

Mateus Rodrigues Leal

Desenvolvimento de modelos mesoscópicos
tridimensionais para o estudo da
desnaturação em DNA

Orientador: Gerald Weber

Belo Horizonte

2021

Mateus Rodrigues Leal

Desenvolvimento de modelos mesoscópicos
tridimensionais para o estudo da
desnaturação em DNA

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

Orientador: Gerald Weber

Belo Horizonte

2021

Dados Internacionais de Catalogação na Publicação (CIP)

L435d Leal, Mateus Rodrigues.
Desenvolvimento de modelos mesoscópicos tridimensionais para o estudo da desnaturação em DNA / Mateus Rodrigues Leal. – 2021.
107f., enc. : il.

Orientador: Gerald Weber.
Tese (doutorado) – Universidade Federal de Minas Gerais,
Departamento de Física.
Bibliografia: f. 84-90.

1. Termodinâmica. 2. DNA. 3. Interações moleculares.
I. Título. II. Weber, Gerald. III. Universidade Federal de Minas Gerais,
Departamento de Física.

CDU – 536.7 (043)

Ficha catalográfica elaborada por Romário Martins – CRB6 3595
Biblioteca Professor Manoel Lopes de Siqueira – Departamento de Física - UFMG



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

ATA DA SESSÃO DE ARGUIÇÃO DA 377ª TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA, DEFENDIDA POR MATEUS RODRIGUES LEAL orientado pelo professor Gerald Weber, para obtenção do grau de **DOUTOR EM CIÊNCIAS, área de concentração física**. Às 09:00 horas de vinte e nove de janeiro de dois mil e vinte um reuniu-se, por videoconferência, a Comissão Examinadora, composta pelos professores **Gerald Weber** (Orientador - Departamento de Física/UFMG), **Ubirajara Agero Batista** (Departamento de Física/UFMG), **João Antonio Plascak** (Departamento de Física/UFMG), **Elso Drigo Filho** (Departamento de Física/UNESP) e **Tauanne Dias Amarante** (Universidade de Cambridge) para dar cumprimento ao Artigo 37 do Regimento Geral da UFMG, submetendo o Mestre **MATEUS RODRIGUES LEAL** à arguição de seu trabalho de Tese de Doutorado, que recebeu o título de "**Desenvolvimento de modelos mesoscópicos tridimensionais para o estudo da desnaturação em DNA**". O candidato fez uma exposição oral de seu trabalho durante aproximadamente 50 minutos. Após esta, os membros da comissão prosseguiram com a sua arguição, e apresentaram seus pareceres individuais sobre o trabalho, concluindo pela aprovação do candidato.

Belo Horizonte, 29 de janeiro de 2021.

Prof. Geral Weber

Orientador do estudante

Departamento de Física/UFMG

Prof. Elso Drigo Filho

Departamento de Física/UNESP

Prof. Ubirajara Agero Batista

Departamento de Física/UFMG

Dra. Tauanne Dias Amarante

Universidade de Cambridge

Prof. João Antonio Plascak

Departamento de Física/UFMG

Candidato



Documento assinado eletronicamente por **Ubirajara Agero Batista, Servidor(a)**, em 29/01/2021, às 16:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gerald Weber, Professor do Magistério Superior**, em 29/01/2021, às 18:42, 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 **Tauanne Dias Amarante, Usuário Externo**, em 30/01/2021, às 00:31, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mateus Rodrigues Leal, Usuário Externo**, em 02/02/2021, às 20:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Elsó Drigo Filho, Usuário Externo**, em 06/02/2021, às 21:44, 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 **João Antonio Plascak, Membro**, em 07/02/2021, às 14:22, 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, informando o código verificador **0544917** e o código CRC **2D1E28EB**.

Agradecimentos

Ao Prof. Dr. Gerald Weber, pela orientação, paciência, apoio sempre que necessitei para o desenvolvimento desse trabalho. Tenho plena consciência que sua participação foi crucial para a realização desse trabalho, por isso carregarei comigo uma imensa gratidão. Também agradeço pelas boas conversas, conselhos e pelos bons cafezinhos.

