^{(w,s)}$ . Usando o vizinho GCpCG em uma sequência homogênea, uma referência de base ortonormal, a matriz  $C^{(s,s)}$  se torna uma matriz diagonal  $\Lambda$  com autovalores  $\lambda_i$ . Para os pares gerais, definimos a matriz  $\Delta^{(a,b)}$  sendo a diferença de interação entre pares vizinhos quaisquer e os vizinhos  $C^{(s,s)}$ , de tal forma que:

$$C^{(a,b)} = \Lambda + \Delta^{(a,b)} \quad (27)$$

A função partição então pode ser escrita como:

$$Z = Tr[(\Lambda + \Delta^{(1,2)})(\Lambda + \Delta^{(2,3)})\dots(\Lambda + \Delta^{(N,1)})] \quad (28)$$

Utilizando as propriedades do traço, temos:

$$Z = \sum_{\omega=0}^N Z_{\omega}(\Lambda) = \sum_{\omega=0}^N Tr[M(\Lambda^{\omega})] \quad , \quad (29)$$

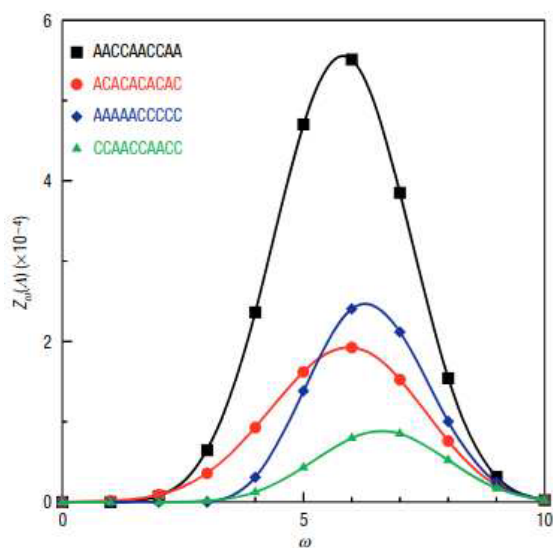
onde  $M[\Lambda^{\omega}]$  são os termos que contêm  $\omega$  multiplicações da matriz  $\Lambda$ . No caso da sequência GC homogênea,  $\Delta^{(s,s)} = 0$ . Nesse caso, todos os termos são de ordem máxima,  $\Lambda^N$ . A Figura 3.2 mostra a função partição  $Z_{\omega}(\Lambda)$  em função do parâmetro  $\omega$  de algumas sequências de mesmo tamanho e com 40-60% de pares GC, calculada em uma temperatura de 370 K. Observa-se também que o valor  $\omega_{max}$ , que maximiza a função, é fortemente dependente da composição da sequência e sendo aproximadamente independente da temperatura, ver Figura 3.3.

A Figura 3.4 mostra a relação linear entre temperaturas de *melting* experimentais e o valor  $\omega_{max}^{1/2}$  de diversas sequências agrupadas por tamanho. Essa relação permite obter temperaturas de *melting* de sequências a partir de uma regressão linear em função de um valor adimensional  $\omega_{max}^{1/2}$  que representa a equivalência térmica, o índice  $\tau$ . A regressão que relaciona o índice  $\tau_i$  com a temperatura de *melting* da sequência  $i$  por grupos de tamanho é:

$$T'_i = a_0(N) + a_1(N)\tau_i \quad , \quad (30)$$

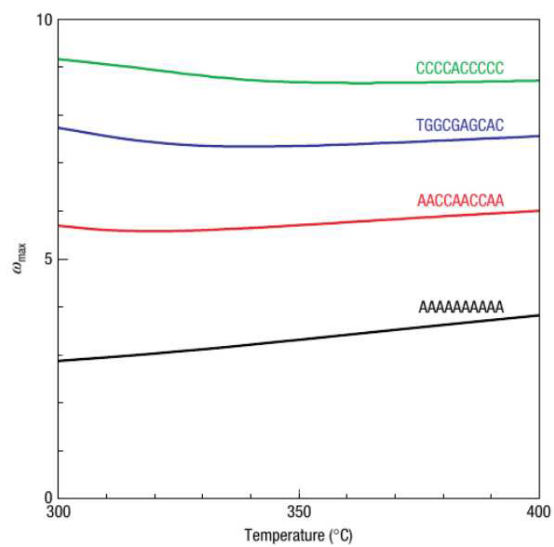
onde  $T'_i$  é a temperatura de *melting* da sequência  $i$ ,  $N$  é o tamanho da sequência e  $a_0$  e  $a_1$  são os coeficientes da regressão linear de cada grupo de tamanho  $N$ . A regressão linear depende também de  $N^{1/2}$ , dessa forma, para sequências com tamanhos sem base de dados experimentais pode-se usar a regressão:

$$a_k(N) = b_{0,k} + b_{1,k}N^{1/2} \quad (31)$$



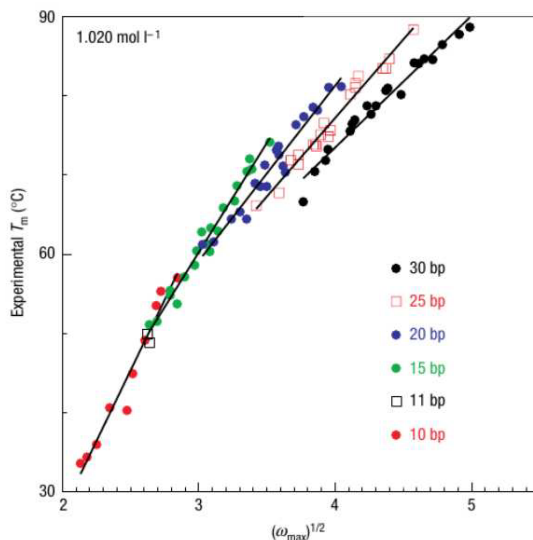
**Figura 3.2**

Distribuição normal da função partição em função do parâmetro de ordem  $\omega$ . Figura retirada de Ref [70].



**Figura 3.3**

Dependência de  $\omega$  com a temperatura. Figura retirada de Ref. [70].



**Figura 3.4**

Temperatura experimental  $T_m$  em função do parâmetro  $\omega_{max}^{1/2}$ . As linhas representam a regressão linear por grupos de sequências com mesmo tamanho, exceto para os dois pontos de sequências de 11 pares de bases. Os coeficientes de regressão linear são utilizados para realizar a predição final de  $T_m$  a partir dos parâmetros de Morse e harmônico. Figura retirada de Ref. [70].

### 3.4 Algoritmo de minimização para obtenção de parâmetros

A predição dos valores de temperatura de *melting* de uma sequência qualquer requer a determinação dos parâmetros  $D$  e  $k$ , determinados por cada par, que definem o hamiltoniano. Esses parâmetros devem estar alinhados com os resultados experimentais. A implementação computacional do modelo PB, o software TfReg [63], calcula os parâmetros do hamiltoniano minimizando a seguinte função:

$$\chi^2 = \sum_{i=1}^N (T'_i - T_i)^2, \quad (32)$$

onde  $T'_i$  é a temperatura de *melting* experimental da  $i$ -ésima sequência e  $T_i$  é a temperatura calculada por um certo conjunto de parâmetros. Para isso, é utilizado o método Nelder-Mead, ou *downhill simplex*, de minimização [71, 72]. Esse algoritmo de minimização foi proposto em 1965 por John Nelder e Roger Mead. A principal vantagem é que não é necessário calcular as derivadas da função para obter valores mínimos. Esse método parte de um conjunto de valores iniciais para os parâmetros  $D$  e  $k$  aplicando pequenas variações nesses valores e recalculando a função  $\chi^2$  até se obter valores que minimizem a função. Esse método permite percorrer o espaço de parâmetros do hamiltoniano e encontrar mínimos da função  $\chi^2$ . Isso permite chegar em valores de  $D$  e  $k$  condizentes com as condições experimentais e permite a predição de temperaturas de *melting* de quaisquer sequências para as quais os valores de  $D$  e  $k$  já tenham sido obtidos.

A dimensão  $N$  do espaço de parâmetros é determinada pela quantidade de parâmetros a serem minimizados. Utiliza-se uma figura geométrica chamada *simplex* que possui  $N+1$  vértices. Em um espaço bidimensional, o *simplex* é um triângulo, em três dimensões, um tetraedro e assim, sucessivamente. O algoritmo consiste em variar o *simplex* comparando os valores dos vértices. Uma estimativa inicial é utilizada para produzir o primeiro ponto  $P_0$ , dado por um vetor de  $N$  variáveis independentes. Para o modelo PB, utilizamos um conjunto padrão de valores iniciais para cada parâmetro que será minimizado, sendo eles:  $D = 38.0 \text{ meV}$ ,  $k = 2.5 \text{ eV/nm}^2$  e  $\lambda_a = 0.29 \text{ nm}$ . O  $i$ -ésimo ponto  $P_i$  do *simplex* é dado por:

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

onde  $e_i$  são vetores de tamanho  $N$  e  $\rho$  é uma constante. O algoritmo consiste em uma série de transformações que atuam sobre o *simplex* comparando os valores da função nos vértices, ver Figura 3.5. Inicialmente, os vértices são classificados como melhores ou piores. No caso da minimização, os melhores são os que resultam no menor valor da função que está sendo avaliada. A partir da ordenação é calculado o centróide ( $x_0$ ) do *simplex*, excluindo o pior ponto. A primeira etapa de transformações é chamada de reflexão e é dada por:

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

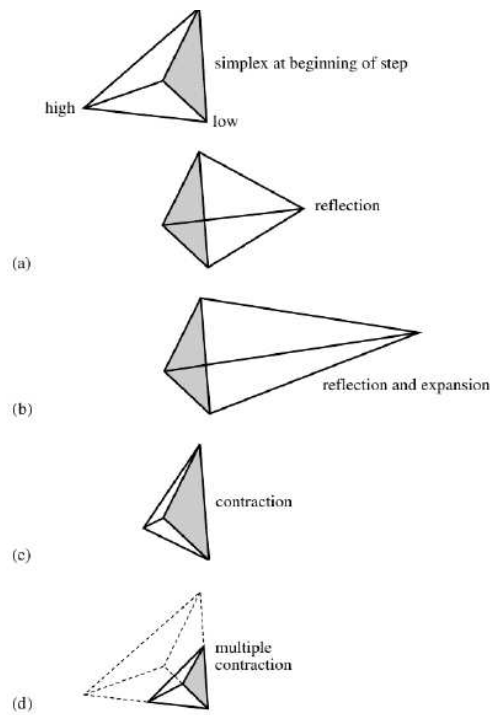
onde  $\alpha$  é um coeficiente de reflexão. O pior ponto é refletido através do centróide e um novo ponto ( $x_r$ ) é estabelecido. Se o valor do novo ponto é comparado com os outros pontos, caso  $x_r$  possua um valor entre o melhor ponto e o de valor intermediário, o pior ponto é então substituído por  $x_r$  e o cálculo é repetido para o novo *simplex*. Caso contrário, segue-se para as próximas etapas de transformações. Se  $x_r$  possui um valor menor que o melhor ponto, utiliza-se a expansão dada por:

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

onde  $\gamma$  é o coeficiente de expansão. Se  $x_e < x_r$ , substitui-se o pior ponto por  $x_e$ , caso contrário, o pior ponto é substituído por  $x_r$ . Caso  $x_r$  seja maior ou igual ao valor dos pontos intermediários, a contração é utilizada:

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

onde  $0 < \beta < 1$  é o coeficiente de contração. Se o valor da função no ponto de contração é menor que o pior ponto, então ocorre a substituição e novamente se inicia o cálculo do *simplex*. Nessa etapa, o volume do *simplex* é reduzido. Caso nenhuma das etapas resultarem em uma melhoria dos pontos do *simplex*, então todos os pontos, exceto o melhor, são redefinidos e o processo é reiniciado.



**Figura 3.5**

Ilustração do simplex e suas transformações. O melhor ponto e o pior são representados por *low* e *high*. As transformações são ilustradas por (a) reflexão, (b) reflexão e expansão, (c) contração e (d) compressão. Similar à contração, a compressão preserva o melhor ponto e altera todos os outros. Figura retirada de Ref. [72].

### 3.5 Lidando com poucos dados

Poucos dados públicos de *melting* estão disponíveis para bases modificadas em RNA. A grande maioria dos dados de *melting* para bases modificadas em RNA é obtida para o modelo NN. Esse modelo necessita um pequeno número de sequências para a parametrização das variáveis. O TfReg [62, 63] é uma ferramenta que funciona melhor quando uma grande quantidade de dados de *melting* para cada pareamento pode ser utilizada. Dessa forma, o estudo de bases modificadas em RNA, utilizando o modelo Peyrard-Bishop, se torna limitado. Essa limitação se deve ao fato de que a  $T_m$  de cada sequência modificada é representada por uma equação de 3 variáveis. Para cada sequência com uma modificação e com um valor experimental de  $T_m$ , três parâmetros devem ser minimizados: o Morse e os harmônicos dos vizinhos 5' e 3'. A parametrização é análoga a um sistema de equações com várias variáveis. O cenário ideal para a minimização é que o número de variáveis que serão minimizadas seja igual ou menor que o de sequências com  $T_m$ 's experimentais disponíveis. Caso haja mais variáveis do que sequências, a minimização se torna um sistema de equações indeterminado, ou seja, existem infinitas soluções e as variáveis podem assumir qualquer valor, isso resulta na perda de sentido físico do resultado. Para evitar esse problema, podemos igualar os parâmetros de variáveis similares, por exemplo, para IUpAU e IUpUA assume-se a inversão AU  $\rightarrow$  UA, ou seja, isso não altera os valores de *stacking* e esses vizinhos são equivalentes. Essa medida reduz o número de variáveis no sistema de equações, permitindo que ele tenha solução única. Apesar disso, se não for verdade que os valores de *stacking* de IUpAU e IUpUA são similares, a precisão do cálculo será reduzida e obteremos parâmetros que resolvem as equações e permitem a predição de  $T_m$ , mas que não representa a realidade física do sistema.

Outra solução para lidar com poucos dados é a separação por contextos. A dependência de contexto é levada em conta ao aplicarmos o que chamamos de regras de contexto. Essas regras diferenciam e agrupam os parâmetros de um pareamento de acordo com as bases vizinhas especificadas. Por exemplo, o valor dos parâmetros de Morse e de harmônico para inosina no contexto 1, definido pelo trímero GIC/CUG, podem diferir significativamente dos valores no contexto 2, definido pelo trímero AIC/UUG. Utilizamos a notação  $IU^1$  e  $IU^2$  para diferenciar os contextos. As regras de contexto atuam agrupando as sequências que contêm os trimeros especificados por cada regra. Uma regra de contexto também pode ser definida por um conjunto de trimeros. Dessa forma, o exemplo acima resulta em um aumento dos parâmetros a serem minimizados, com os parâmetros de Morse  $IU^1$  e  $IU^2$ , e de harmônico  $GCpIU^1$ ,  $AUpIU^2$ ,  $IU^1pCG$  e  $IU^2pCG$ . Sem o uso dessa separação, teríamos apenas os parâmetros de Morse  $IU$  e de harmônico  $GCpIU$ ,  $AUpIU$  e  $IUpCG$ . Geralmente, as variações dos parâmetros devido ao contexto são negligenciáveis, mas, em alguns casos, podem ocorrer grandes variações entre um contexto e outro [73].

A separação de contexto aumenta o número de variáveis do sistema, de forma que o

sistema de equações pode novamente se tornar indeterminado. A escolha correta de contextos, no entanto, permite uma maior precisão no cálculo dos parâmetros. Essa precisão maior ao separar os contextos se deve ao fato de o TfReg calcular valores médios para os parâmetros que minimizam os sistemas de equação. Por outro lado, ao realizar uma minimização sem o uso do contexto para esses mesmos trimeros, obteremos um valor de Morse IU que representa um valor médio dos Morses em cada contexto. Nesse caso, se os Parâmetros de Morse forem muito diferentes entre os dois contextos, mais uma vez o resultado deixa de representar a realidade física do sistema e se torna apenas um valor para a predição de  $T_m$ .

Neste trabalho, foram utilizadas as duas técnicas para tratar os dados, tanto em conjunto quanto separadas. O resultado que melhor se adequou aos dados foram os que utilizaram apenas a regra de contexto. Para isso, foram escolhidas regras que representavam grandes variações das variáveis em cada contexto e que mantinham o número de variáveis menor que o número de equações.

## 4 Inosina

Neste capítulo, será abordada a aplicação do modelo PB às sequências de RNA com pelo menos uma modificação de inosina; os procedimentos de minimização para obtenção dos parâmetros e concluindo com a discussão dos resultados obtidos.

### 4.1 Introdução

Vários estudos [35, 38, 40, 56–60] já foram conduzidos para obtenção de parâmetros termodinâmicos para inosina em RNA. Apesar disso, o modelo NN usado nesses estudos não descreve as interações intramoleculares como as ligações de hidrogênio e interações de *stacking*. Neste capítulo, descrevemos os procedimentos e os parâmetros obtidos para inosina em RNA no modelo Peyrard-Bishop. Foram estudados os pares IC, IG, II e IU, incluindo IC e IU tandem [35, 38, 59]. Infelizmente, dados públicos de *melting* para RNA com bases modificadas são muito limitados. Dessa forma, não foi possível explorar todos os pareamentos e contextos possíveis.

### 4.2 Procedimentos de minimização

Neste trabalho, foram utilizados dados publicados de temperaturas de *melting* com as mesmas condições experimentais (NaCl 1 M, cacodilato de sódio 20 mM, Na<sub>2</sub>EDTA 0,5 mM, pH 7,0 e concentração de oligonucleotídeos de 100  $\mu$  M), dos quais obtivemos 71 sequências [35, 38, 59] contendo IC (30 sequências), IG (9 sequências), II (5 sequências) e IU (27 sequências), incluindo sequências IU e IC tandem, ver Tabela A.1, no Apêndice A. As sequências foram divididas em pares únicos IC, IG, II e IU, quando são não-tandem, podendo haver mais de uma modificação na sequência; IC e IU tandem, quando há duas modificações consecutivas; e, posteriormente, divididas em grupos de acordo com a base com a qual fazem par.

Os procedimentos de minimização foram conduzidos para obter parâmetros de potencial de Morse ( $D$ ) e potencial harmônico ( $k$ ), ligações de hidrogênio e de interação de *stacking*, respectivamente, para inosina em RNA. O conjunto de dados foi dividido em quatro principais grupos, um para cada pareamento IC, IG, II e IU. A implementação computacional VarGibbs [65] para o modelo NN foi utilizado para produzir temperaturas de *melting* para sequências canônicas de referência. As sequências de referência foram produzidas substituindo a I por seu equivalente canônico. As sequências de referência para o par IG foram produzidas substituindo o par GI pelo par GC. Duas sequências de referência foram produzidas para os pares II, uma substituindo o par II por AU e a outra por GC. A escolha da referência para IG e II se deve a falta de uma parametrização canônica para o *mismatch* equivalente. A lista completa de sequências de referência é mostrada na Tabela A.2.

As minimizações foram feitas a partir de um valor inicial padrão para os parâmetros ( $D = 38.0 \text{ meV}$ ,  $k = 2.5 \text{ eV/nm}^2$  e  $\lambda_a = 0.29 \text{ nm}$ ) para a primeira rodada de minimização. O valor de  $\lambda_a$  foi mantido fixo em todas as minimizações. Em cada rodada de minimização, foi utilizada uma randomização dos parâmetros iniciais de 20%. A estabilidade da inosina na fita é fortemente dependente do contexto, em especial, a inosina na posição terminal da fita [35, 38]. Os contextos interno e terminal foram considerados para o pareamento C e U, durante as minimizações,  $IC^I$  e  $IU^I$  para internos e  $IC^*$  e  $IU^*$  para terminais. Também foi possível identificar e analisar contextos para o qual o par interno IU reduz a estabilidade do duplexo quando comparado com outros contextos para IU.

A cada rodada global foram realizadas 100 minimizações locais e o resultado final de parâmetros é a média dos 20 melhores resultados. As minimizações locais foram realizadas até que a diferença de  $\chi^2$  entre dois passos locais seja menor que  $0.1 \text{ }^\circ\text{C}^2$ . As rodadas globais conduzidas após a primeira rodada utilizaram como parâmetros iniciais os resultados da rodada anterior. Minimizações globais foram conduzidas até que não houvesse melhora significativa no valor de  $\chi^2$  que é utilizado como um parâmetro de controle de qualidade para cada rodada de minimização, ver Tabela 4.1.

**Tabela 4.1**

Valor de  $\chi^2$  para cada grupo e em cada rodada de minimização.

Rounds	Progressão de $\chi^2$ ( $^\circ\text{C}^2$ )					
	Interno/Terminal				Tandem	
	IC	IU	IG	II	IU <sup>†</sup>	IC <sup>†</sup>
r1	13.5	44.2	3.7	14.5	3.8	3.4
r2	13.3	44.1	3.7	14.4	3.6	3.3
r3	13.3	16.6	3.7	14.5	–	3.2
r4	13.2	16.6	3.7	–	–	3.2
r5	–	16.5	–	–	–	–

#### 4.2.1 Grupo IC

Os parâmetros IC foram divididos em grupos de contexto para interações internas e terminais,  $IC^I$  e  $IC^*$ , respectivamente. Sequências tandem foram minimizadas separadamente levando em conta a dependência de contexto,  $IC^\dagger$ .

**Rodadas interno e terminal** Quatro rodadas foram conduzidas para esse grupo. A partir da terceira rodada a randomização dos valores iniciais foi modificada de 20% para 10%, uma vez que o valor de  $\chi^2$  estava se aproximando do mínimo.

**Rodada tandem** As sequências  $IC^\dagger$  tandem foram minimizadas em quatro rodadas. Os parâmetros iniciais utilizados na primeira rodada foram os resultados obtidos na última rodada

de internos e terminais. Consideramos que os parâmetros de harmônico para vizinhos não-tandem são iguais, e não variam, aos parâmetros obtidos na minimização interna, por exemplo, os vizinhos  $CGpIC^I$  e  $CGpIC^*$ . A partir da terceira rodada utilizamos a randomização inicial de parâmetros de 10%, devido ao valor de  $\chi^2$  estar próximo do mínimo.

#### 4.2.2 Grupo IG

Neste grupo, nenhuma regra de contexto foi utilizada. O resultado final foi obtido em três rodadas de minimização. Além disso, apenas IG internos foram minimizados, pois não há pares terminais ou tandem nesse grupo.

#### 4.2.3 Grupo II

Neste grupo, nenhuma regra de contexto foi utilizada. Além disso, apenas II internos foram minimizados, pois não há pares terminais ou tandem nesse grupo. Três rodadas foram feitas para esse grupo.

#### 4.2.4 Grupo IU

**Rodada interno e terminal** Os parâmetros IU foram divididos em dois grupos considerando inteiros e terminais,  $IU^I$  e  $IU^*$ , respectivamente. Sequências tandem foram minimizadas separadamente, como no grupo IC. Duas rodadas principais foram feitas considerando apenas o contexto interno e terminal. A partir desses resultados, realizamos um estudo de contexto para o pareamento IU.

**Rodada de separação dos contextos internos** Após a rodada interno e terminal foi identificada a necessidade de realizar estudos de dependência do contexto. Este grupo de minimizações incorpora a regra de contexto para os pares internos. Os contextos utilizados são apresentados na Tabela A.3, no Apêndice A. Três rodadas de minimização foram feitas utilizando todas as sequências. A primeira rodada desse grupo utilizou resultados obtidos na rodada de contextos interno e terminal como parâmetros iniciais. Os parâmetros utilizados são apresentados na Tabela A.6, no Apêndice A.

**Rodada tandem** Duas rodadas foram realizadas para esse grupo. Apenas os parâmetros de *stacking*,  $IUpUI$ ,  $IUpIU$  e  $UIpIU$  foram minimizados. Utilizamos os parâmetros  $IU^B$ , obtidos a partir do grupo anterior, como parâmetros iniciais para a minimização.

#### 4.2.5 Desvio padrão

Uma última rodada foi realizada para cada grupo a fim de determinar os valores do desvio padrão. Nesta rodada, foram realizadas 100 minimizações. Como nenhum erro experimental foi fornecido para o conjunto de dados utilizados neste trabalho, consideramos que o erro experimental das medições de  $T_m$  é aproximadamente  $\pm 0,5^\circ C$ . Cada minimização utilizou uma randomização do conjunto de dados de  $T_m$  em  $\pm 0,5^\circ C$  e nenhuma randomização inicial para os parâmetros de entrada. Este procedimento nos permite simular vários cenários experimentais, cada minimização ajusta ligeiramente os parâmetros para o novo conjunto de dados. A média final e o desvio padrão, para cada parâmetro, foram obtidos a partir dos resultados desta rodada.

### 4.3 Resultados

Parâmetros de ligação de hidrogênio e de *stacking*, respectivamente, potenciais de Morse e harmônico, foram obtidos a partir de resultados de *melting* de 71 sequências de RNA contendo pares IC, IG, II e IU, incluindo IC e IU tandem. A escolha inicial de parâmetros foi feita considerando que a inosina forma duas ligações de hidrogênio com as bases às quais se liga. Para isso, foi utilizado o parâmetro de Morse do par AU. Os resultados obtidos na minimização estão de acordo com os resultados experimentais disponíveis. A comparação entre os resultados preditos pelo modelo Peyrard-Bishop e os resultados experimentais é mostrada na Tabela A.1. Além disso, consideramos o contexto ao redor da inosina uma vez, que seu comportamento é dependente do ambiente da sequência [35–38]. Os resultados finais da minimização são apresentados nas Tabelas A.4 e A.5, no Apêndice A.

Os resultados do modelo PB apontam para uma alta instabilidade dos pares IU internos,  $IU^A$  e  $IU^B$ , sendo consideravelmente mais instável que os pares AU e GU, concordando com outros estudos [35, 56, 73], ver Figuras 4.1 e 4.2 e Tabelas A.1, A.4 e A.5, no Apêndice A. I e G são bases muito similares estruturalmente, com exceção do grupo N2-amino em G. Ambas ligações de hidrogênio dos pares GU e IU são deslocadas em relação às ligações do par AU. Ao invés da ligação N6-amino de A para O4-carbonila de U e N1 de A para N3-imino de U, ocorre O6-carbonila de I (ou G) para N3-imino de IU e N1-imino de I (ou G) para O2-carbonila de U [35]. A ligação de hidrogênio interna mais fraca dos pares IU pode ser explicada pelas mudanças conformacionais da geometria da ligação [39] e sua similaridade com potencial de Morse para GU com uma ligação de hidrogênio [73] sugere que na maioria dos casos o par IU interno possui apenas uma ligação de hidrogênio. A contribuição dos pares IU para a estabilidade da sequência é na média 2 kcal/mol a menos que pares GU nas mesmas posições [35], o que é consistente com a perda de uma ligação de hidrogênio. Esse valor também foi utilizado para avaliar contextos internos do par IU. Outras variáveis podem ser responsáveis pelo valor reduzido do Morse de  $IU^I$ , como hidratação e a diferença eletrostática entre I e G devido ao

grupo 2-amino [35].

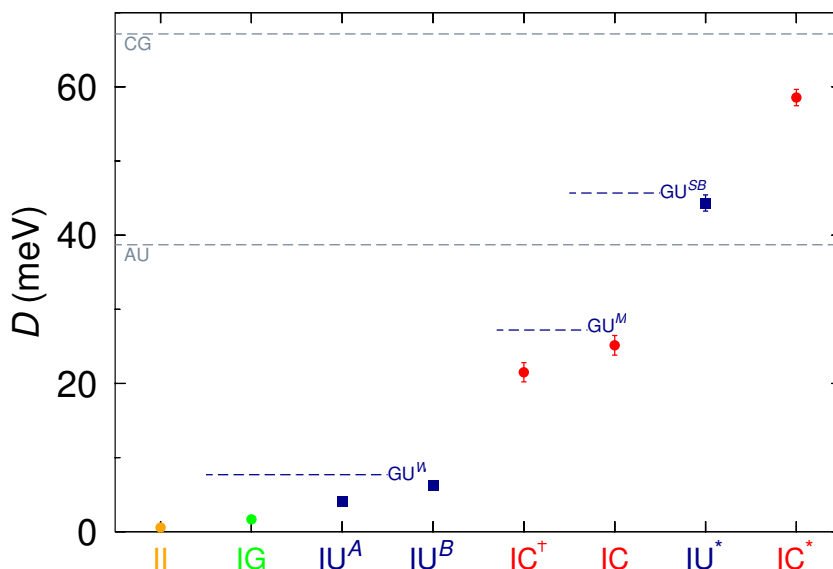
É possível que a inosina também afete toda a estrutura tridimensional e eletrônica da fita dupla, alterando de forma significativa parâmetros de Morse e *stacking* dos seus vizinhos próximos. Esta possível influência da inosina nos parâmetros de seus vizinhos não seria detectável pelos modelos NN e PB, podendo introduzir vieses no resultado final da minimização. Não foram encontradas, nem na literatura nem em nossa análise, evidências que suportem essa hipótese. A única diferença entre I e G é o grupo N2-amino, que contribui para a eletronegatividade da base. É possível que a falta desse grupo induza uma redistribuição eletrônica da fita. Isso poderia levar a uma variação dos parâmetros de Morse e *stacking* dos outros pares da sequência, principalmente nos primeiros e segundos vizinhos da inosina. Se essa redistribuição ocorresse, ela seria mais influente em sequências pequenas, como as utilizadas nesse estudo.

O contexto IU terminal,  $IU^*$ , por sua vez, apresenta altos valores para os potenciais de Morse e harmônicos, contribuindo com maior estabilidade que os pares AU e GU, sendo condizente com outros estudos [35, 73]. Além disso, identificamos uma regra de contexto para os pares IU internos,  $IU^A$  e  $IU^B$ , que diferem entre suas contribuições para a estabilidade da fita. A separação de contextos internos demonstra a importância do contexto de sequências contendo inosina. Apesar da pouca diferença na contribuição dos parâmetros de Morse, os contextos internos apresentam grandes diferenças entre os dois conjuntos.

Apenas quatro sequências com  $IU^\dagger$  foram analisadas, por esse motivo, realizamos a minimização utilizando os resultados obtidos para  $IU^B$  a fim de obter apenas os parâmetros de *stacking*. Obtivemos parâmetros de *stacking* para os vizinhos  $IU^\dagger pIU^\dagger$ ,  $UI^\dagger pIU^\dagger$  e  $IU^\dagger pUI^\dagger$ . Os resultados sugerem que esses vizinhos desestabilizam o duplexo.

Os pares IC apresentam resultados similares aos já observados [38, 56], sendo menos estável que um par GC e energeticamente equivalente aos pares GU [56, 73]. Os resultados apontam que o potencial de Morse dos pares  $IC^I$  é mais fraco quando comparado ao canônico AU. Por outro lado, pares  $IC^*$  possui um potencial mais fraco que o canônico GU e similar ao par GU terminal [73], ver Figura 4.1. Os pares IC tandem são menos estáveis que pares GC tandem e são similares às modificações não tandem, em concordância com outras evidências [38]. Além disso, foi observado que os vizinhos  $CI^\dagger pCI^\dagger$  contribuem de forma mais forte para a estabilidade com um valor de harmônico mais elevado do que os vizinhos  $CI^\dagger pIC^\dagger$  e  $IC^\dagger pCI^\dagger$ , ver Figura 4.2 e Tabela A.5, no Apêndice A.

Os efeitos terminais da inosina foram reportados previamente [35, 38]. O aumento da estabilidade do par  $IU^*$  foi atribuída à flexibilidade conformacional, que poderia permitir que o par IU assumisse conformações mais estáveis que não são acessíveis em contextos internos,  $IU^A$  e  $IU^B$  [35]. Os resultados da minimização sugerem que os pares IU contribuem com apenas uma ligação de hidrogênio a menos que o par AU e que a maior estabilidade dos contextos terminais pode ser atribuída à restauração dessa ligação. A estabilidade mais elevada



**Figura 4.1**

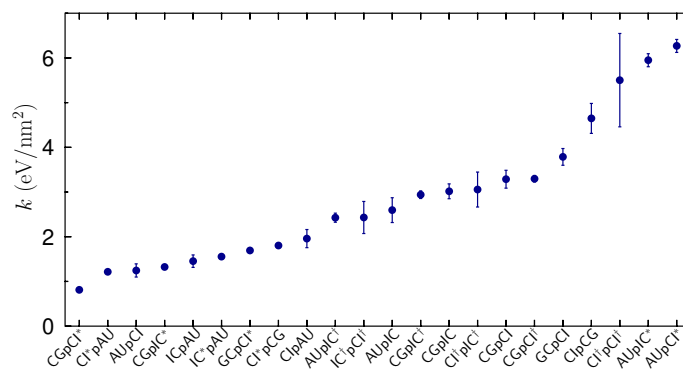
Potencial de Morse calculado para os pares IU, IC e IG em duplexo de RNA. Parâmetros dos pares  $GU^W$  (Weak),  $GU^M$  (Medium) e  $GU^S$  (Strong) retirados de Ref. [73].

do par  $IC^*$  também é atribuída à flexibilidade conformacional [38]. Nesse caso, a variação não pode ser atribuída ao ganho de uma ligação de hidrogênio, mas sim a um fortalecimento das ligações existentes. Observamos valores de harmônicos elevados para vizinhos terminais  $AUpIC^*$ ,  $AUpCI^*$ ,  $AUpIU^*$  e  $AUpUI^*$ . Essas configurações podem ser utilizadas para fim de estabilizar *primers* de RNA degenerados. Os valores de harmônico, na maioria dos casos, tendem a ser mais altos quando na vizinhança de pares GC, ver Tabela A.5, no Apêndice A.

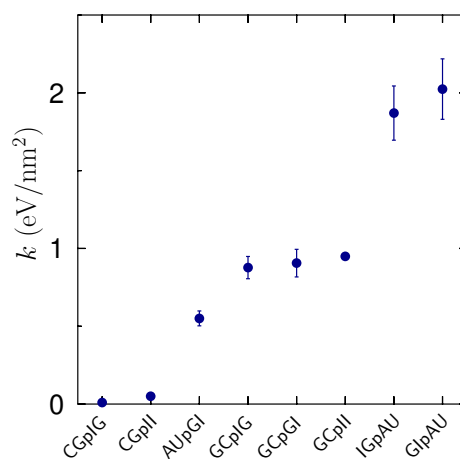
Os pares IG e II são muito similares e contribuem fracamente para estabilizar o duplexo, em concordância com estudos experimentais [59]. Parâmetros para a ligação de hidrogênio dos pares IG são mais fracos que  $IU^B$ , sugerindo que IG também possui apenas uma ligação de hidrogênio. O par II, por sua vez, apresenta um Morse aproximadamente trinta vezes menor que o par IG, o que também sugere apenas uma ligação de hidrogênio. Em geral, parâmetros de *stacking* dos pares IG e II também são menores que outros pares de inosina, o que também contribui para maior instabilidade do duplexo, ver Figuras 4.1 e 4.2 e Tabelas A.4 e A.5, no Apêndice A.

#### 4.4 Inosina em DNA e RNA

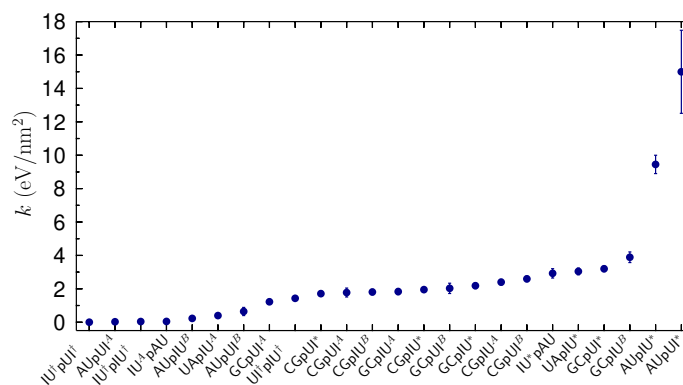
Parâmetros do modelo PB já foram obtidos para inosina em DNA em um trabalho similar [49], ver Tabela 4.2. Nesse trabalho, foram analisadas 97 sequências de DNA com modificação de inosina [48, 50, 51]. Antes de compararmos os dois estudos para cada tipo de oligonucleotídeo, precisamos considerar as principais diferenças entre os estudos. O número



(a)



(b)



(c)

**Figura 4.2**

Parâmetros de *stacking* dos pares IC, IG, II e IU. a) Par IC, b) Pares IG e II, c) par IU.

de sequências e contextos disponíveis para cada pareamento difere de forma considerável entre os estudos. Assim, é possível que cada um dos estudos possua vieses específicos associados à variáveis, como os contextos disponíveis e o tamanho das sequências. Outra diferença que deve ser considerada é a maior estabilidade do duplexo de DNA relativamente ao de RNA [2]. Isso se deve à menor reatividade da desoxirribose, permitindo que os pares estabeleçam e mantenham conformações mais estáveis. O estudo da inosina em RNA também contou com análise de contextos, separando interações internas e terminais para os pares IC e IU, o que não foi possível no estudo feito por Maximiano and Weber [49]. Por esses motivos, a comparação apresentada aqui é apenas qualitativa.