À minha esposa Dayse, pelo companheirismo, compreensão e por acreditar nas minhas escolhas. Aos meus pais Geraldo e Helena, que sempre me deram apoio e o incentivo necessário para que eu pudesse trilhar meus estudos até aqui.

Aos professores e amigos do departamento de física pela amizade e momentos agradáveis que me proporcionaram ao longo do curso.

Ao CNPq e à Fapemig pela bolsa concedida.

Resumo

O modelo Peyrard-Bishop (PB) é parte de um conjunto de modelos chamados mesoscópicos, que consiste em uma descrição clássica das interações moleculares com potenciais simples e que permite o cálculo da temperatura de desnaturação com eficiência em moléculas como o DNA e RNA. Nesse modelo, a dupla fita é considerada perfeitamente plana, ou seja, sem a característica torção helicoidal e também não são considerados quaisquer parâmetros relacionados à dimensão molecular. O modelo permite a integração trivial de um dos seus graus de liberdade, restando apenas um grau para ser resolvido por técnicas de integração numérica como a técnica de integral de transferência, razão pela qual é mais conhecido como modelo unidimensional (1D). Na Hamiltoniana original, o modelo produz uma curva de desnaturação muito suave enquanto seria esperado uma transição mais abrupta, o que requer a adição de termos não-lineares artificiais. As definições usadas na Hamiltoniana PB tornam bastante difícil a comparação dos seus resultados com os de técnicas microscópicas como dinâmica molecular atômica. Desde a introdução do modelo em 1989, há um esforço em obter Hamiltonianas tridimensionais, mas, em geral, elas não são redutíveis à equações 1D, isto é, a um modelo onde reste apenas uma dimensão a ser integrada pela técnica de integral de transferência. Obter um modelo tridimensional que possa ser reduzido a apenas uma dimensão integrável tem sido um problema em aberto.

Nosso trabalho consistiu em fazer uma análise detalhada do modelo Peyrard-Bishop e elaborar um modelo tridimensional com características geométricas mais próximas dos modelos atômicos que o PB bidimensional. Desenvolvemos algumas aproximações que nos permitiram obter a redução da dimensão integrável, efetivamente obtendo um novo modelo 1D, e que nos permite utilizar a técnica de integral de transferência. Diferente do modelo oriundo da dimensão bidimensional, o novo modelo resulta em transições abruptas sem necessitar de potenciais artificialmente introduzidos. De fato, foi possível demonstrar que as transições abruptas resultam diretamente por termos considerando três graus de liberdade. Para quantificar os efeitos das aproximações, realizamos a integração numérica da Hamiltoniana sem aproximação alguma. Com isso mostramos que o novo modelo é adequado à condição de pequenos ângulos que é similar à situação da fita dupla quando se encontra próximo à temperatura de desnaturação. Uma vantagem adicional do novo modelo é que pode ser aplicado sem modificação ao conjunto de técnicas desenvolvidas para sequências de DNA heterogêneas.

Palavras-chave: Peyrard-Bishop, Peyrard-Bishop tridimensional, convergência, DNA, Integral de Transferência.

Abstract

The Peyrard-Bishop model is part of a class of mesoscopic models which consists in the description of the molecular interactions by simple potentials and which allows the efficient calculation of the melting temperatures of molecules such as DNA and RNA. In this model the double strand is considered as perfectly flat, that is, without its characteristic helicoidal torsion and neither considers any structural parameters. The model can be trivially integrated in one of its degrees of freedom such that only one degree remains to be solved by numerical integration techniques such as the transfer integral method, for which it is often referred to as a 1D model. For the original Hamiltonian the model results in a melting curve which increases slowly with temperature while a more abrupt transition would have been expected and requires the addition of artificial non-linear terms. The definitions used by the PB Hamiltonian makes it difficult to compare its results with those of microscopic techniques such as atomistic molecular dynamics. Since its introduction in 1989 there has been an effort to obtain tridimensional Hamiltonians, however they could not be reduced to a 1D model, that is, a model where only one of its degrees of freedom remains to be integrated by the transfer integral method. Obtaining a tridimensional model that could be reduced to a single integrable dimension has since been an open problem.

Here, we performed a detailed analysis of the Peyrard-Bishop model and proposed a tridimensional model with geometric definitions which are closer to atomistic models than that of the 2D model. We developed several approximations which allowed us to obtain a reduction of the integrable dimensions, effectively obtaining a new 1D model which allows is to use the transfer integral method. Differently from the bidimensional model, the new model results in abrupt transitions without the need to add artificial potentials. In fact, we were able to show that the abrupt transitions are a result of the three degrees of freedom. To evaluate the effect of our approximations we performed the numerical integration of the non-approximated Hamiltonian. In this way we could show that the new model is appropriate under the conditions of small angles which is similar to the double strand close to the melting temperature. One additional advantage of the new model is that it can be used without modifications for the existing techniques which were developed for heterogeneous DNA.

Keyword: Peyrard-Bishop, tridimensional Peyrard-Bishop, convergence, DNA, transfer integral.

Lista de Figuras

1	Difração de raio-X	12
2	Formas A, B e Z do DNA	13
3	Composição química de um nucleosídeo e das base nitrogenadas canônicas do DNA	14
4	Estrutura do B-DNA	15
5	Representação planar da estrutura química do DNA	16
6	Esquema do arranjo experimental que permite medir parâmetros de rigidez na molécula de DNA	16
7	Rigidez da molécula de DNA	17
8	Esquema simplificado do aparato experimental para absorção UV	19
9	Absorção UV para uma sequência C-G.	19
10	Comparativo entre dados experimentais e teóricos.	23
11	Desnaturação baseada na equação de Langevin	26
12	Esquema do DNA descrito pelo modelo PB	26
13	Potencial de Morse	29
14	Região de integração	30
15	Comparação das curvas de desnaturação usando o potencial harmônico e anarmônico.	37
16	Divergência no modelo Peyrard-Bishop 1D	37
17	Eliminação da divergência no modelo Peyrard-Bishop 1D	38
18	Superfície descrita pelo núcleo da função de partição.	39
19	Degenerescência dos autovalores.	40
20	Comportamento de Z em função de ω_{\max}	43
21	Comportamento de ω_{\max}	43
22	Comportamento de $\sqrt{\omega_{\max}}$ com a temperatura de desnaturação.	44
23	Esquema de um modelo Peyrard-Bishop 3D	47
26	Esquema do modelo Peyrard Bishop 3D	55

27	Modelo tridimensional aproximado	56
28	Projeção de $J_{n,n-1}$ no plano xy	57
29	Exponencial do potencial elástico com $N = 2$ em termos de $f(r_1, r_2)$	61
30	Região de integração sem a torção	64
31	$\langle r \rangle \times b$ usando $\omega = 0$ rad e diferentes níveis de aproximação	65
32	$\langle r \rangle \times b$ usando $\omega = 0.0$ rad e diferentes níveis de aproximação	66
33	$\langle r \rangle \times \phi_0$	67
34	$\{\langle r \rangle, Z\} \times \zeta$	68
35	$\langle r \rangle \times T$ com $\omega = 0,05$ rad e diferentes níveis de aproximação	69
36	$\{\langle r \rangle, Z\} \times J_0$	72
37	$\{\langle r \rangle, Z\} \times k$	73
38	Potencial de Morse com diferentes valores de D	73
39	$\{\langle r \rangle, Z\} \times D$	74
40	Potencial de Morse com diferentes valores de a	75
41	$\{\langle r \rangle, Z\} \times a$	75
42	$\lambda \times b$	76
43	$\lambda \times T$	80
44	$\langle r \rangle \times T$ com diferentes valores de N	81