Inicialmente, iremos comparar os pares IC, IU e IT, pois são um pareamento do tipo purina-pirimidina, sendo os mais próximos dos canônicos AU, AT, CG. O par IC interno é o que apresenta potencial de Morse mais similar entre DNA e RNA, sugerindo que esse par estabiliza o duplexo com duas ligações de hidrogênio e em conformações similares em ambos os casos. Os pares IU em RNA e IT em DNA diferem, principalmente, pelo grupo C5-metil presente na timina. Esse grupo permite interações mais intensas com os pares vizinhos, discutiremos isso mais profundamente no capítulo 5. Existe uma grande diferença entre os parâmetros de Morse de IT e IU. O baixo valor de Morse de IU indica que esse par possui uma ligação de hidrogênio a menos que o par IT. Os pares IG e II são os que apresentam as maiores diferenças entre os estudos. Os parâmetros de Morse para esses pares obtidos neste trabalho sugerem que eles não formam ligações de hidrogênio ou formam ligações muito fracas em RNA. O número de dados de *melting* para sequências de RNA com estes pares é o menor disponível no conjunto utilizado para as minimizações realizadas, de tal forma, que é possível que exista uma grande influência de contexto nas sequências estudadas. Não foi possível encontrar dados publicados de resultados experimentais de  $T_m$  para o par IA em RNA. Dessa forma, não é possível estabelecer uma comparação entre os resultados. É possível, no entanto, estimar que esse par se deva comportar de forma similar aos pares II e IG devido ao fato de os três serem *mismatches* do tipo pirimidina-pirimidina.

**Tabela 4.2**

Parâmetros para o potencial de Morse para inosina em DNA. Dados retirados de Ref. [49].

Potencial de Morse para Inosina em DNA (meV)				
IA	IC	IG	IT	II
31(4)	34(2)	38(2)	29(5)	24(5)

A inosina em DNA possui a capacidade de estabelecer ligações estáveis e com contribuições de estabilidade similares ao par AU com todas as bases canônicas A, C, G e T [48, 50, 51], o que é reforçado pelos resultados de Maximiano and Weber [49]. É geralmente encontrada na forma livre, desoxiinosina (dI), atuando como uma molécula central na síntese e degradação de purinas [27]. Pode ser incorporada de forma errônea, com baixa taxa de ocorrência, no DNA nuclear dos eucariontes, sendo rapidamente reconhecida e removida pelos mecanismos

de reparo, por meio de enzimas como reparo de excisão de base (*base excision repair*, BER) e endonuclease V [27, 47]. Apesar da alta afinidade da inosina em DNA, geralmente, ela é incorporada no duplexo em oposição a uma citosina [47]. IC possui uma ligação de hidrogênio a menos que GC promovendo aberturas da fita e podendo levar a alterações nos sítios de reconhecimento de proteínas que se ligam em DNA [47]. Isso pode tanto viabilizar quanto inviabilizar as ligações DNA-proteína em uma determinada região genômica, potencialmente afetando a expressão gênica. Não encontramos nenhum estudo sobre como a inosina é reconhecida durante o processo de transcrição. No entanto, os resultados apresentados por Maximiano and Weber [49] sugerem que durante a transcrição pode ser reconhecida como qualquer uma das quatro bases canônicas. Essa possibilidade implicaria que a transcrição de uma sequência de DNA modificada com I poderia resultar em  $4^n$  transcritos, onde  $n$  é a quantidade de inosinas presentes na sequência. São necessários mais estudos para determinar se realmente é possível a transcrição de sequências de DNA contendo inosina em contextos biológicos e como se dá a codificação da informação nesses casos. No RNA, por sua vez, a inosina apresenta uma menor afinidade se ligando, preferencialmente, às bases C e U. É uma modificação essencial, como já foi discutido, podendo ser interpretada durante a tradução de forma majoritária como uma guanina e em menor escala como uma adenina, sendo um mecanismo de variabilidade genética. A ocorrência da inosina em DNA é rapidamente corrigida por mecanismos celulares de correção, por outro lado, em RNA, é uma modificação abundante [47]. Podemos levantar a hipótese de que as diferenças entre os pareamentos mais favoráveis para a inosina em DNA, quando comparados aos de RNA, sejam um dos fatores para a ocorrência dessa base ser evolutivamente preservada em RNA, ao mesmo tempo que é descartada no DNA.

## 4.5 Perfis de abertura e contextos biológicos

Os parâmetros obtidos foram utilizados para calcular os chamados perfis de abertura. Os perfis de abertura são as distâncias médias  $\langle y_n \rangle$ , na direção da ligação de hidrogênio, entre duas bases de um par. Assim, é possível estudar de forma qualitativa a influência da inosina em diferentes duplexos e seu impacto na estrutura e seus papéis biológicos.

A modificação não sinônima de inosina em mRNA foi associada com preeclampsia severa precoce quando ocorre na posição 49360951 do cromossomo 3 (chr3:49360951) do gene Membro da Família Homóloga Ras A (*Ras Homolog Family Member A*, RHOA) e foi identificado como danosa à proteína pelo software *Sorting Intolerant from Tolerant* (SIFT) [74]. A tradução não sinônima nessa posição altera o códon UAC, que codifica tirosina, para um códon UIC, que codifica cisteína. Apesar do mRNA ser mais comumente encontrado na conformação de fita simples, existem mecanismos de degradação como, RISC, que dependem da hibridização de alvos de mRNA com miRNA e siRNA [75]. Na Figura 4.3, vemos o comportamento da modificação em um duplexo de mRNA do gene RHOA com uma distância de 10 pb da posição

chr3:49360951, uma sequência de 21 pb, similar a um siRNA. Observa-se que o par IU introduz uma grande abertura interna na região modificada do duplexo, enquanto o par IC apresenta uma abertura pequena e não muito diferente da abertura do par AU. A instabilidade produzida pela inosina quando pareada com U pode permitir que sequências modificadas escapem do RISC e outros mecanismos regulatórios, possibilitando uma expressão ainda maior das proteínas modificadas. Por outro lado, o pareamento IC poderia ser aplicado como uma ferramenta para a hibridização dessa região em terapias baseadas em miRNA e siRNA para o tratamento desse tipo de condição. O modelo PB foi utilizado para prever temperaturas de *melting* da interação de um duplexo com 7 pb na região chr3:49360951, ver Tabela 4.3.

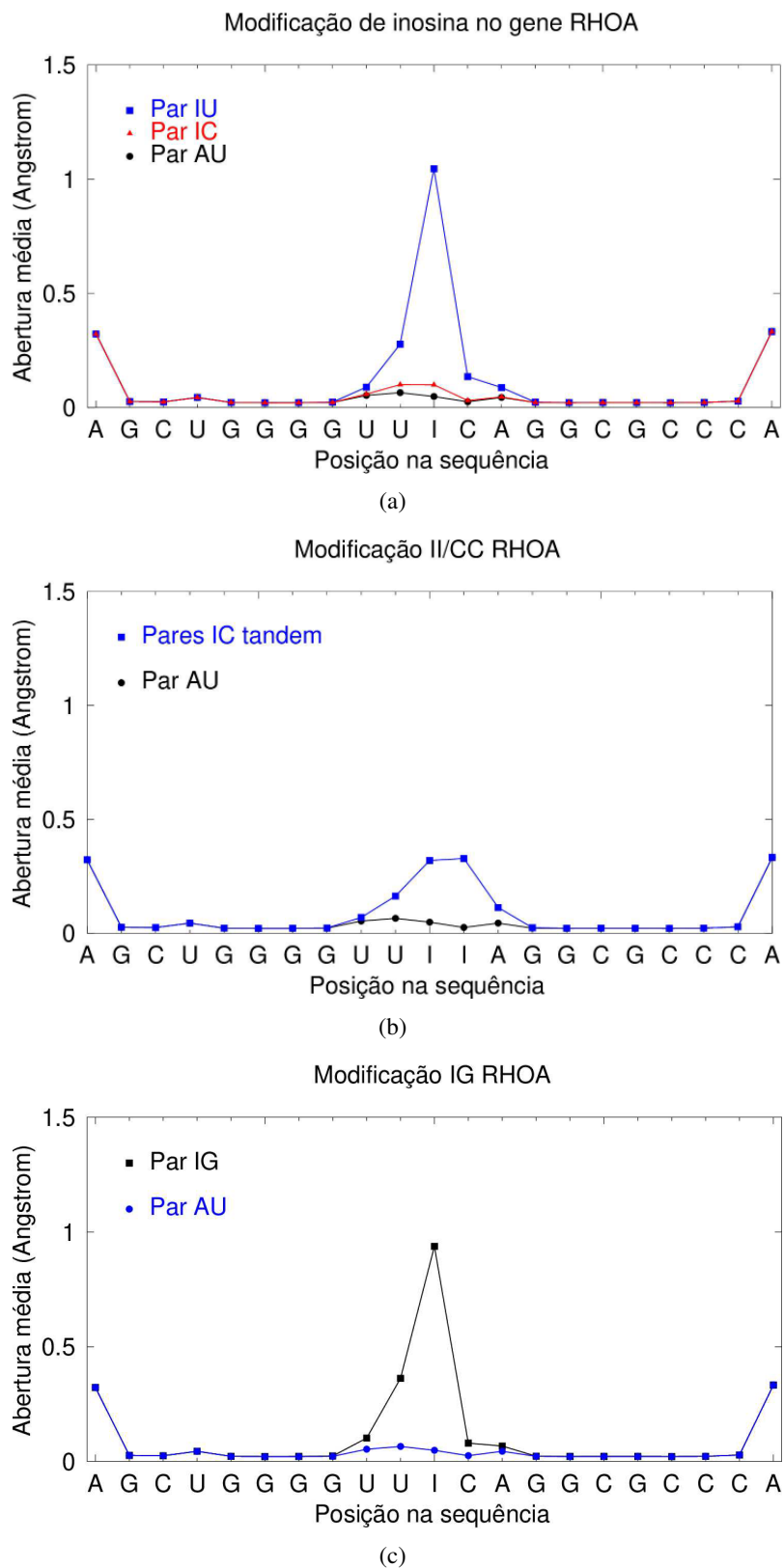
**Tabela 4.3**

Temperatura de *melting*  $T_{\text{pred.}}$  predita para sequências não modificadas e modificadas em uma hibridização de RNA de 7 pares de bases do gene RHOA na posição chr3:49360951. Cálculo realizado para concentração da solução de 1000 M e concentração das fitas de 100  $\mu\text{M}$ . A base alterada está em negrito.

Sequências 5' $\rightarrow$ 3' / 3' $\rightarrow$ 5'	$T_{\text{pred.}}$ ( $^{\circ}\text{C}$ )
GUUACAG/CAAUGUC	41.1
GUUGCAG/CAACGUC	50.9
GUU <b>I</b> CAG/CAACGUC	35.5
GUU <b>I</b> CAG/CAAUGUC	23.5

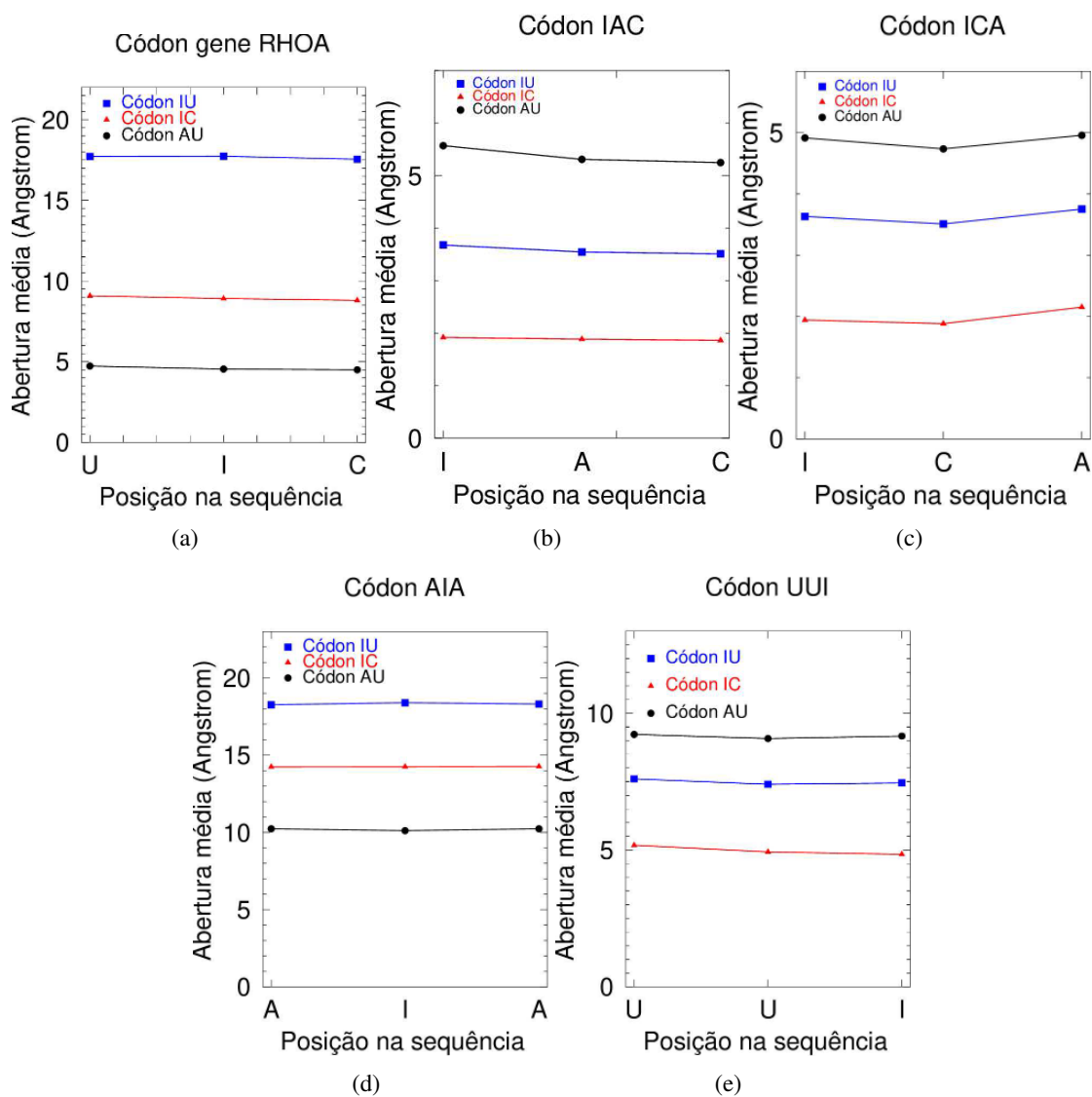
A inosina pode ser interpretada como A, G ou U durante a tradução e pode induzir a parada do ribossomo [23]. A fim de estudar a tradução de sequências modificadas, consideramos que durante o processo de tradução a interação códon-anticódon se comporta como dsRNA. O perfil de abertura do códon modificado UIC do RHOA destaca a ruptura da interação causada pelos pares IU internos, ver Figura 4.4. Além disso, esse códon causa parada ribossomal com uma taxa de 5% [23]. Pares IC também introduzem instabilidade, no entanto, possuem uma interação mais forte, sendo suficiente para favorecer a ligação com o tRNA ACG permitindo a ocorrência da tradução não sinônima. Apesar de nesse caso a modificação introduzir instabilidade e permitir traduções não sinônimas, ela também pode introduzir estabilidade e ser traduzida tanto de forma sinônima quanto não sinônima [23]. Em especial, em contextos terminais nos quais a inosina possui um valor mais alto de Morse.

Inosina na posição terminal possui uma taxa menor de parada ribossomal [23], exceto no caso UUI. Os parâmetros PB obtidos neste trabalho também foram utilizados para analisar abertura de alguns códons contendo modificação e suas taxas de parada associadas [23], ver Figura 4.4 e Tabela 4.4. A maioria das diferenças nas taxas de parada ribossomal pode ser explicada pelas diferenças nas interações internas e terminais, e a abertura média da seção códon-anticódon também parece influenciar esse processo. O trímero UUI possui um comportamento atípico, quando comparado com outros trimeros estudados por Licht et al. [23]. Apesar de nesse códon pares IU e IC serem mais fortes que AU, ele possui uma taxa de parada de 49%.



**Figura 4.3**

Perfil de abertura de duplexos de mRNA do gene RHOA com e sem modificações, flanqueados por 10 pares de bases em ambos os lados, na posição chr3:49360951. Destacando a sequência com pareamentos a) IU (preto), pareamento IC (vermelho) e canônico AU (azul), b) IC tandem (preto), c) IG (preto).

**Figura 4.4**

Perfil de abertura para alguns códons com e sem a modificação. Pareamento de inosina com U (azul) e C (vermelho) e códon com pareamento canônico AU (preto).

Esse comportamento não é esperado e sugere que a abertura códon-anticódon não é suficiente para explicar completamente a parada ribossomal induzida pela modificação.

**Tabela 4.4**

Abertura média de códons contendo uma inosina, onde Tx significa "taxa".

Aberturas dos códons					
Códons	Abertura média (Å)			Tx parada	Tx sinônima
	IC	IU	AU		
UIC	8.9	17.6	4.6	5%	–
IAC	1.9	3.6	5.4	1%	25%
ICA	2.0	3.6	4.9	3%	2%
AIA	14.3	10.2	3.7	7%	–
UUI	5.0	7.5	9.2	49%	–

O códon IAC possui uma taxa de tradução sinônima de 25% [23], ver Figura 4.4. Quando comparado com o códon ICA (taxa de tradução sinônima de 1%), ambos os códons são similares com taxas de parada similares e seus perfis de abertura não diferem muito. A inosina em ambos os casos cumpre um papel estabilizador para o códon, apresentando perfis muito parecidos. No entanto, o códon IAC é mais fortemente estabilizado pela modificação do que o códon ICA. As diferenças nos perfis de abertura surgem da diferença entre as contribuições dos vizinhos AU<sub>upCG</sub> quando comparados com os vizinhos CG<sub>pAU</sub> e também do menor potencial de Morse do par AU ao invés do GC na terceira posição do códon.