Sumário

Resumo	5
Abstract	6
1 Introdução	12
1.1 A conformação e a composição química dos ácidos desoxirribonucleicos . . .	14
1.2 Função biológica	17
1.3 Desnaturação Térmica	18
1.4 Modelos teóricos de DNA	23
1.5 O modelo Peyrard-Bishop bidimensional para o DNA	25
1.5.1 Sequências homogêneas	28
1.5.2 Técnica de Integral de Transferência (TIT)	32
1.5.3 Resultados obtido pelo Modelo Peyrard Bishop e algumas de suas variações	36
1.5.4 Modelo Peyrard-Bishop para sequência heterogênea	39
2 Modelo Peyrard-Bishop tridimensional	46
2.1 Desenvolvimento Matemático do cálculo de Z e $\langle r \rangle$ para a versão tridimensional	50
3 Modelo Peyrard-Bishop tridimensional com três graus de liberdade	54
3.1 Modelo 3D aproximado	55
3.2 Análise numérica	63
3.3 Comparação de Z e $\langle r \rangle$ usando as hamiltonianas exata e aproximada.	64
3.3.1 Variação dos Parâmetros dos potenciais	68
3.3.2 Autovalores	76
3.3.3 Estudo numérico de sequências grandes usando o Modelo 3D aproxi- mado	81
4 Conclusão	82

Referências	83
A Apêndice	90
Artigo associado	94

1 Introdução

Em 1951, Rosalind Franklin realizou estudos com DNA em meios onde a cristalização pudesse ser variada, o que permitiu a ela obter das imagens de difração de raios-X (figura 1) obtidas melhores que aquelas existentes até então. Em meios onde a umidade era 75%, os experimentos mostravam uma cristalização já observada anteriormente (o que ela denominou de forma A), mas quando a umidade era 92%, as fitas tornavam-se mais alongadas (forma B) e produziam imagens mais nítidas [1]. As imagens obtidas por Franklin (a forma B) foram utilizadas por Watson e Crick (por meios considerados no mínimo questionável [2]) para obter a famosa estrutura de dupla hélice, trabalho que lhes rendeu o prêmio Nobel de Fisiologia em 1962 [3]. Nos anos seguintes, estabeleceu-se a classificação das fibras de DNA analisadas por difração de Raios-X em A-DNA ou B-DNA, tornando o modelo de dupla hélice um consenso na comunidade científica (revisado em [4]).

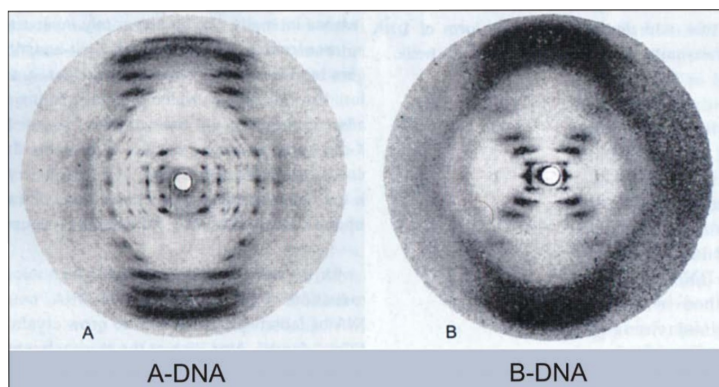


Figura 1

Imagens obtidas por difração de raio-X. As regiões escuras indicam onde houve interferência destrutiva e os claros, interferência construtiva. Figura retirada de [5].

As amostras utilizadas nos estudos com difração de raios-X eram obtidas diretamente de tecidos e por isso eram repletas de impurezas e continham tamanhos sem padronização. Isso não permitia uma análise mais refinada, como por exemplo diferenciar os variados tipos de bases, mas apenas fazer proposições de um modelo estrutural idealizado que representasse a



Figura 2

Esquema representativo de dupla hélice nas configurações A-DNA, B-DNA e Z-DNA. A estrutura geométrica das formas A e B se diferenciam principalmente no passo da hélice, já a forma Z possui a torção no sentido contrário das duas primeiras. Além disso, o passo da hélice é maior que nos casos A e B. Figura retirada de [5].

média das estruturas analisadas [6].

A síntese de DNA foi desenvolvida na década de 1970 e isso possibilitou que pequenos fragmentos de DNA puros, com tamanhos e sequências definidos, fossem cristalizados. Com isso, finalmente estava-se diante da real possibilidade de ver experimentalmente com detalhes o modelo proposto por Watson e Crick [7]. Uma enorme expectativa foi criada acerca da análise dos primeiros seguimentos de DNA sintetizado e, em 1979, a análise de uma sequência poli(dC-dG) mostrou que havia uma conformação diferente das formas A-DNA e B-DNA. A nova configuração possuía o sentido da dupla hélice invertido em relação às formas até então conhecidas e a cadeia principal possuía um formato de zigue-e-zague [8], razão pelo qual foi denominado Z-DNA[9]. As três formas conformacionais estão esquematizadas na figura 2.

Com o passar do tempo, o aumento do número de estruturas de DNA analisadas possibilitou estabelecer que a forma mais abundante *in vivo* é a forma B-DNA e que as outras podem ser induzidas por fatores como baixa umidade relativa, presença de álcoois ou íons na solução [10, 11].

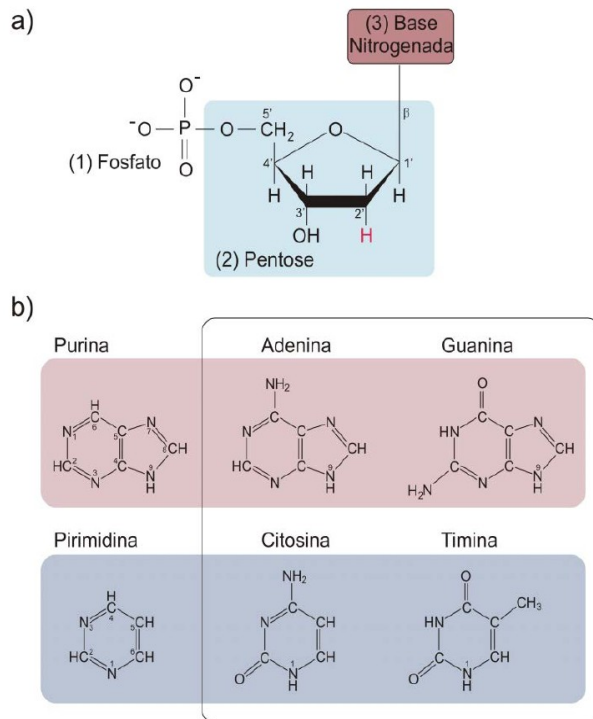


Figura 3

(a) Estrutura de um nucleosídeo, mostrando a numeração dos carbonos da pentose. A figura (b) mostra a estrutura das quatro bases nitrogenadas canônicas do DNA e o composto que caracteriza as Purinas e Pirimidinas. Figura retirada de [5].

1.1 A conformação e a composição química dos ácidos desoxirribonucleicos

O DNA canônico é composto de uma unidade básica denominada nucleotídeo, que por sua vez é composto por um nucleosídeo e um grupo fosfatado, conforme esquematizado na figura 3. O nucleosídeo é formado por uma base nitrogenada (as canônicas são a citosina, adenina, guanina, timina ou uracila) e um açúcar (ribose no caso do RNA ou desoxirribose no caso do DNA). As bases nitrogenadas são divididas em dois grupos: Purinas e Pirimidinas, sendo que no DNA ocorre com maior frequência adenina e guanina (Purinas) e timina e citosina (Pirimidinas).

O empilhamento dos nucleotídeos ocorre devido à ligações fosfodiéster, sempre no sentido da hidroxila ligada ao carbono C3' para a hidroxila ligada ao carbono C5' do açúcar. Cada molécula de açúcar liga-se com uma base nitrogenada por uma ligação β -glicosídica entre o carbono C1' do anel da pentose e um dos nitrogênios das bases, sendo N1 nas Pirimidinas e

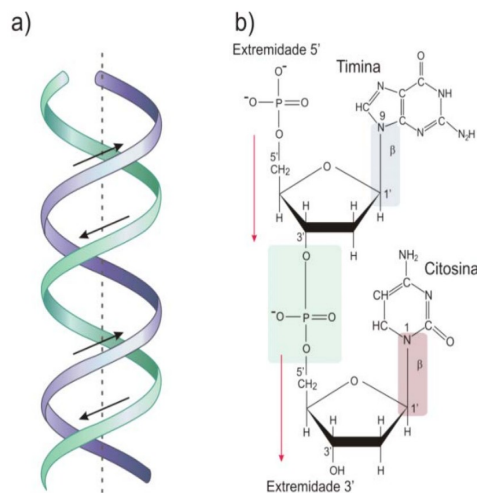


Figura 4

(a) Estrutura esquemática da forma B-DNA. (b) Mostra a orientação das ligações em cada fita, formando o empilhamento. Figura retirada de [5].