## 4.6 Conclusão

O modelo PB foi utilizado para obter parâmetros de interação para inosina em sequências de RNA. Foram analisadas 71 sequências contendo inosina ligada com as bases C, G, I e U, incluindo pares tandem [35, 38, 59]. Observou-se que a inosina é capaz de estabilizar duplexos de RNA, no entanto, na maioria dos casos, essa base vai induzir instabilidade na sequência. Apesar de vizinhos GC aumentarem a contribuição da inosina para a estabilidade, pares contendo inosina serão menos estáveis que suas contrapartes canônicas, exceto no caso terminal. Além disso, mostramos que a contribuição da inosina é dependente do contexto. Os parâmetros obtidos e apresentados aqui permitem uma visão qualitativa de sua função biológica, como o comportamento de códons modificados de sequências de mRNA em nível celular. Os resultados apresentados aqui podem ser aplicados para melhorar técnicas moleculares, como design de *primers* de PCR e terapias baseadas em mRNA. No entanto, ainda faltam contextos completos de vizinhança e emparelhamento que possam estender esta investigação termodinâmica.

## 5 Pseudouridina

Neste capítulo, será abordada a aplicação do modelo PB às sequências de RNA com uma modificação de pseudouridina; os procedimentos de minimização para obtenção dos parâmetros e concluindo com a discussão dos resultados obtidos.

### 5.1 Introdução

As contribuições de estabilidade da pseudouridina ( $\Psi$ ) para o duplexo de RNA já foram avaliadas utilizando o modelo NN e modelos de mecânica quântica/mecânica molecular (*quantum mechanics/molecular mechanics*, QM/MM) [76–78]. Dutta et al. [31] estudou interações intramoleculares de N1-metilpseudouridina ( $m^1\Psi$ ) em RNA através de modelos QM/MM. O modelo NN não descreve de forma satisfatória as interações intramoleculares dos pares e seus vizinhos e seus parâmetros fornecem pouca informação sobre a estrutura do duplexo. Modelos de QM/MM são capazes de avaliar as interações intramoleculares, mas seus resultados são muito dependentes da parametrização utilizada nos campos de força de cada software. Apesar de confiáveis para predição estrutural e representação das interações intramoleculares, os modelos de QM/MM, geralmente, são baseados apenas em cálculos computacionais, sem capacidade de comparação direta com resultados experimentais, o que não permite análise de contextos nem a predição de  $T_m$ . O modelo PB permite obter informações acerca das interações intramoleculares, a predição de temperaturas de *melting*, análise de contextos e o estudo da estrutura de duplexos com modificações. Foram realizados protocolos para obtenção de parametrização dos pares  $\Psi A$ ,  $m^1\Psi A$ ,  $\Psi G$ ,  $\Psi C$  e  $\Psi U$ .

### 5.2 Procedimentos de minimização

A fim de obter parâmetros de Morse ( $D$ ) e de *stacking* ( $k$ ) para pseudouridina e N1-metilpseudouridina em RNA, foram utilizados procedimentos similares aos usados para inosina. Foram coletados dados publicados de temperaturas de *melting* com as mesmas condições experimentais (NaCl 1 M, cacodilato de sódio 20 mM, Na<sub>2</sub>EDTA 0,5 mM, pH 7,0 e concentração de oligonucleotídeos de 100  $\mu$  M) dos quais obtivemos 41 sequências [77, 78] contendo o par  $\Psi A$  (29 sequências) e os *mismatches*  $\Psi C$  (3),  $\Psi G$  (6) e  $\Psi U$  (3), ver Tabela B.1, no Apêndice B. Dados experimentais de *melting* para sequências contendo N1-metilpseudouridina foram obtidos em colaboração com o KAS Group, liderado pela Profa. Astakhova. Foram produzidas 9 sequências de  $m^1\Psi A$  para as quais foram obtidas temperaturas de *melting*. As temperaturas de *melting* para essas sequências foram calculadas utilizando a mesma concentração de solução das temperaturas publicadas. Os experimentos foram realizados em duas concentrações de nucleotídeos, 5 e 10  $\mu$ M, e as minimizações foram realizadas para cada concentração. As

sequências foram divididas em cinco grupos,  $\Psi A$  e  $\Psi C$ ,  $\Psi G$ ,  $\Psi U$  e  $m^1\Psi A$ . Temperaturas de *melting* para sequências canônicas equivalentes de referência para cada grupo foram produzidas utilizando o modelo NN. As sequências de referência para os grupos  $\Psi A$  e  $m^1\Psi A$  foram produzidas substituindo a modificação por uridina. Nos grupos  $\Psi C$ ,  $\Psi G$  e  $\Psi U$ , as sequências de referência foram obtidas substituindo os pares por um par CG. Resultados experimentais obtidos por Kierzek et al. [77] para sequências canônicas, com o par AU substituindo a modificação, também foram usadas como referência quando possível. A lista completa de sequências de referência utilizadas é apresentada na Tabela B.2, no Apêndice B.

As minimizações foram feitas a partir do conjunto padrão de parâmetros ( $D = 38.0 \text{ meV}$ ,  $k = 2.5 \text{ eV/nm}^2$  e  $\lambda_a = 0.29 \text{ nm}$ ) para a primeira rodada de minimização. O valor de  $\lambda_a$  foi mantido constante em todas as minimizações. Para cada rodada de minimização, foi utilizada uma randomização dos parâmetros iniciais de 20%. A cada rodada global foram realizadas 100 minimizações locais e o resultado final de parâmetros é a média dos 20 melhores resultados. As rodadas globais conduzidas após a primeira utilizaram como parâmetros iniciais os resultados da rodada anterior. Minimizações locais foram realizadas até que a diferença do  $\chi^2$  entre dois passos locais seja menor que  $0.1 \text{ }^\circ\text{C}^2$ . Minimizações globais foram conduzidas até que não houvesse melhora significativa no valor de  $\chi^2$ , ver Tabela 5.1.

**Tabela 5.1**

O valor de  $\chi^2$  em cada rodada de minimização para cada grupo contendo pseudouridina. a) grupo  $\Psi A$  com separação de contexto interno  $\Psi^I A$  e terminal  $\Psi^* A$ .

Rounds	Progressão de $\chi^2$ ( $^\circ\text{C}^2$ )					
	$\Psi A$	$\Psi A^a$	$m^1\Psi A$	$\Psi C$	$\Psi G$	$\Psi U$
r1	228.7	106.2	42.9	0.009	22.7	0.007
r2	184.9	103.6	49.1	0.01	22.7	0.013
r3	179.6	102.5	18.9	$8.0 \cdot 10^{-7}$	22.7	0.0001
r4	–	–	–	–	4.0	–
r5	–	–	–	–	4.2	–

### 5.3 Design das sequências de N1-metilpseudouridina

As sequências de N1-metilpseudouridina e os experimentos de *melting* utilizados neste estudo foram produzidas pelo KAS Group. Foram produzidas 9 sequências de 7 nucleotídeos com uma  $m^1\Psi$  central, com conteúdo GC e concentrações similares às utilizadas em Hudson et al. [78]. Com objetivo de avaliar de forma mais eficiente os contextos, escolhemos o trímero central como  $X(m^1\Psi)Y$ , com X e Y podendo ser A, C, G ou U. Os trímeros foram escolhidos de forma que não houvesse sequências com trímeros repetidos. Dessa forma, fomos capazes de pelo menos duas sequências representantes de cada parâmetro necessário para minimização, assim evitamos o problema descrito na seção 3.5.

### 5.3.1 Grupo $\Psi A$ sem contexto

Três rodadas de minimização foram realizadas para esse grupo. Nesse grupo não foi utilizado separações de contexto.

**Grupo  $\Psi A$  com contextos interno e terminal** Três rodadas de minimização foram realizadas para esse grupo. Essa linha de minimização utilizou as regras de contextos para separar as variáveis em dois grupos, terminal e interno,  $\Psi^*A$  e  $\Psi^I A$ , respectivamente.

**Grupo  $m^1\Psi A$**  Neste grupo, nenhuma regra de contexto foi utilizada. Para cada concentração, três rodadas foram realizadas.

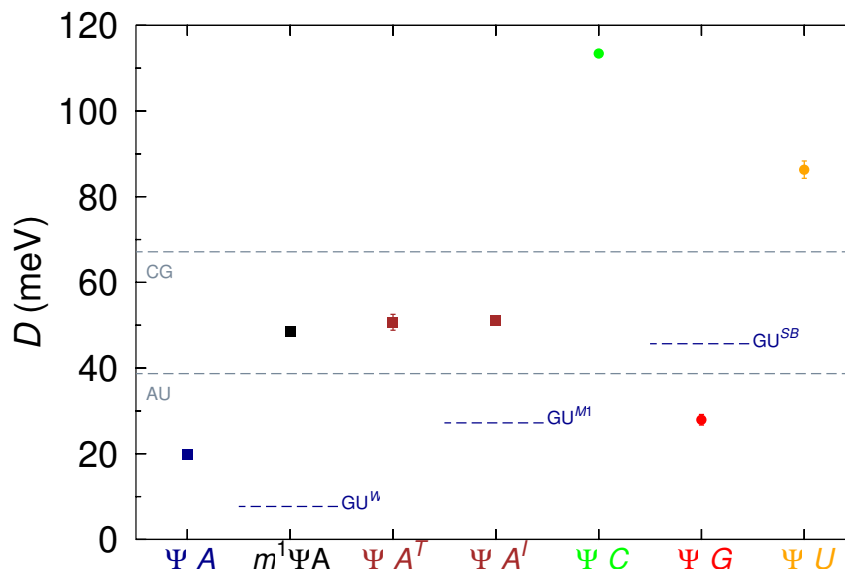
**Grupo  $\Psi G$**  Cinco rodadas de minimização foram realizadas para esse grupo. Neste grupo, haviam mais parâmetros a serem minimizados do que sequências disponíveis. Além disso, as sequências eram muito poucas para realizar adequadamente a separação por contexto. Dessa forma, os parâmetros de *stacking* dos vizinhos  $AUpG\Psi$  e  $AUp\Psi G$ ,  $CGpG\Psi$  e  $CGp\Psi G$ ,  $GCpG\Psi$  e  $GCp\Psi G$  foram associados, sendo variados em conjunto durante o processo de minimização, como explicado na seção 3.5. A partir da quarta rodada, os parâmetros associados foram permitidos variarem separadamente, enquanto os parâmetros de Morse  $\Psi G$  e de *stacking* para os vizinhos  $G\Psi pAU$  e  $UApG\Psi$  foram mantidos fixos. A quinta e última rodada utilizou os resultados da quarta rodada como parâmetros iniciais.

**Grupo  $\Psi C$  e Grupo  $\Psi U$**  Esses grupos foram minimizados separadamente, mas pela similaridade dos dados foram utilizados o mesmos procedimentos. Três rodadas de minimização foram realizadas para esses grupos. Nestes grupos, as sequências contendo as variáveis  $\Psi C$  e  $\Psi U$  foram minimizadas.

### 5.3.2 Desvio padrão

O desvio padrão de cada grupo  $\Psi$  foi determinado de forma similar ao apresentado na seção 4.2.5. Uma rodada extra de minimização foi realizada para cada grupo, exceto o grupo  $\Psi G$ . Nesta rodada, foram realizadas 100 minimizações. Nenhum erro experimental foi fornecido para o conjunto de dados utilizados neste trabalho. Então, em cada minimização foi utilizada uma randomização do conjunto de dados de  $T_m$  em  $\pm 0,5^\circ C$  e nenhuma randomização inicial para os parâmetros de entrada. A média final e o desvio padrão, de cada parâmetro, foram obtidos a partir dos resultados desta rodada.

Devido a associação dos parâmetros, duas rodadas foram necessárias para obtermos o desvio padrão para os parâmetros do grupo  $\Psi G$ . Apenas os parâmetros de *stacking*  $AUpG\Psi$ ,



**Figura 5.1**

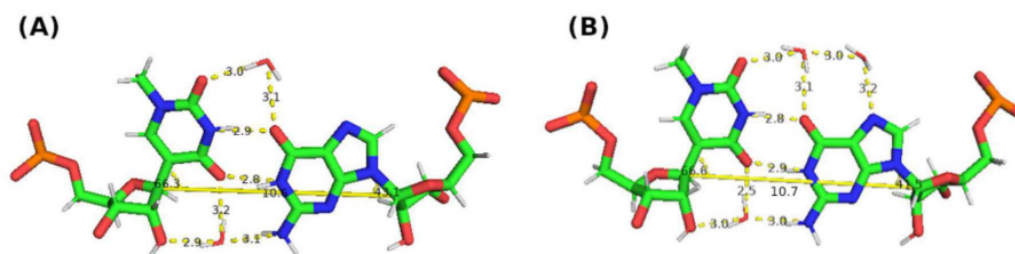
Potencial de Morse calculado para os pares  $\Psi A$ ,  $\Psi C$ ,  $\Psi G$ ,  $\Psi U$  e  $m^1\Psi A$  em um duplexo de RNA. Parâmetros dos pares  $GU^W$  (Weak),  $GU^M$  (Medium) e  $GU^S$  (Strong) retirados de Ref. [73].

$AU\psi G$ ,  $CG\psi G$ ,  $CG\psi\Psi$ ,  $GC\psi\Psi$  e  $GC\psi G$  foram minimizados na primeira rodada. Os parâmetros restantes, Morse  $\Psi G$  e *stacking*  $G\Psi pAU$  e  $UApG\Psi$ , foram minimizados a partir dos resultados da primeira rodada.

## 5.4 Resultados

Obtivemos parâmetros para o potencial de Morse e harmônico para pseudouridina e N1-pseudouridina a partir de resultados experimentais de *melting* de 50 sequências de RNA contendo os pares  $\Psi A$ ,  $m^1\Psi A$ ,  $\Psi C$ ,  $\Psi G$  e  $\Psi U$ . A comparação entre os resultados preditos pelo modelo Peyrard-Bishop e os resultados experimentais é mostrado na Tabela B.1, no Apêndice B. Os resultados finais da minimização são apresentados nas Figuras 5.1 e 5.3. Os parâmetros apresentados para  $m^1\Psi$  foram obtidos para concentração de fitas de  $5 \mu M$ . Tabelas com os resultados finais de todos os conjuntos de minimização são apresentadas no Apêndice B, ver Tabelas B.1 B.4, B.5, e B.4, incluindo os resultados para  $m^1\Psi$  com concentração das fitas de  $10 \mu M$ .

Experimentos de *melting* [77–79], ressonância magnética nuclear (*nuclear magnetic resonance*, NMR) e estudos de Dinâmica Molecular (*Molecular dynamics*, MD) [31, 76] demonstram que a pseudouridina e N1-metilpseudouridina contribuem fortemente para a estabilidade do duplexo, inclusive no caso dos *mismatches*. Muitos fatores parecem contribuir para a maior estabilidade proporcionada pela pseudouridina em relação à uridina. A ligação C5-glicosídeo proporciona aumento dos ângulos torcionais possíveis [22, 77], permitindo que a base assum



**Figura 5.2**

Pontes de água podem agir como intermediárias para ligações de hidrogênio. A imagem ilustra a interação mediada por água de um par  $m^1\Psi G$ : a) duas moléculas chave de água atuam na interação; b) três moléculas de água mediam a interação. Figura retirada de Ref. [31]

conformações mais estáveis sem afetar a estrutura geral do duplexo. Além disso, a pseudouridina conta com um grupo imino na posição N1, sendo um doador de ligações de hidrogênio, o grupo metil de  $m^1\Psi$  atua de forma similar [22, 31, 76, 77]. Dessa forma, a pseudouridina possui dois doadores e dois aceptores para ligações de hidrogênio. Isso faz dela uma base promíscua permitindo estabilizar duplexos pareando com outras bases além de adenina [22, 77]. Dutta et al. [31] e Deb et al. [76] também indicam que os grupos imino de  $\Psi$  e metil de  $m^1\Psi$  permitem ligações internas com moléculas de água, chamadas pontes de água. Nessa interação, a molécula de  $H_2O$  atua como um intermediário para ligações de hidrogênio, permitindo ligações entre as bases que de outra forma seriam impossíveis [31, 76], ver Figura 5.2.

Os experimentos de melting [77, 78] indicam pouca diferença energética de ligação entre os pares  $\Psi A$  internos e terminais, aproximadamente 0.7 kcal/mol. Esse valor indica que não existem perdas ou ganhos na quantidade de ligações de hidrogênio desse par. É esperado que pares em posições terminais contribuam de forma mais fraca com a estabilidade da fita, exceto no caso dos pares IC e IU, como visto no capítulo 4. Isso se deve a falta de um vizinho que estabilize o par, proporcionando uma abertura da fita na posição terminal. Duas linhas de minimização foram realizadas a fim de avaliar diferenças entre o contextos internos e terminais. O primeiro conjunto de rodadas de minimização considera que as contribuições internas e terminais são as mesmas, o segundo considera que contextos internos e terminais possuem contribuições diferentes.

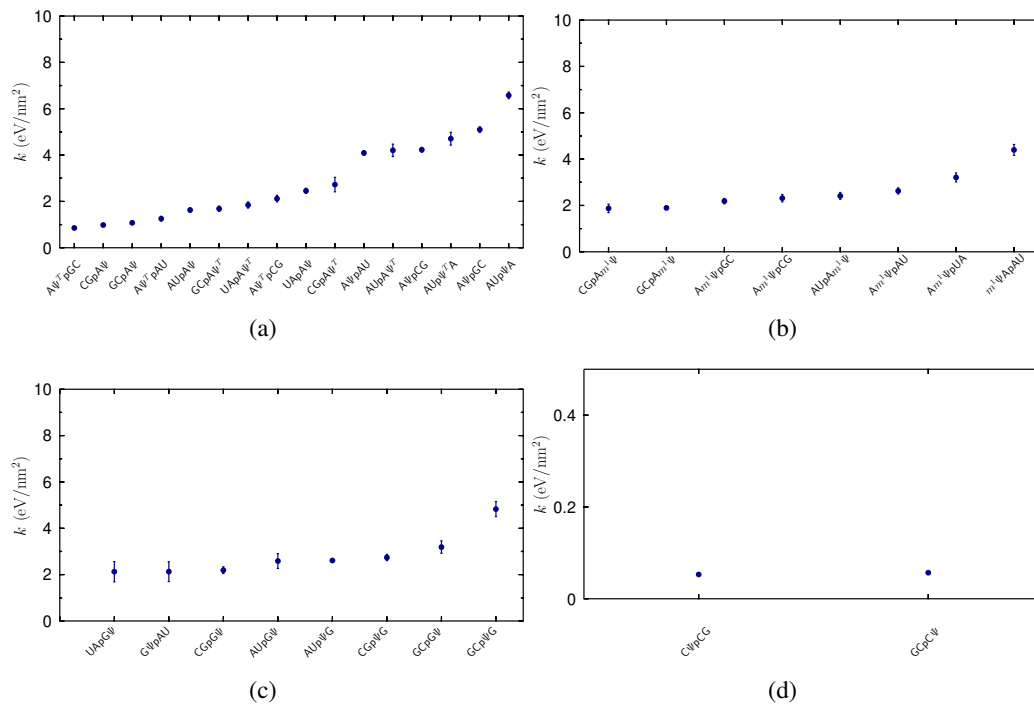
O conjunto sem contexto apresenta valores abaixo do par canônico AU para o potencial de Morse e valores muito elevados para os potenciais harmônicos. Esse resultado destoa dos resultados do modelo de próximos vizinhos obtidos por Kierzek et al. [77] e Hudson et al. [78]. O valor de  $\chi^2$  para esse conjunto é alto em relação ao conjunto que utiliza a separação por contextos, ver Tabela 5.1. Dessa forma, consideramos esse resultado pouco representativo do sistema físico real. Por esse motivo, a partir de agora vamos discutir apenas os resultados para o grupo com contexto. Os resultados desse grupo são apresentados no Apêndice B.