N9 nas Purinas, conforme a figura 4.

A disposição dos nucleotídeos na hélice são de tal forma que os fosfatos e os açúcares ficam na parte mais externa e as bases nitrogenadas localizam na parte mais interna, possibilitando a formação de ligações de hidrogênio entre as bases nitrogenadas complementares nas duas fitas, conforme a figura 5.

O número grande de ligações de hidrogênio entre as bases e disposição dos grupos fosfatos para a parte externa da hélice faz com que a molécula adquira uma alta polaridade, ficando a parte externa negativa e a interna positiva. Essa distribuição de cargas na parte interna faz com que haja repulsão entre as fitas, criando um espaço entre elas. A estabilidade é alcançada devido à interação com íons positivos presentes ao seu redor, como por exemplo Na^+ ou Mg^{2+} [12], com isso a estrutura torna-se bastante rígida [13–15].

Uma das possibilidades de se obter experimentalmente os parâmetros da rigidez de uma determinada molécula de DNA é prender suas extremidades e movimentar uma delas, conforme a figura 6 [14].

No caso da figura 6, para fixar a molécula de DNA usou-se uma esfera de polistireno, cujo diâmetro mede 520 ± 10 nm, na extremidade capturada pela armadilha óptica, já a outra extremidade foi ligada a um pedaço de RNA fixado na base de vidro. Ao movimentar a base,

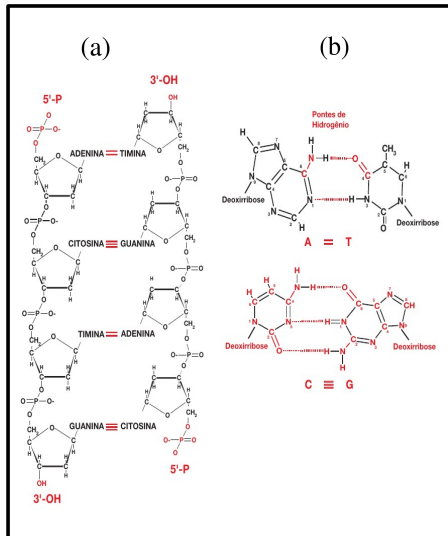


Figura 5
 Representação planar das quatro bases canônicas do DNA, mostrando as ligação de hidrogênio que ocorre entre elas. Figura retirada de [5].

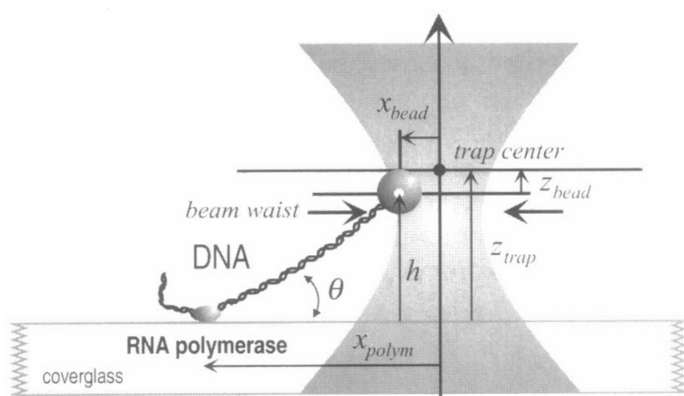


Figura 6
 Esquema do arranjo experimental que permite medir parâmetros de rigidez na molécula de DNA. Figura retirada de [14].