O conjunto que utiliza regras de contexto apresenta um Morse bem elevado quando com-

parado com o par AU. Tanto no caso terminal quanto no caso interno, as principais diferenças são observadas nos parâmetros harmônicos. Os resultados completos são apresentados no Apêndice B. Os resultados obtidos para o conjunto onde há a separação de contextos são condizentes com os resultados experimentais [77, 78] e com a presença de duas ligações de hidrogênio para o par  $\Psi A$ . Além disso, são uma melhor representação do sistema físico real. Assim, esse é considerado o resultado final e o que será utilizado nas versões futuras do TfReg. Apesar dos parâmetros de Morse  $\Psi^* A$  e  $\Psi^I A$  serem similares, os valores dos potenciais harmônicos são bem diferentes entre esses dois contextos. Isso se deve, provavelmente, ao grupo imino, por ser um doador para ligações de hidrogênio, esse grupo pode competir com pares vizinhos por seus aceptores de ligações de hidrogênio. Também pode ser o caso desse grupo interagir de forma repulsiva com outros doadores. A pseudouridina no contexto terminal estabiliza o duplexo de forma diferente a depender de qual ponta, 5' ou 3', ela está e de seus vizinhos.

O dímero GCpA $\Psi^*$ , que representa um par  $\Psi A$  na ponta 5' com um vizinho 3'-CG, possui o dobro do valor para o potencial harmônico do dímero equivalente à A $\Psi^*$ pGC,  $\Psi A$  na ponta 3' com vizinho CG, por exemplo. Essa tendência se mantém para os outros pares terminais, um par  $\Psi A$  na ponta 5' do duplexo contribui mais fortemente para a estabilidade do duplexo do que o mesmo par na ponta 3'. O *stacking* do par  $\Psi^I A$  também apresentou dependência relativa às bases vizinhas 5' e 3'. Estudos de estabilidade utilizando NMR e QM/MM [76] demonstram que as interações de *stacking* são estabilizadas quando  $\Psi$  está na posição 5' relativa às bases A, G e U, e quando está na posição 3' de G. Também foi observado que a interação de *stacking* não é estabilizada quando  $\Psi$  está na posição 3' de A, C e U, e quando está na posição 5' de C. Os valores para os potenciais harmônicos obtidos aqui são condizentes com o que foi demonstrado por Deb et al. [76], ver Figura 5.3.

A  $m^1\Psi$  tem se mostrado muito eficiente para aplicações biotecnológicas, em especial, em vacinas de mRNA [32]. Apesar disso, até o desenvolvimento deste trabalho, nenhum estudo de *melting* havia sido publicado para essa modificação. Os únicos estudos encontrados acerca da estabilidade térmica da N1-metilpseudouridina são estudos de QM/MM [31]. Temperaturas de *melting* para sequências de RNA com  $m^1\Psi$  foram obtidas em colaboração com o grupo da Profa. Astakhova. O potencial de Morse obtido para  $m^1\Psi$  é próximo, porém menor que o Morse da pseudouridina sem metilação, indicando que a quantidade de ligações de hidrogênio é preservada após a modificação, em concordância com estudos anteriores [31, 76–78]. Os valores de *stacking* sofrem a maior variação quando comparamos os pares  $m^1\Psi A$  e  $\Psi A$ . Os resultados indicam que a metilação contribui com a estabilidade dos vizinhos na posição 5' da modificação, por outro lado, reduz a contribuição de estabilidade dos vizinhos 3', uma tendência similar foi observada por Dutta et al. [31]. Apesar disso, nossos resultados sugerem que os pares  $m^1\Psi A$  contribuem de forma mais fraca para a estabilidade do duplexo quando comparada com pares  $\Psi A$ , o que entra em contradição com o resultado apresentado por Dutta et al. [31]. A aparente



**Figura 5.3**

Parâmetros de *stacking* para os vizinhos dos pares (a)  $\Psi A$ , (b)  $m^1\Psi A$ , (c)  $\Psi G$  e (d)  $\Psi C$  e  $\Psi U$ . As barras nos pontos indicam o desvio padrão.

contradição entre os dois estudos pode ser explicada pela pouca amostragem experimental disponível para  $m^1\Psi$  nesse estudo, pelas diferenças entre os modelos utilizados e por possíveis regras de contexto não identificadas. Apesar disso, ambos os estudos apontam para tendências similares das interações de  $\Psi$  e  $m^1\Psi$ .

O grupo de *mismatches* é o que possui menor quantidade de dados de  $T_m$  experimentais. Isso dificulta o processo de minimização e pode, muitas vezes, levar à resultados que só se aplicam às sequências utilizadas. Observamos valores muito elevados para os Morses  $\Psi C$  e  $\Psi U$ , ver Figura 5.1. Esses valores são maiores do que os valores do par GC que possui três ligações de hidrogênio. Kierzek et al. [77] mostra que as contribuições desses pares para a estabilidade é maior que as contribuições de UC e UU, mas não apresenta comparação com o pareamento canônico. No entanto, isso não justifica os altos potenciais de Morse observados neste estudo. Nesses dois grupos, apenas três sequências de cada estavam disponíveis, além disso, duas delas são sequências terminais. O algoritmo de minimização ao percorrer um espaço de parâmetros com poucas variáveis converge de forma muito rápida para mínimos globais de  $\chi^2$ . As poucas sequências e a dificuldade de separar os contextos internos e terminais dessas sequências podem introduzir vieses no protocolo de minimização. Esse parece ser o caso para  $\Psi C$  e  $\Psi U$ . Para avaliar se esse é o caso, basta a comparação direta com a  $T_m$  da sequência análoga canônica disponível em Kierzek et al. [77], ver Figura 5.2.

Vemos que tanto  $\Psi$ C quanto  $\Psi$ U afetam muito fracamente a estabilidade da fita quando ocorrem nas posições terminais. Por outro lado, afetam significativamente a estabilidade quando ocorrem na posição interna. O valor do Morse domina a equação, sendo relevante quando se calcula  $T_m$ . Como as quatro sequências com modificações terminais apresentam valores muito próximos da sequência canônica, a minimização rapidamente converge para valores muito altos de Morse, enquanto os valores de *stacking* se tornam muito baixos. Nesse caso, os valores obtidos para esses pares não podem ser considerados como representantes do sistema físico real. Assim, não podemos tirar conclusões sobre as interações desses pares e seus impactos estruturais. Também não é possível dizer se estes parâmetros são suficientes para o cálculo de  $T_m$  de outras sequências com os mesmos contextos das utilizadas aqui.

**Tabela 5.2**

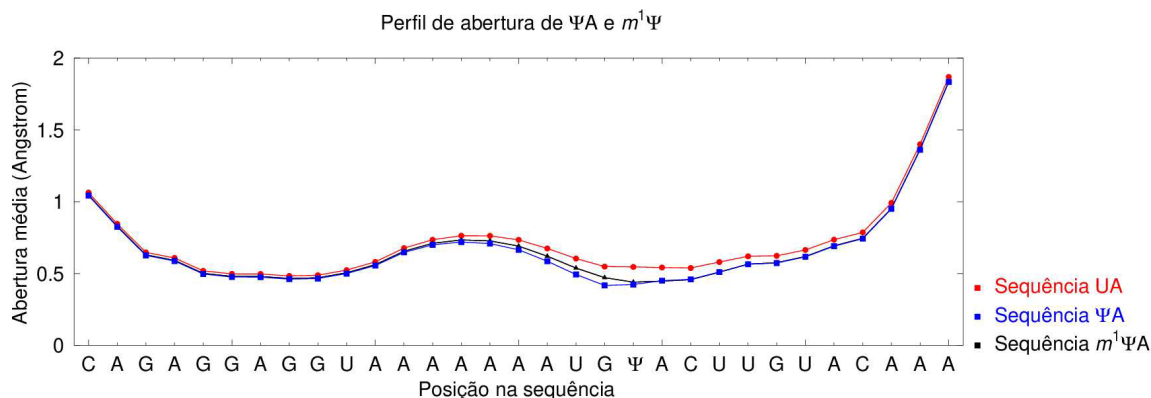
Temperaturas de *melting* de sequências com  $\Psi$ C,  $\Psi$ U e o análogo canônico UA.

Sequência 5' → 3' / 3' → 5'	$T_m$ (°C)
UCAGUCAGU/AGUCAGUCA	59.1
UCAG $\Psi$ CAGU/AGUC $\Psi$ GUCA	45.6
$\Psi$ CAGUCAGU/CGUCAGUCA	57.2
UCAGUCAG $\Psi$ /AGUCAGUCC	57.4
UCAG $\Psi$ CAGU/AGUCUGUCA	49.9
$\Psi$ CAGUCAGU/UGUCAGUCA	57.5
UCAGUCAG $\Psi$ /AGUCAGUCU	57.5

A minimização para o par  $\Psi$ G, apesar de também possuir poucas sequências, apresentou resultados mais consistentes com o esperado para esse *mismatch* [77].  $\Psi$ G apresenta duas ligações de hidrogênio muito similares ao par AU, são elas, o grupo N3-imino de  $\Psi$  com o O6-carbonila de G e O2-carbonila de  $\Psi$  com N1 de G e N3-imino de U com N1 de A N-amino A com O2-carbonila de U. O valor de Morse para esse par é similar ao de um par AU, porém menor, o que é condizente com duas ligações de hidrogênio. Os valores de *stacking* também estão dentro do esperado. Os resultados para esse par apontam para uma contribuição relativamente alta para a estabilidade do duplexo e pode ser uma forma de justificar a capacidade da pseudouridina promover a supressão de códons de parada [21].

## 5.5 Perfis de abertura e vacinas de mRNA

Vacinas de mRNA e outras terapias de mRNA são sequências específicas de RNA que codificam proteínas de interesse utilizando o ribossomo da própria célula do paciente [22, 32]. Essas técnicas utilizam  $\Psi$  e  $m^1\Psi$  para reduzir a resposta imune que degrada os primers, garantindo que a síntese proteica ocorra, além de proporcionar estabilidade e aumentar a eficiência desses protocolos. Vacinas que utilizam  $\Psi$  são mais eficientes que vacinas que utilizam ape-



**Figura 5.4**

Perfil de abertura da sequência de mRNA utilizada por Kim et al. [32]. Perfis foram calculados para a temperatura de 220K. Em vermelho, sequência sem modificação, em azul, sequência com modificação  $\Psi$  e em preto,  $m^1\Psi$ .

nas U, e vacinas que utilizam  $m^1\Psi$  são mais eficientes que as que utilizam  $\Psi$ , como reportado por Kim et al. [32]. Os parâmetros calculados aqui foram utilizados para obter perfis de abertura de uma sequência de mRNA utilizada por Kim et al. [32] para estudar as diferenças biológicas entre as sequências modificadas e não modificadas.

O perfil de abertura para o duplexo utilizado por Kim et al. [32] reforça a noção de que tanto a pseudouridina quanto a N1-metilpseudouridina atuam aumentando a estabilidade da dupla fita. Apesar disso, pouca diferença é notada entre as curvas de  $\Psi$  e  $m^1\Psi$ , ver Figura 5.4. Assim como para a inosina, descrito no capítulo 4, também foram calculadas as temperaturas de *melting* para um intervalo de três nucleotídeos ao redor da modificação, conforme mostrado na Tabela 5.3. Para isso, utilizamos os parâmetros de regressão obtidos para  $\Psi A$  no cálculo de  $T_m$  das sequências contendo  $m^1\Psi$ , uma vez que os parâmetros de regressão para  $m^1\Psi A$  foram determinados para concentrações de fita de  $5 \mu M$ . As temperaturas calculadas deixam evidente o ganho de estabilidade das modificações.

**Tabela 5.3**

Temperatura de *melting* predita  $T_{\text{pred.}}$  de sequências não modificadas e modificadas para uma hibridização de RNA de 7 pares de bases da sequência apresentada em Kim et al. [32]. As bases alteradas estão em negrito. Cálculo realizado para concentração da solução de 1000 M e concentração das fitas de 100  $\mu\text{M}$ .

Sequência 5' $\rightarrow$ 3' / 3' $\rightarrow$ 5'	$T_{\text{pred.}}$ ( $^{\circ}\text{C}$ )
AUGUACU/UACAUGA	38.3
AUGCACU/UACGUGA	48.4
AUG $\mathbf{m}^1\Psi$ ACU/UACAUGA	38.5
AUG $\Psi$ ACU/UACAUGA	40.9
AUG $\Psi$ ACU/UACUUGA	30.0
AUG $\Psi$ ACU/UACCUGA	32.1
AUG $\Psi$ ACU/UACGUGA	33.4

A redução na imunogenicidade proporcionada pela  $\Psi$  e  $\mathbf{m}^1\Psi$  ainda não é completamente compreendida. Os principais mecanismos conhecidos envolvem a família de receptores Receptor do tipo toll (*toll-like receptor*, TLR) [20, 80, 81], que atuam no reconhecimento e degradação de RNAs virais. O TLR3 atua no reconhecimento e degradação de duplexos ricos em uridina [80], enquanto os TLR7 e TLR8 atuam em mRNAs de fita simples [81, 82]. Além disso, sequências contendo  $\Psi$  são mais resistentes à degradação mediada por Ribonuclease L (RNase L) [83]. Outros mecanismos também estão associados ao reconhecimento e degradação de sequências contendo uridina.

## 5.6 Conclusão

Utilizamos o modelo PB e obtivemos parâmetros de interação para pseudouridina e N1-metilpseudouridina em RNA. Foram analisadas 41 sequências de RNA contendo pseudouridina ligada com as bases A, C, G e U [77, 78]. Observamos que a pseudouridina e a N1-metilpseudouridina apresentam uma forte contribuição para a estabilidade do duplexo quando pareadas com adenina. Os pares  $\Psi\text{A}$  e  $\mathbf{m}^1\Psi\text{A}$  são capazes de estabilizar o duplexo com contribuições maiores que o par canônico AU e de forma similar ao par GC. A análise de contexto para  $\Psi\text{A}$  apresentou valores equivalentes de Morse dos pares terminais e internos. Por outro lado, as interações entre os vizinhos apresentam grandes variações entre os dois conjuntos. Também observamos que nesses contextos a contribuição de  $\Psi\text{A}$  depende da posição relativa (5' ou 3') de seus vizinhos.  $\mathbf{m}^1\Psi\text{A}$ , por outro lado, apresentou pouca variação associada à posição de seus vizinhos. Os resultados obtidos aqui também demonstram que o par  $\Psi\text{G}$  contribui de forma significativa para a estabilidade do duplexo e sua contribuição é similar ao par GU. Poucos dados de temperatura de *melting* estavam disponíveis para os pares  $\Psi\text{C}$  e  $\Psi\text{U}$  levando a um resultado não conclusivo para estes pares, sendo necessário mais dados para uma análise completa destes pares. Os parâmetros de interação obtidos aqui podem ser utilizados para estudar

sequências que contenham os contextos utilizados neste estudo, permitindo o desenvolvimento e melhoramento de *primers* e técnicas terapêuticas baseadas em RNA.

## 6 Conclusão

O estudo de bases modificadas em RNA é um campo ainda em desenvolvimento. As modificações são diversas e possuem muitas funções associadas. A primeira modificação de RNA foi descoberta em 1956 [8] e sua descoberta levou à produção eficiente de vacinas de mRNA [1, 22, 32]. As vacinas de mRNA demonstram a versatilidade e alta eficiência das modificações para o uso em diferentes aplicações. *Primers* degenerados para RT-PCR de RNA modificado podem ser utilizados na detecção de variantes virais, sem perda de eficiência e sem a necessidade de produzir novos *primers* para cada variante emergente. Outras aplicações vêm sendo desenvolvidas, apesar disso, poucos estudos avaliando os impactos físicos, estruturais e de estabilidade térmica das modificações foram realizados.

Neste trabalho, estudamos três modificações de bases nitrogenadas em RNA e avaliamos suas contribuições para a estabilidade de duplexos à luz do modelo Peyrard-Bishop (PB) utilizando dados experimentais publicados [35, 38, 59, 77, 78]. O software TfReg [62, 63] foi utilizado para calcular parâmetros do modelo PB, potenciais de Morse e *stacking*, associados à cada modificação. Diversas rodadas de minimização foram realizadas para cada par, levando em conta as especificidades de cada conjunto de dados. Obtivemos parâmetros de estabilidade para o modelo PB de pares modificados com inosina (I), pseudouridina ( $\Psi$ ) e N1-metilpseudouridina ( $m^1\Psi$ ).

O modelo PB se mostrou útil para o estudo de modificações de RNA, desde que os dados  $T_m$  disponíveis sejam suficientes. Além disso, utilizamos e validamos protocolos de minimização que permitem estudar um conjunto de poucos dados. A separação de contexto utilizada se mostrou uma ferramenta essencial para obter resultados fisicamente válidos e para uma interpretação precisa dos parâmetros obtidos.

Os parâmetros obtidos para os pares IC e IU, demonstram a capacidade da inosina de estabilizar o duplexo e também corroboram conhecimentos biológicos. O fato de I ser, majoritariamente, interpretada como G durante a síntese proteica [14] pode ser confirmado comparando os potenciais de Morse obtidos para IU e IC. Nossos resultados também sugerem que as ligações de hidrogênio do par IU são consideravelmente fracas e similares à pareamentos com apenas uma ligação. Além disso, pares IU e IC em posições terminais apresentaram parâmetros de Morse e harmônicos elevados, independente da posição relativa 5' ou 3'. Os valores elevados para os parâmetros de Morse na posição terminal indicam possibilidades de aplicações para estabilização de *primers*. Os pares IG e II, por sua vez, apresentaram potenciais de Morse e harmônicos muito similares, isso se deve à similaridade de I com G. Esses pares, nos contextos estudados, atuam contribuindo para uma alta instabilidade do duplexo. Também comparamos nossos resultados com parâmetros de inosina em DNA de Maximiano and Weber [49]. Essa comparação demonstra uma grande diferença entre a ocorrência em DNA e RNA. A inosina em DNA apresenta a capacidade de ligar, de forma similar ao par AU, com todas as bases canônicas,

enquanto em RNA, ela desestabiliza o duplexo na maioria dos casos.

Os pares  $\Psi A$  e  $m^1\Psi A$  apresentaram potenciais de Morse muito similares, um pouco menor no caso de  $m^1\Psi A$ , em concordância com Dutta et al. [31]. Os valores de *stacking* desses pares também apresentaram tendências similares aos reportados por Dutta et al. [31] e Deb et al. [76]. Enquanto  $\Psi A$  apresenta valores de *stacking* dependentes da posição de seus vizinhos,  $m^1\Psi A$  remove de forma significativa essa dependência. Contextos terminais também foram estudados para  $\Psi A$ . Esse contexto também apresentou dependências da posição relativa dos pares vizinhos. O *mismatch*  $\Psi G$  apresentou uma contribuição significativa para estabilidade do duplexo tanto em termos de *stacking* quanto de Morse. Os resultados obtidos para  $\Psi C$  e  $\Psi U$  são inconclusivos, sendo necessário um conjunto maior de dados para um estudo completo desses pares.

Tanto para I quanto para  $\Psi$  foram produzidos gráficos de perfis de abertura de sequências de interesse. Os gráficos de I, no caso do gene RHOA, sugerem que sua presença permite que a sequência passe pelos mecanismos de degradação por miRNA, levando à produção de proteínas defeituosas [41]. Sendo esse o caso, seria possível o desenvolvimento de terapias gênicas que atuem degradando sequências nas quais a presença de inosina é prejudicial. Os perfis de abertura utilizados para estudar os códons apresentados por Licht et al. [23] são capazes de explicar algumas das taxas de tradução sinônima, mas falham em outros casos. Isso implica que apenas os perfis de abertura não são suficientes para um estudo completo da interação códon-anticódon.

Para a pseudouridina, comparamos os pares UA,  $\Psi A$  e  $m^1\Psi A$  em uma sequência utilizada em um estudo de eficiência das vacinas de mRNA [32]. Observamos uma grande influência na estrutura do duplexo com  $\Psi$  e  $m^1\Psi$  atuando de formas muito similares. U, por sua vez, apresenta uma pequena abertura do duplexo na região em que está presente. O perfil de abertura não ilustra diferenças significativas entre as modificações e o pareamento canônico. No entanto, as temperaturas preditas para cada caso demonstram as relações esperadas para essas sequências. Observamos que sequências com pares UA são menos estáveis que os pares  $\Psi A$  e  $m^1\Psi A$ . O par  $m^1\Psi A$ , por sua vez, apresenta uma  $T_m$  intermediária entre  $\Psi A$  e UA, justificando a capacidade de  $m^1\Psi$  produzir as proteínas esperadas.

As bases modificadas se tornam cada dia mais relevantes para aplicações tecnológicas e o entendimento de sistemas biológicos. Entender as contribuições dessas bases para a formação do duplexo e de suas estruturas é fundamental para o desenvolvimento de novas tecnologias e suas aplicações. Em nosso trabalho, obtivemos parâmetros de interação para bases modificadas em RNA. Encontramos resultados consistentes com outros trabalhos e que auxiliam o entendimento do comportamento dessas modificações. Os resultados obtidos aqui podem ser utilizados para predição de  $T_m$  e para avaliar os impactos estruturais dessas modificações em duplexos de RNA.

## 7 Perspectivas

O estudo de bases modificadas em RNA é necessário para a compreensão dos mecanismos essenciais à vida e para o avanço de futuras aplicações biotecnológicas. Os principais resultados deste trabalho podem ser aplicados em diversos contextos e combinados com outros modelos para investigar diferentes modificações, *mismatches* de bases canônicas e variados cenários biológicos. Atualmente, há um grande interesse em pesquisas que buscam prever os sítios de ocorrência de modificações em sequências de RNA biológicas. Os modelos PB e NN podem auxiliar na compreensão dos processos envolvidos na ocorrência dessas modificações. Futuramente, caso mais dados se tornem disponíveis, será possível otimizar e reavaliar os resultados obtidos aqui. Além disso, esperamos poder analisar outras modificações que possuam dados de  $T_m$  disponíveis.

Durante o desenvolvimento deste trabalho, ficou evidente a limitação do modelo PB quando se dispõe de poucos dados de temperaturas de *melting*. Isso exigiu a criação de protocolos de minimização capazes de lidar com a quantidade limitada de dados disponíveis. No futuro, pretendemos desenvolver uma integração entre os modelos PB e NN, permitindo análises mais aprofundadas de conjuntos com poucas temperaturas de *melting* e comparando e, possivelmente, otimizando os resultados obtidos neste estudo.

## REFERÊNCIAS

- [1] Tobias von der Haar, Thomas E Mulrone, Fabio Hedayioglu, Sathishkumar Kurusamy, Maria Rust, Kathryn S Lilley, James E Thaventhiran, Anne E Willis, and C Mark Smales. Translation of in vitro-transcribed rna therapeutics. *Frontiers in Molecular Biosciences*, 10:1128067, 2023.
- [2] Jerry L Wilson. Biochemistry; (stryer, lubert), 1988.
- [3] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [4] Rosalind E Franklin and Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.
- [5] David L Nelson and Michael M Cox. *Princípios de bioquímica de Lehninger*. Artmed Editora, 2022.
- [6] Lauren A Levine, Matthew Junker, Myranda Stark, and Dustin Greenleaf. A DNA melting exercise for a large laboratory class. *J. Chem. Educ.*, 92:1928–1931, 2015. doi:[10.1021/acs.jchemed.5b00049](https://doi.org/10.1021/acs.jchemed.5b00049).
- [7] John SantaLucia, Jr., H T Allawi, and P A Seneviratne. Improved nearest-neighbour parameters for predicting DNA duplex stability. *Biochem.*, 35:3555–3562, 1996. doi:[10.1021/bi951907q](https://doi.org/10.1021/bi951907q).
- [8] Frank F Davis and Frank Worthington Allen. Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem*, 227(2):907–15, 1957.
- [9] Waldo E Cohn. Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *Journal of Biological Chemistry*, 235(5): 1488–1498, 1960.
- [10] Hao Chen, Wenhao Luo, Xiaoyue Lu, and Taiping Zhang. Regulatory role of rna modifications in the treatment of pancreatic ductal adenocarcinoma (pdac). *Heliyon*, 9(11), 2023.
- [11] Christian Lorenz, Christina E Lünse, and Mario Mörl. trna modifications: impact on structure and thermal adaptation. *Biomolecules*, 7(2):35, 2017.
- [12] Wenbin Zhou, Daniel Karcher, and Ralph Bock. Importance of adenosine-to-inosine editing adjacent to the anticodon in an arabidopsis alanine trna under environmental stress. *Nucleic Acids Research*, 41(5):3362–3372, 2013.

- [13] Tomoko Hamma and Adrian R Ferré-D'Amaré. Pseudouridine synthases. *Chemistry & biology*, 13(11):1125–1135, 2006.
- [14] Boris Zinshteyn and Kazuko Nishikura. Adenosine-to-inosine rna editing. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(2):202–209, 2009.
- [15] Michael S Paul and Brenda L Bass. Inosine exists in mrna at tissue-specific levels and is most abundant in brain mrna. *The EMBO journal*, 17(4):1120–1127, 1998.
- [16] Adrian Gabriel Torres, David Piñeyro, Marta Rodríguez-Escribà, Noelia Camacho, Oscar Reina, Adélaïde Saint-Léger, Liudmila Filonava, Eduard Batlle, and Lluís Ribas de Pouplana. Inosine modifications in human trnas are incorporated at the precursor trna level. *Nucleic acids research*, 43(10):5145–5157, 2015.
- [17] Paz Nombela, Borja Miguel-López, and Sandra Blanco. The role of m6a, m5c and  $\psi$  rna modifications in cancer: Novel therapeutic opportunities. *Molecular cancer*, 20(1):1–30, 2021.
- [18] Xiaochen Xue, Zhiying Wang, Yafen Wang, and Xiang Zhou. Disease diagnosis based on nucleic acid modifications. *ACS Chemical Biology*, 2023.
- [19] Steve M Green, Paul R Lambden, Yu Deng, Ian N Clarke, J Andrew Lowes, Sarah Lineham, Julie Bushell, John Rogers, E Owen Caul, and Charles R Ashley. Polymerase chain reaction detection of small round-structured viruses from two related hospital outbreaks of gastroenteritis using inosine-containing primers. *Journal of medical virology*, 45(2):197–202, 1995.
- [20] Pedro Morais, Hironori Adachi, and Yi-Tao Yu. The critical contribution of pseudouridine to mrna covid-19 vaccines. *Frontiers in cell and developmental biology*, 9:3187, 2021.
- [21] Hadieh Monajemi and Sharifuddin M Zain. How stop codon pseudouridylation induces nonsense suppression. *Computational and Theoretical Chemistry*, 1204:113414, 2021.
- [22] Lyana LY Ho, Gabriel HA Schiess, Pâmella Miranda, Gerald Weber, and Kira Astakhova. Pseudouridine and n 1-methylpseudouridine as potent nucleotide analogues for rna therapy and vaccine development. *RSC Chemical Biology*, 2024.
- [23] Konstantin Licht, Markus Hartl, Fabian Amman, Dorothea Anrather, Michael P Janisiw, and Michael F Jantsch. Inosine induces context-dependent recoding and translational stalling. *Nucleic acids research*, 47(1):3–14, 2019. doi:[10.1093/nar/gky1163](https://doi.org/10.1093/nar/gky1163).
- [24] Maria Sofia Basile, Placido Bramanti, and Emanuela Mazzon. Inosine in neurodegenerative diseases: from the bench to the bedside. *Molecules*, 27(14):4644, 2022.

- [25] Dan Dominissini, Sharon Moshitch-Moshkovitz, Ninette Amariglio, and Gideon Rechavi. Adenosine-to-inosine rna editing meets cancer. *Carcinogenesis*, 32(11):1569–1577, 2011.
- [26] William Slotkin and Kazuko Nishikura. Adenosine-to-inosine rna editing and human disease. *Genome medicine*, 5(11):1–13, 2013.
- [27] Sundaramoorthy Srinivasan, Adrian Gabriel Torres, and Lluís Ribas de Pouplana. Inosine in biology and disease. *Genes*, 12(4):600, 2021.
- [28] Xiaoxue Yang, Jing Yang, Xiaozhen Liang, Qian Chen, Sijia Jiang, Haihua Liu, Yue Gao, Zhonglu Ren, Yi-Wu Shi, Sheng Li, et al. Landscape of dysregulated placental rna editing associated with preeclampsia. *Hypertension*, 75(6):1532–1541, 2020.
- [29] Michael Charette and Michael W Gray. Pseudouridine in rna: what, where, how, and why. *IUBMB life*, 49(5):341–352, 2000.
- [30] Marianna Penzo and Lorenzo Montanaro. Turning uridines around: role of rna pseudouridylation in ribosome biogenesis and ribosomal function. *Biomolecules*, 8(2):38, 2018.
- [31] Nivedita Dutta, Indrajit Deb, Joanna Sarzynska, and Ansuman Lahiri. Structural and thermodynamic consequences of base pairs containing pseudouridine and n1-methylpseudouridine in rna duplexes. *BioRxiv*, pages 2023–03, 2023.
- [32] Kyusik Q Kim, Bhagyashri D Burgute, Shin-Cheng Tzeng, Crystal Jing, Courtney Junegers, Junya Zhang, Liewei L Yan, Richard D Vierstra, Sergej Djuranovic, Bradley S Evans, et al. N1-methylpseudouridine found within covid-19 mrna vaccines produces faithful protein products. *Cell Reports*, 40(9), 2022.
- [33] Oliwia Andries, Séan Mc Cafferty, Stefaan C De Smedt, Ron Weiss, Niek N Sanders, and Tasuku Kitada. N1-methylpseudouridine-incorporated mrna outperforms pseudouridine-incorporated mrna by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *Journal of Controlled Release*, 217:337–344, 2015.
- [34] Brenda L. Bass. RNA editing by adenosine deaminases that act on RNA. *Annual Review of Biochemistry*, 71(1):817–846, 2002. doi:[10.1146/annurev.biochem.71.110601.135501](https://doi.org/10.1146/annurev.biochem.71.110601.135501). URL <https://doi.org/10.1146/annurev.biochem.71.110601.135501>. PMID: 12045112.
- [35] Daniel J Wright, Jamie L Rice, Dawn M Yanker, and Brent M Znosko. Nearest Neighbor Parameters for Inosine·Uridine Pairs in RNA Duplexes. *Biochem.*, 46(15):4625–4634, 2007. doi:[10.1021/bi0616910](https://doi.org/10.1021/bi0616910).
- [36] Douglas H Turner, Naoki Sugimoto, Ryszard Kierzek, and Scott D Dreiker. Free energy increments for hydrogen bonds in nucleic acid base pairs. *Journal of the American Chemical Society*, 109(12):3783–3785, 1987.

- [37] Niels Langkjær, Jesper Wengel, and Anna Pasternak. Watson–Crick hydrogen bonding of unlocked nucleic acids. *Bioorganic & Medicinal Chemistry Letters*, 25(22):5064–5066, 2015.
- [38] Daniel J Wright, Christopher R Force, and Brent M Znosko. Stability of RNA duplexes containing inosine·cytosine pairs. *Nucleic Acids Research*, 46(22):12099–12108, 10 2018. ISSN 0305-1048. doi:[10.1093/nar/gky907](https://doi.org/10.1093/nar/gky907).
- [39] Naďa Špačková and Kamila Réblová. Role of inosine–uracil base pairs in the canonical RNA duplexes. *Genes*, 9(7):324, June 2018. ISSN 2073-4425. doi:[10.3390/genes9070324](https://doi.org/10.3390/genes9070324). URL <http://dx.doi.org/10.3390/genes9070324>.
- [40] Martin J. Serra, Patricia E. Smolter, and Eric Westhof. Pronounced instability of tandem IU base pairs in RNA. *Nucleic Acids Research*, 32(5):1824–1828, 03 2004. ISSN 0305-1048. doi:[10.1093/nar/gkh501](https://doi.org/10.1093/nar/gkh501). URL <https://doi.org/10.1093/nar/gkh501>.
- [41] ADJ Scadden and Christopher WJ Smith. Rnai is antagonized by a? i hyper-editing. *EMBO reports*, 2001.
- [42] Cyril X George, Zhenji Gan, Yong Liu, and Charles E Samuel. Adenosine deaminases acting on rna, rna editing, and interferon action. *Journal of Interferon & Cytokine Research*, 31(1):99–117, 2011.
- [43] H Grosjean, F Constantinesco, D Foiret, and N Benachenhou. A novel enzymatic pathway leading to 1-methylinosine modification in haloferax volcanii trna. *Nucleic acids research*, 23(21):4312–4319, 1995.
- [44] Adrian Gabriel Torres, David Piñeyro, Liudmila Filonava, Travis H Stracker, Eduard Battlle, and Lluís Ribas de Pouplana. A-to-i editing on trnas: biochemical, biological and evolutionary implications. *FEBS letters*, 588(23):4279–4286, 2014.
- [45] Michael Hollmann, Melissa Hartley, and Stephen Heinemann. Ca<sup>2+</sup> permeability of kampa-gated glutamate receptor channels depends on subunit composition. *Science*, 252(5007):851–853, 1991.
- [46] Dirk Feldmeyer, Kalev Kask, Rossella Brusa, Hans-Christian Kornau, Rohini Kolhekar, Andrei Rozov, Nail Burnashev, Vidar Jensen, Øivind Hvalby, Rolf Sprengel, et al. Neurological dysfunctions in mice expressing different levels of the q/r site–unedited ampar subunit glur–b. *Nature neuroscience*, 2(1):57–64, 1999.
- [47] Ingrun Alseth, Bjørn Dalhus, and Magnar Bjørås. Inosine in dna and rna. *Current opinion in genetics & development*, 26:116–123, 2014.

- [48] Jr Watkins, Norman E. and Jr SantaLucia, John. Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.*, 33(19):6258–6267, 2005. doi:[10.1093/nar/gki918](https://doi.org/10.1093/nar/gki918). URL <http://nar.oxfordjournals.org/cgi/content/abstract/33/19/6258>.
- [49] Rodolfo Vieira Maximiano and Gerald Weber. Deoxyinosine mismatch parameters calculated with a mesoscopic model result in uniform hydrogen bonding and strongly variable stacking interactions. *Chem. Phys. Lett.*, 631–632:87–91, 2015. doi:[10.1016/j.cplett.2015.04.045](https://doi.org/10.1016/j.cplett.2015.04.045).
- [50] Francis H Martin, Miguel M Castro, Fareed Aboul-ela, and Ignacio Tinoco. Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.*, 13(24):8927–8938, 1985. doi:[10.1093/nar/13.24.8927](https://doi.org/10.1093/nar/13.24.8927).
- [51] Donald E Bergstrom, Peiming Zhang, and W Travis Johnson. Comparison of the base pairing properties of a series of nitroazole nucleobase analogs in the oligodeoxyribonucleotide sequence 5'-d(CGCXAATTYGCG)-3'. *Nucleic Acids Research*, 25(10):1935–1942, 1997.
- [52] E Ohtsuka, S Matsuki, M Ikehara, Y Takahashi, and K Matsubara. An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.*, 260(5):2605–2608, 1985. URL <http://www.jbc.org/cgi/content/abstract/260/5/2605>.
- [53] Yoosuke Takahashi, Kikuya Kato, Yoshihide Hayashizaki, Toshiaki Wakabayashi, Eiko Ohtsuka, Shigeru Matsuki, Morio Ikehara, and Kenichi Matsubara. Molecular cloning of the human cholecystokinin gene by use of a synthetic probe containing deoxyinosine. *Proceedings of the National Academy of Sciences*, 82(7):1931–1935, 1985.
- [54] Rajkumar V Patil and Eugene E Dekker. PCR amplification of an Escherichia coli gene using mixed primers containing deoxyinosine at ambiguous positions in degenerate amino acid codons. *Nucleic acids research*, 18(10):3080, 1990.
- [55] Airi Palva, Gabriele Vidgren, and Lars Paulin. Application of PCR with oligonucleotide primers containing deoxyinosine for gene detection, isolation and sequencing. *Journal of microbiological methods*, 19(4):315–321, 1994.
- [56] Scott A Strobel, Thomas R Cech, Nassim Usman, and Leonid Beigelman. The 2,6-diaimnopyrimidine riboside 5-methylisocytidine wobble base pair: An isoenergetic substitution for the study of GU pairs in RNA. *Biochemistry*, 33(46):13824–13835, 1994.
- [57] John SantaLucia Jr, Ryszard Kierzek, and Douglas H Turner. Functional group substitutions as probes of hydrogen bonding between GA mismatches in RNA internal loops. *Journal of the American Chemical Society*, 113(11):4313–4322, 1991.

- [58] Magdalena Broda, Elżbieta Kierzek, Zofia Gdaniec, Tadeusz Kulinski, and Ryszard Kierzek. Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry*, 44(32):10873–10882, 2005.
- [59] Mark E Burkard and Douglas H Turner. NMR structures of r (GCA G GC G UGC)<sub>2</sub> and determinants of stability for single guanosine-guanosine base pairs. *Biochemistry*, 39(38):11748–11762, 2000. doi:[10.1021/bi000720i](https://doi.org/10.1021/bi000720i).
- [60] Shun Sakuraba, Junichi Iwakiri, Michiaki Hamada, Tomoshi Kameda, Genichiro Tsuji, Yasuaki Kimura, Hiroshi Abe, and Kiyoshi Asai. Free-energy calculation of ribonucleic inosines and its application to nearest-neighbor parameters. *Journal of chemical theory and computation*, 16(9):5923–5935, 2020.
- [61] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62(23):2755–2757, 1989. doi:[10.1103/PhysRevLett.62.2755](https://doi.org/10.1103/PhysRevLett.62.2755).
- [62] Gerald Weber. TfReg: Calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models. *Bioinformatics*, 29:1345–1347, 2013. doi:[10.1093/bioinformatics/btt133](https://doi.org/10.1093/bioinformatics/btt133). URL <http://bioinformatics.oxfordjournals.org/content/29/10/1345>.
- [63] Gerald Weber. Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures. *Nucleic Acids Res.*, 41:e30, 2013. doi:[10.1093/nar/gks964](https://doi.org/10.1093/nar/gks964). URL <http://nar.oxfordjournals.org/content/41/1/e30>.
- [64] Philip N Borer, Barbara Dengler, Ignacio Tinoco Jr, and Olke C Uhlenbeck. Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.*, 86(4):843–853, 1974. doi:[10.1016/0022-2836\(74\)90357-X](https://doi.org/10.1016/0022-2836(74)90357-X).
- [65] 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](https://doi.org/10.1093/bioinformatics/btu751). URL <http://bioinformatics.oxfordjournals.org/content/31/6/871>.
- [66] John SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95(4):1460–1465, 1998. doi:[10.1073/pnas.95.4.1460](https://doi.org/10.1073/pnas.95.4.1460). URL <http://www.pnas.org/cgi/content/abstract/95/4/1460>.
- [67] K. J. Breslauer, R Frank, H Blocker, and L. A. Marky. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83(11):3746–3750, 1986. doi:[10.1073/pnas.83.11.3746](https://doi.org/10.1073/pnas.83.11.3746).

- [68] Tianbing Xia, John SantaLucia, Jr., Mark E. Burkard, Ryszard Kierzek, Susan J. Schroeder, Xiaoqi Jiao, Christopher Cox, and Douglas H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochem.*, 37:14719–14735, 1998. doi:[10.1021/bi9809425](https://doi.org/10.1021/bi9809425).
- [69] Philip M Morse. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical Review*, 34(1):57, 1929.
- [70] Gerald Weber, Niall Haslam, Nava Whiteford, Adam Prügel-Bennett, Jonathan W. Essex, and Cameron Neylon. Thermal equivalence of DNA duplexes without melting temperature calculation. *Nat. Phys.*, 2:55–59, 2006. doi:[10.1038/nphys189](https://doi.org/10.1038/nphys189).
- [71] John A Nelder and Roger Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.
- [72] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.
- [73] Tauanne D. Amarante and Gerald Weber. Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model. *J. Chem. Inf. Model.*, 56(1):101–109, 2016. doi:[10.1021/acs.jcim.5b00571](https://doi.org/10.1021/acs.jcim.5b00571). URL <http://dx.doi.org/10.1021/acs.jcim.5b00571>.
- [74] Xiaoxue Yang, Jing Yang, Xiaozhen Liang, Qian Chen, Sijia Jiang, Haihua Liu, Yue Gao, Zhonglu Ren, Yi-Wu Shi, Sheng Li, et al. Landscape of dysregulated placental rna editing associated with preeclampsia. *Hypertension*, 75(6):1532–1541, 2020.
- [75] Marco Antonio Valencia-Sanchez, Jidong Liu, Gregory J Hannon, and Roy Parker. Control of translation and mrna degradation by mirnas and sirnas. *Genes & development*, 20(5): 515–524, 2006.
- [76] Indrajit Deb, Łukasz Popenda, Joanna Sarzyńska, Magdalena Małgowska, Ansuman Lahiri, Zofia Gdaniec, and Ryszard Kierzek. Computational and nmr studies of rna duplexes with an internal pseudouridine-adenosine base pair. *Scientific Reports*, 9(1):16278, 2019.
- [77] Elzbieta Kierzek, Magdalena Malgowska, Jolanta Lisowiec, Douglas H Turner, Zofia Gdaniec, and Ryszard Kierzek. The contribution of pseudouridine to stabilities and structure of rnas. *Nucleic acids research*, 42(5):3492–3501, 2014.
- [78] Graham A Hudson, Richard J Bloomingdale, and Brent M Znosko. Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *Rna*, 19(11):1474–1482, 2013.

- [79] Darrell R Davis. Stabilization of rna stacking by pseudouridine. *Nucleic acids research*, 23(24):5020–5026, 1995.
- [80] Katalin Karikó, Michael Buckstein, Houping Ni, and Drew Weissman. Suppression of RNA recognition by toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, 23(2):165–175, aug 2005. doi:[10.1016/j.immuni.2005.06.008](https://doi.org/10.1016/j.immuni.2005.06.008). URL <https://doi.org/10.1016%2Fj.immuni.2005.06.008>.
- [81] Sandra S Diebold, Catherine Massacrier, Shizuo Akira, Carine Paturel, Yannis Morel, and Caetano Reis e Sousa. Nucleic acid agonists for toll-like receptor 7 are defined by the presence of uridine ribonucleotides. *European journal of immunology*, 36(12):3256–3267, 2006.
- [82] Florian Heil, Hiroaki Hemmi, Hubertus Hochrein, Franziska Ampenberger, Carsten Kirschning, Shizuo Akira, Grayson Lipford, Hermann Wagner, and Stefan Bauer. Species-specific recognition of single-stranded rna via toll-like receptor 7 and 8. *Science*, 303(5663):1526–1529, 2004.
- [83] Bart R Anderson, Hiromi Muramatsu, Babal K Jha, Robert H Silverman, Drew Weissman, and Katalin Karikó. Nucleoside modifications in rna limit activation of 2'-5'-oligoadenylate synthetase and increase resistance to cleavage by rnase l. *Nucleic acids research*, 39(21):9329–9338, 2011.

## APÊNDICE A - Dados utilizados e resultados obtidos para inosina em RNA

**Tabela A.1**

Sequências de RNA contendo modificações de inosina (I).  $T_{\text{ref.}}$  e  $T_{\text{pred.}}$  são, respectivamente, as temperaturas experimentais e preditas.

Sequências 5' → 3' / 3' → 5'	$T_{\text{ref.}}$ (°C)	$T_{\text{pred.}}$ (°C)
Wright et al. [35]		
GCAICGC/CGUUGCG	45.2	44.7
GCAIGGC/CGUUCCG	48.7	49.2
GCCIAGC/CGGUUCG	47.1	46.6
GCCICGC/CGGUGCG	54.7	55.2
GCCIGGC/CGGUCCG	59.1	58.7
GCCIUGC/CGGUACG	44.6	45.0
GCGICGC/CGCUGCG	48.0	49.2
GCGIAGC/CGCUUCG	43.6	42.4
GCGIGGC/CGCUCCG	59.2	59.1
GCGIUGC/CGCUACG	48.6	48.7
GCUIAGC/CGAUUCG	34.6	37.0
GCUICGC/CGAUGCG	45.1	43.9
GCUIGGC/CGAUCCG	48.5	48.4
GCUIUGC/CGAUACG	37.3	36.2
GCGCAI/CGCGUU	55.8	55.9
GCGCCI/CGCGGU	56.6	56.7
GCGCGI/CGCGCU	52.2	52.3
GCGCUI/CGCGAU	47.9	47.9
IAGCGC/UUCGCG	47.5	47.5
ICGCGC/UGC GCG	55.8	55.3
IGGCGC/UCCGCG	55.8	55.9

(continua na próxima página)

(Tabela A.1 continuação)

Sequências 5' → 3'/3' → 5'	$T_{\text{ref.}}$ (°C)	$T_{\text{pred.}}$ (°C)
IUGC/GC/UACGCG	57.6	57.5
ICCGGU/UGGCC	53.5	53.9
GGCIUGCC/CCGUICGG	42.4	43.9
GGCUIGCC/CCGIUCGG	47.9	47.6
GGCIIIGGC/CCGUUCCG	46.1	45.1
GUCUIGAC/CAGIUCUG	34.9	34.7
Wright et al. [38]		
GCAICGC/CGUCGCG	55.3	55.6
GCAIGGC/CGUCCCG	59.4	59.3
GCAIUGC/CGUCACG	47.5	47.4
GCCIAGC/CGGCUCG	56.9	57.4
GCCICGC/CGGCGCG	64.8	64.9
GCCIGGC/CGGCCCG	67.3	68.2
GCCIUGC/CGGCACG	57.1	56.9
GCGIAGC/CGCCUCG	57.0	55.8
GCGICGC/CGCCGCG	60.7	63.0
GCGIGGC/CGCCCCG	66.4	66.8
GCGIUGC/CGCCACG	55.8	55.3
GCUIAGC/CGACUCG	46.8	47.4
GCUICGC/CGACGCG	56.9	55.3
GCUIGGC/CGACCCG	58.4	58.7
GCUIUGC/CGACACG	46.2	46.8
GCGCAI/CGCGUC	56.2	56.2
GCGCCI/CGCGGC	57.5	57.9
GCGCGI/CGCGCC	54.8	54.9
GCGCUI/CGCGAC	46.4	46.6
IAGCGC/CUCGCG	47.6	47.7
ICGCGC/CGCGCG	55.4	54.3

(continua na próxima página)

(Tabela A.1 continuação)

Sequências 5' → 3'/3' → 5'	$T_{\text{ref.}}$ (°C)	$T_{\text{pred.}}$ (°C)
<b>IGGCGC/CCCGCG</b>	54.5	54.6
<b>IUGCGC/CACGCG</b>	56.7	56.7
<b>ICCGGC/CGGCC</b>	51.3	51.9
<b>CGGCC/ICCGGC</b>	54.8	54.7
<b>GUCCIGAC/CAGICCU</b>	51.4	51.6
<b>GGCICGCC/CCGCICGG</b>	65.9	65.7
<b>GGCCIGCC/CCGICCGG</b>	67.7	67.2
<b>GGCIIIGGC/CCGCCCCG</b>	68.6	68.3
Burkard and Turner [59]		
<b>GUGGCAG/CACIGUC</b>	34.6	35.8
<b>GUGICAG/CACGGUC</b>	35.1	35.8
<b>CGGCAUG GCIGUAC</b>	31.3	31.4
<b>CGICAUG/GCGGUAC</b>	31.7	31.4
<b>GGCUGAG/CIGACUC</b>	37.7	36.4
<b>GICUGAG/CGGACUC</b>	36.7	36.4
<b>GCAGGCIUGC/CGUICGGACG</b>	44.6	44.6
<b>GAGGAUICUC/CUCIUAGGAG</b>	36.4	36.4
<b>GAGIAUGCUC/CUCGUAIGAG</b>	36.0	36.0
<b>GUGICAG/CACIGUC</b>	29.4	31.9
<b>CGICAUG/GCIGUAC</b>	26.8	26.9
<b>GICUGAG/CIGACUC</b>	35.5	32.7
<b>GCGICGC/CGCIGCG</b>	47.7	47.8
<b>CGCIGCG/GCGICGC</b>	40.3	40.4

**Tabela A.2**

Sequências canônicas de referência para os pares IC, IG e IU, incluindo o par IC tandem (as últimas cinco sequências).  $T_{\text{pred.}}$  é a temperatura predita.

Sequências 5' → 3' / 3' → 5'	$T_{\text{pred.}}$ (°C)
Sequências de referência para os grupos IC e IU	
GCAACGC/CGUUGCG	56.7
GCAAGGC/CGUUCCG	59.8
GCCAAGC/CGGUUCG	59.8
GCCACGC/CGGUGCG	66.3
GCCAGGC/CGGUCCG	69.1
GCCAUGC/CGGUACG	59.9
GCGACGC/CGCUGCG	63.4
GCGAAGC/CGCUUCG	56.8
GCGAGGC/CGCUCCG	66.3
GCGAUGC/CGCUACG	57.0
GCUAAGC/CGAUUCG	51.1
GCUACGC/CGAUGCG	58.4
GCUAGGC/CGAUCCG	61.5
GCUAUGC/CGAUACG	51.5
GCGCAA/CGCGUU	44.5
GCGCCA/CGCGGU	57.4
GCGCGA/CGCGCU	53.8
GCGCUA/CGCGAU	46.8
AAGCGC/UUCGCG	44.3
ACGCGC/UGCGCG	53.4
AGGCGC/UCCGCG	57.2
AUGCGC/UACGCG	45.2
ACCGGU/UGGCCA	48.9
GCAGCGC/CGUCGCG	66.1

(continua na próxima página)

(Tabela A.2 continuação)

Sequências 5' → 3' / 3' → 5'	$T_{\text{pred.}} (^{\circ}\text{C})$
GCAGGGC/CGUCCCG	69.1
GCAGUGC/CGUCACG	60.1
GCCGAGC/CGGCUCG	66.3
GCCGCGC/CGGCGCG	72.0
GCCGGGC/CGGCCCG	74.8
GCCGUGC/CGGCACG	66.3
GCGGAGC/CGCCUCG	66.3
GCGGCGC/CGCCGCG	72.0
GCGGGGC/CGCCCCG	74.8
GCGGUGC/CGCCACG	66.3
GCUGAGC/CGACUCG	60.2
GCUGCGC/CGACGCG	66.1
GCUGGGC/CGACCCG	69.1
GCUGUGC/CGACACG	60.1
GCGCAG/CGCGUC	52.8
GCGCCG/CGCGGC	60.6
GCGCUG/CGCGAC	52.8
GAGCGC/CUCGCG	53.8
GCGCGC/CGCGCG	63.1
GGGCGC/CCCGCG	64.1
GUGCGC/CACGCG	53.6
GCCGGC/CGGCCG	64.1
CGGCCG/GCCGGC	59.6
Sequências de referência para o par IC tandem	
GUCCGGAC/CAGGCCUG	67.7
GGCGCGCC/CCGCGCGG	80.2
GGCCGGCC/CCGGCCGG	83.0
GGCGGGGC/CCGCCCG	83.0

(continua na próxima página)

(Tabela A.2 continuação)

Sequências 5' → 3' / 3' → 5'	$T_{\text{pred.}}$ (°C)
GCAGCUGC/CGUCGACG	71.2
Sequências de referência para o par IU tandem	
GGCAUGCC/CCGUACGG	71.0
GGCUAGCC/CCGAUCGG	71.6
GGCAAGGC/CCGUUCCG	67.3
GUCUAGAC/CAGAUCUG	54
Sequências de referência para o par IG	
GUGGCAG/CACCGUC	60.2
GUGCCAG/CACGGUC	60.2
CGGCAUG/GCCGUAC	56.3
CGCCAUG/GCGGUAC	56.3
GGCUGAG/CCGACUC	60.1
GCCUGAG/CGGACUC	60.1
GCAGGCCUGC/CGUCCGGACG	85.9
GAGGAUCCUC/CUCCUAGGAG	71.9
GAGCAUGCUC/CUCGUACGAG	72.4
Sequências de referência para o par II	
GUGGCAG/CACCGUC	60.3
CGGCAUG/GCCGUAC	56.3
GGCUGAG/CCGACUC	60.1
GCGGCGC/CGCCGCG	71.8
CGCGGCG/GCGCCGC	68.6
GUGACAG/CACUGUC	49.3
CGACAUG/GCUGUAC	45.4
GACUGAG/CUGACUC	49.6
GCGACGC/CGCUGCG	62.7
CGCAGCG/GCGUCGC	62.6

**Tabela A.3**

Regras de contexto usadas para os pares IU e IC. As sequências são agrupadas por trímeros de inosina e as regras são aplicadas aos trímeros especificados.

Pares	Regras
IU <sup>A</sup>	AIC/UUG, AIG/UUC, AUC/UIG, CIC/GUG, CIG/GUC, CUC/-GIG
IU*	UU*/AI*, UI*/AU*, GU*/CI*, GI*/CU*, CU*/GI*, CI*/GU*, AI*/UU*, AU*/UI*
IU <sup>B</sup>	AUA/UIU, AUG/UIC, CIA/GUU, CUA/GIU, GIA/CUU, GIC/-CUG, GUA/CIU, UIA/AUU
IU <sup>†</sup>	CCU/GGI, CII/GUU, CIU/GUI, CUI/GIU, CUU/GII
IC <sup>I</sup>	ACA/UIU, ACC/UIG, ACG/UIC, ACU/UIA, AIC/UCG, AIG/UCC, CCA/GIU, CCC/GIG, CCG/GIC, CGC/GCG, CIA/GCU, CIC/GCG, GCA/CIU, GCC/CIG, GIA/CCU, UCA/AIU
IC*	UI*/AC*, UC*/AI*, GI*/CC*, GC*/CI*, CI*/GC*, CC*/GI*, AI*/UC*, AC*/UI*
IC <sup>†</sup>	AIC/UCI, CCC/GII, CCI/GIC, CIC/GCI, CCG/IIC

**Tabela A.4**

Os valores preditos do potencial de Morse ( $D$ ), que representa as ligações de hidrogênio, para cada par.

Parâmetro ( $D$ )	(meV)
IC	25.2(7)
IC*	59.4(7)
IC <sup>†</sup>	21.9(3)
IU <sup>A</sup>	5.8(6)
IU <sup>B</sup>	3.5(5)
IU*	44(2)
IU <sup>†</sup>	3.5(5)
II	$5.6(2) \cdot 10^{-2}$

**Tabela A.5**

Os valores preditos do potencial harmônico ( $k$ ), que representa a interação de empilhamento.

Parâmetro ( $k$ )	(eV/nm <sup>2</sup> )	Parâmetro ( $k$ )	(eV/nm <sup>2</sup> )
GCpUI <sup>A</sup>	2.0(1)	GCpIU <sup>A</sup>	4.0(2)
AUpIU <sup>A</sup>	0.24(3)	AUpUI <sup>A</sup>	0.64(8)
CGpUI <sup>A</sup>	2.6(1)	CGpIU <sup>A</sup>	$1.9(1) \cdot 10^{-1}$
AUpIU*	9.4(7)	AUpUI*	14(3)
CGpIU*	2.0(1)	CGpUI*	1.8(1)
GCpIU*	2.2(2)	GCpUI*	3.3(2)
IU*pAU	2.9(2)	UApIU*	3.1(2)
AUpUI <sup>B</sup>	0.036(4)	CGpUI <sup>B</sup>	1.9(1)
CGpIU <sup>B</sup>	2.24(7)	GCpIU <sup>B</sup>	1.76(9)
GCpUI <sup>B</sup>	1.33(9)	IU <sup>B</sup> pAU	0.57(4)
UApIU <sup>B</sup>	0.34(4)	IU <sup>†</sup> pIU <sup>†</sup>	0.04(2)
IU <sup>†</sup> pUI <sup>†</sup>	$0.1(0) \cdot 10^{-3}$	UI <sup>†</sup> pIU <sup>†</sup>	1.63
AUpIC	2.4(1)	AUpCI	1.19(5)
CGpCI	3.30(7)	CGpIC	2.94(9)
CIpAU	1.85(8)	CIpCG	4.7(1)
GCpCI	3.74(8)	ICpAU	1.43(6)
AUpCI*	6.3(2)	AUpIC*	5.9(2)
CGpCI*	0.81(2)	CGpIC*	$1.32(3) \cdot 10^{-1}$
CI*pAU	1.21(4)	CI*pCG	1.80(5)
GCpCI*	1.69(4)	IC*pAU	1.55(5)
AUpIC <sup>†</sup>	2.4(1)	CGpCI <sup>†</sup>	3.29(7)
CGpIC <sup>†</sup>	2.94(9)	CI <sup>†</sup> pCI <sup>†</sup>	5.3(2)
CI <sup>†</sup> pIC <sup>†</sup>	2.7(1)	IC <sup>†</sup> pCI <sup>†</sup>	2.05(9)
AUpGI	0.55(5)	CGpIG	0.0009(1)
GCpGI	0.91(9)	GCpIG	0.88(7)
GIpAU	2.0(2)	IGpAU	1.9(2)
CGpII	0.05(2)	GCpII	1.01

**Tabela A.6**Valores iniciais usados como parâmetros para os contextos  $IU^A$  e  $IU^B$ .

$k$	(eV/nm <sup>2</sup> )
AUpIU <sup>A</sup>	0.4
AUpUI <sup>A</sup>	2.5
GCpUI <sup>A</sup>	1.4
GCpIU <sup>A</sup>	3.5
CGpUI <sup>A</sup>	2.7
CGpIU <sup>A</sup>	2.2
AUpUI <sup>B</sup>	0.1
CGpUI <sup>B</sup>	2.5
CGpIU <sup>B</sup>	2.2
GCpIU <sup>B</sup>	3.5
GCpUI <sup>B</sup>	1.4
IU <sup>B</sup> pAU	$0.4 \cdot 10^{-1}$
UApIU <sup>B</sup>	0.2
Morse	(meV)
IU <sup>A</sup>	38.7
IU <sup>B</sup>	2.4

## APÊNDICE B - Dados utilizados e resultados obtidos para pseudouridina em RNA

**Tabela B.1**

$T_m$  preditas para sequências de RNA contendo modificações de pseudouridina  $\Psi A$ ,  $m^1\Psi$ ,  $\Psi G$ ,  $\Psi U$  and  $\Psi C$ .  $T_{ref.}$  e  $T_{pred.}$  são as temperaturas experimentais e preditas, respectivamente.

Sequências 5' → 3' / 3' → 5'	$T_{ref.}$ (°C)	$T_{pred.}$ (°C)
Hudson et al. [78]		
CGA $\Psi$ ACG/GCUAUGC	52.4	50.5
CGC $\Psi$ ACG/GCGAUGC	58.5	60.1
CGG $\Psi$ ACG/GCCAUGC	60.7	57.9
CGU $\Psi$ ACG/GCAAUGC	47.6	46.4
CGA $\Psi$ CCG/GCUAGGC	59.5	55.8
CGC $\Psi$ CCG/GCGAGGC	67.4	65.1
CGG $\Psi$ CCG/GCCAGGC	59.1	62.8
CGU $\Psi$ CCG/GCAAGGC	49.3	51.7
CGA $\Psi$ GCG/GCUACGC	55.7	56.3
CGC $\Psi$ GCG/GCGACGC	65.9	65.6
CGG $\Psi$ GCG/GCCACGC	61.8	63.1
CGU $\Psi$ GCG/GCAACGC	55.1	52.2
CGA $\Psi$ UCG/GCUAAGC	50.2	49.2
CGG $\Psi$ UCG/GCCAAGC	58.1	56.6
CGU $\Psi$ UCG/GCAAAGC	48	45.0
GCGCA $\Psi$ /CGCGUA	53.9	54.5
GCGCC $\Psi$ /CGCGGA	54.5	55.1
GCGCG $\Psi$ /CGCGCA	55.1	55.9
GCGCU $\Psi$ /CGCGAA	46.1	47.6
$\Psi$ AGCGC/AUCGCG	47.9	49.2

(continua na próxima página)

(Tabela B.1 continuação)

Sequências 5' → 3'/3' → 5'	$T_{\text{ref.}} (^{\circ}\text{C})$	$T_{\text{pred.}} (^{\circ}\text{C})$
ΨCGCGC/AGCGCG	52.5	54.3
ΨGGCGC/ACCGCG	55.3	56.0
ΨUGCGC/AACGCG	53	53.8
Kierzek et al. [77]		
UCAGΨCAGU/AGUCAGUCA	61.4	60.0
UCACΨGAGU/AGUGACUCA	62.4	64.1
UCAAΨUAGU/AGUUAUUA	45	47.5
UCAUΨAAGU/AGUAAUUA	45.1	44.6
ΨCAGUCAGU/AGUCAGUCA	59.9	57.6
UCAGUCAGΨ/AGUCAGUCA	59.7	59.4
UCAGΨCAGU/AGUCCGUCA	45.6	45.6
ΨCAGUCAGU/CGUCAGUCA	57.2	57.3
UCAGUCAGΨ/AGUCAGUCC	57.4	57.4
UCAGΨCAGU/AGUCGGUCA	61.8	60.2
UCACΨGAGU/AGUGGCUCA	58.1	58.1
UCAAΨUAGU/AGUUGAUA	42.7	42.3
UCAUΨAAGU/AGUAGUUA	40.5	40.9
ΨCAGUCAGU/GGUCAGUCA	58.2	58.9
UCAGUCAGΨ/AGUCAGUCG	60.7	61.6
UCAGΨCAGU/AGUCUGUCA	49.9	49.9
ΨCAGUCAGU/UGUCAGUCA	57.5	57.5
UCAGUCAGΨ/AGUCAGUCU	57.5	57.6
Sequências de $m^1\Psi$		
GCA( $m^1\Psi$ )GAG/CGUACUC	40	42.8
ACG( $m^1\Psi$ )ACG/UGCAUGC	40	40.1
CGU( $m^1\Psi$ )CAG/GCAAGUC	34.3	35.3
GAC( $m^1\Psi$ )UCG/CUGAAGC	37.5	39.3
CGA( $m^1\Psi$ )GUC/GCUACAG	42.5	39.8

(continua na próxima página)

(Tabela B.1 continuação)

Sequências 5' → 3'/3' → 5'	$T_{\text{ref.}}$ (°C)	$T_{\text{pred.}}$ (°C)
CAG(m <sup>1</sup> Ψ)AGC/GUCAUCG	44.1	44.2
GCU(m <sup>1</sup> Ψ)CGA/CGAAGCU	42.5	41.5
AGC(m <sup>1</sup> Ψ)UCG/UCGAAGC	43.8	41.9
CGA(m <sup>1</sup> Ψ)ACG/GCUAUGC	39.8	39.5

**Tabela B.2**

Sequências canônicas de referência para pares  $\Psi$ A,  $m^1\Psi$ ,  $\Psi$ G,  $\Psi$ U e  $\Psi$ C.  $T_{\text{pred.}}$  é a temperatura predita.

Sequências 5' $\rightarrow$ 3' / 3' $\rightarrow$ 5'	$T_{\text{pred.}}$ (°C)
Sequências de referência para o par $\Psi$	
CGAUACG/GCUAUGC	43.2
CGCUACG/GCGAUGC	53.9
CGGUACG/GCCAUGC	54.0
CGUUACG/GCAAUGC	42.0
CGAUCCG/GCUAGGC	52.9
CGCUCCG/GCGAGGC	62.5
CGGUCCG/GCCAGGC	62.7
CGUUCG/GCAAGGC	52.3
CGAUGC/GCUACGC	52.6
CGCUGC/GCGACGC	62.3
CGGUGC/GCCACGC	62.5
CGUUGC/GCAACGC	52.0
CGAUUCG/GCUAAGC	41.6
CGGUUCG/GCCAAGC	52.3
CGUUUCG/GCAAAGC	40.3
GCGCAU/CGCGUA	45.1
GCGCCU/CGCGGA	57.2
GCGCGU/CGCGCA	53.4
GCGCUU/CGCGAA	44.3
UAGCGC/AUCGCG	46.8
UCGCGC/AGCGCG	53.8
UGGCGC/ACCGCG	57.4
UUGCGC/AACGCG	44.5
Sequências com resultados experimentais de Kierzek et al. [77]	

(continua na próxima página)

(Tabela B.2 continuação)

Sequências 5' → 3' / 3' → 5'	$T_{\text{pred.}}$ (°C)
UCAGUCAGU/AGUCAGUCA	59.1
UCACUGAGU/AGUGACUCA	58.5
UCAAUUAGU/AGUUAUUA	43.2
UCAUUAAGU/AGUAAUUA	42.3
UCAGUCAGU/AGUCAGUCA	59.1
UCAGUCAGU/AGUCAGUCA	59.1
Sequências de referência para os pares $\Psi\text{C}$ , $\Psi\text{G}$ e $\Psi\text{U}$	
UCAGCCAGU/AGUCGGUCA	70.6
UCACCGAGU/AGUGGCUCA	68.3
UCAACUAGU/AGUUGAUA	54.9
UCAUCAAGU/AGUAGUUA	53.8
UCAGACAGU/AGUCUGUCA	62.8
UCAGGCAGU/AGUCCGUCA	70.6
CCAGUCAGU/GGUCAGUCA	67.2
UCAGUCAGC/AGUCAGUCG	67.5
ACAGUCAGU/UGUCAGUCA	62.7
UCAGUCAGA/AGUCAGUCU	63.0
GCAGUCAGU/CGUCAGUCA	67.5
UCAGUCAGG/AGUCAGUCC	67.3
Sequências de referência para os pares $m^1\Psi$	
GUGGCAG/CACCGUC	60.2
GUGCCAG/CACGGUC	60.2
CGGCAUG/GCCGUAC	56.3
CGCCAUG/GCGGUAC	56.3
GGCUGAG/CCGACUC	60.1
GCCUGAG/CGGACUC	60.1
GCAGGCCUGC/CGUCCGGACG	85.9
GAGGAUCCUC/CUCCUAGGAG	71.9

(continua na próxima página)

(Tabela B.2 continuação)

Sequências 5' → 3' / 3' → 5'	$T_{\text{pred.}}$ (°C)
GAGCAUGCUC/CUCGUACGAG	72.4

**Tabela B.3**

Regras de contexto utilizadas para os pares  $\Psi^I$  e  $\Psi^*$ . As sequências são agrupadas por trímeros de pseudouridina e as regras são aplicadas aos trímeros especificados.

Pares	Regras
$A\Psi^*$	AA*/U $\Psi^*$ , A $\Psi^*$ /UA*, C $\Psi^*$ /GA*, GA*/C $\Psi^*$ , G $\Psi^*$ /CA*, UA*/A $\Psi^*$ , U $\Psi^*$ /AA*
$A\Psi^I$	AAA/U $\Psi^I$ U, AAC/U $\Psi^I$ G, AAU/U $\Psi^I$ A, A $\Psi^I$ A/UAU, A $\Psi^I$ C/UAG, A $\Psi^I$ G/UAC, CAA/G $\Psi^I$ U, CAC/G $\Psi^I$ G, CAG/G $\Psi^I$ C, C $\Psi^I$ A/GAU, C $\Psi^I$ C/GAG, GAA/C $\Psi^I$ U, GAC/C $\Psi^I$ G, G $\Psi^I$ A/CAU, UAA/A $\Psi^I$ U

**Tabela B.4**

Os valores preditos do potencial de Morse ( $D$ ), que representa as ligações de hidrogênio, para cada par. Mostrados estão pares ( $I$ ) internos e ( $T$ ) terminais.

Parâmetro	(meV)
$m^1\Psi A$	4.9(1)
$\Psi^I A$	5.1(1)
$\Psi^* A$	5.1(2)
$\Psi G$	27.9 (1)
$\Psi C$	11.3(2)
$\Psi U$	8.6(2)

**Tabela B.5**

Os valores preditos do potencial harmônico ( $k$ ), que representa a interação de empilhamento. Mostrados estão pares ( $I$ ) internos e ( $T$ ) terminais.

Parâmetro ( $k$ )	(eV/nm <sup>2</sup> )	Parâmetro ( $k$ )	(eV/nm <sup>2</sup> )
Am <sup>1</sup> ΨpAU	2.6(1)	Am <sup>1</sup> ΨpCG	2.3(2)
Am <sup>1</sup> ΨpGC	2.2(1)	Am <sup>1</sup> ΨpUA	3.2(2)
AUpAm <sup>1</sup> Ψ	2.4(2)	CGpAm <sup>1</sup> Ψ	1.9(2)
GCpAm <sup>1</sup> Ψ	1.9(1)	m <sup>1</sup> ΨApAU	4.4(2)
AUpAΨ <sup>I</sup>	1.6(1)	AUpAΨ*	4.2(3)
AUpΨ <sup>I</sup> A	6.6(1)	AUpΨ*A	4.7(3)
AΨ*pAU	1.3(1)	AΨ*pCG	2.1(1)
AΨ*pGC	0.85(6)	AΨ <sup>I</sup> pAU	4.1(1)
AΨ <sup>I</sup> pCG	4.2(1)	AΨ <sup>I</sup> pGC	5.1(1)
CGpAΨ <sup>I</sup>	0.98(4)	CGpAΨ*	2.7(3)
GCpAΨ <sup>I</sup>	1.1(1)	GCpAΨ*	1.7(1)
UApAΨ <sup>I</sup>	2.5(1)	UApAΨ*	1.8(1)
GCpUΨ	0.22(2)	GCpΨU	0.20(2)
CΨpCG	5.3(4) · 10 <sup>-1</sup>	GCpCΨ	6.0(6) · 10 <sup>-1</sup>
AUpGΨ	2.6(3)	AUpΨG	2.6(1)
CGpGΨ	2.2(1)	CGpΨG	2.7(1)
GCpGΨ	3.2(3)	GCpΨG	4.8(3)
GΨpAU	2.1(4)	UApGΨ	2.1(4)

**Tabela B.6**

Valores de Morse  $D$  e Harmônico  $k$  obtidos para o par  $\Psi A$  sem regras de contexto.

Harmônico	(eV/nm <sup>2</sup> )
AUpA $\Psi$	8.6(3)
AUp $\Psi A$	10.5(3)
A $\Psi$ pAU	5.2(2)
A $\Psi$ pCG	6.4(2)
A $\Psi$ pGC	7.55(4)
CGpA $\Psi$	4.1(2)
GCpA $\Psi$	4.9(2)
UApA $\Psi$	8.7(2)
Morse (meV)	19.8(2)

**Tabela B.7**Valores de Morse  $D$  e Harmônico  $k$  obtidos para o par  $m^1\Psi A$  em concentração das fitas de  $10 \mu M$ 

$k$	(eV/nm <sup>2</sup> )	$k$	(eV/nm <sup>2</sup> )
Am <sup>1</sup> ΨpAU	2.6(1)	Am <sup>1</sup> ΨpCG	2.3(2)
Am <sup>1</sup> ΨpGC	2.2(1)	Am <sup>1</sup> ΨpUA	3.2(2)
AUpAm <sup>1</sup> Ψ	2.4(2)	CGpAm <sup>1</sup> Ψ	1.9(2)
GCpAm <sup>1</sup> Ψ	1.9(1)	m <sup>1</sup> ΨApAU	4.4(2)
AUpAΨ <sup>I</sup>	1.6(1)	AUpAΨ*	4.2(3)
AUpΨ <sup>I</sup> A	6.6(1)	AUpΨ*A	4.7(3)
AΨ*pAU	1.3(1)	AΨ*pCG	2.1(1)
AΨ*pGC	0.85(6)	AΨ <sup>I</sup> pAU	4.1(1)
AΨ <sup>I</sup> pCG	4.2(1)	AΨ <sup>I</sup> pGC	5.1(1)
CGpAΨ <sup>I</sup>	0.98(4)	CGpAΨ*	2.7(3)
GCpAΨ <sup>I</sup>	1.1(1)	GCpAΨ*	1.7(1)
UApAΨ <sup>I</sup>	2.5(1)	UApAΨ*	1.8(1)
GCpUΨ	0.22(2)	GCpΨU	0.20(2)
CΨpCG	$5.3(4) \cdot 10^{-1}$	GCpCΨ	$6.0(6) \cdot 10^{-1}$
AUpGΨ	2.6(3)	AUpΨG	2.6(1)
CGpGΨ	2.2(1)	CGpΨG	2.7(1)
GCpGΨ	3.2(3)	GCpΨG	4.8(3)
GΨpAU	2.1(4)	UApGΨ	2.1(4)
Morse (meV)	47.2(9)		