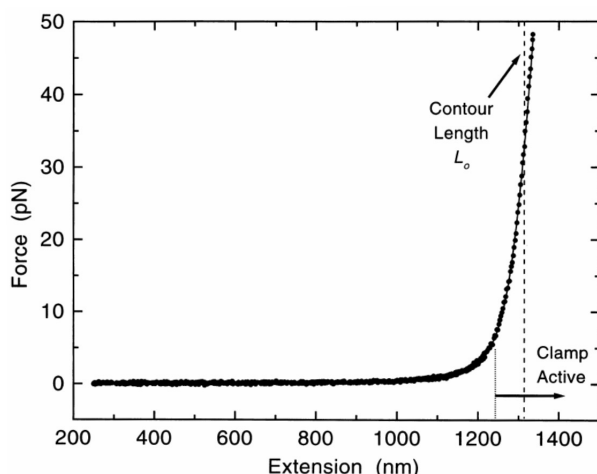


Figura 7

A figura mostra como o comprimento de uma molécula de DNA varia ao ser “esticada”. Nesse caso específico, o comprimento da molécula pode se aumentar até por volta dos 1200 nm sem esforço e, a partir desse comprimento, observamos um aumento acentuado na força necessária para promover um pequeno incremento no comprimento. Figura retirada de [14].

pode-se medir o estiramento e a força necessária para segurar a esfera de poliestireno e, assim, obter a relação de elasticidade da molécula, mostrada na figura 7.

1.2 Função biológica

As proteínas são as macromoléculas biológicas mais importantes, sendo responsáveis, em média, por mais da metade da massa de uma célula e estão presentes em todo ser vivo. Uma proteína é um polímero de aminoácidos e pode atuar como enzima, veículo de outras moléculas ou íons, com função motora para auxiliar no movimento em células e tecidos, participam na regulação gênica, desempenham papéis importantes no sistema imunológico, entre outras tantas funções. As proteínas são formadas por aminoácidos unidos por ligações covalentes (também conhecida por ligação peptídica) formando longos polímeros.

O ordenamento dos aminoácidos que compõe a proteína é determinado pelo DNA, sendo que a cada três bases nitrogenadas consecutivas, temos a menor unidade do código genético, o Códon. Por exemplo, uma sequência formada por citosina-timina-timina (C-T-T) é o código para o aminoácido leucina, já a trinca C-A-C codifica o aminoácido histidina. É muito importante destacar que a relação códon-aminoácido não é 1:1, ou seja, um mesmo aminoácido pode ser produzido por mais de um códon, como por exemplo o aminoácido histidina, que também é codificado pelo códon C-A-T. O conjunto de códons necessários para montar uma proteína chama-se gene e o conjunto de genes de um ser vivo forma seu código genético.

Para realizar suas principais funções (tradução, replicação e transcrição) a molécula de DNA necessita desconectar as duas fitas, uma vez que o RNA precisa se ligar ao DNA para executar os processos. Sendo assim, o estudo do fenômeno de abertura das fitas é imprescindível para desvendarmos os processos celulares envolvendo o DNA. O processo de replicação, por exemplo, é feito de forma coordenada, semi conservativa e antiparalela e pode incluir outras moléculas, como por exemplo proteínas estabilizadoras como as girases e topoisomerases. No início da replicação, o DNA é desenrolado e a dupla fita é separada, formando o que chamamos de forquilha de replicação. É nesse espaço que ocorre o acréscimo de nucleotídeos no sentido oposto da abertura, formando uma nova fita. Assim o novo duplexo é formado por uma fita original outra nova (fita filha) com orientação antiparalela (sentido oposto).

1.3 Desnaturação Térmica

Conforme descrito anteriormente, para que o DNA realize suas principais funções (replicação, tradução e transcrição) é necessário que as duas fitas que compõem a hélice se desconectem, isto é, uma fita deve se separar da outra. A separação pode ocorrer devido aos diferentes fatores como concentração salina [16], força aplicada as fitas (experimentos usando pinça óptica) [17] e temperatura [18]. Nesse texto vamos focar na desnaturação térmica, pois o nosso trabalho depende da utilização de dados experimentais termodinâmicos envolvendo o DNA para ajustar os parâmetros do modelo que propomos.

A baixas temperaturas predomina, em solução, DNA formado por duplexos, mas à medida que aquecemos a solução, as ligações de hidrogênio entre os pares de bases nitrogenadas vão se rompendo, formando duas fitas simples. Os pares A-T e C-G possuem, respectivamente, duas e três ligações de hidrogênio, portanto necessitam de temperaturas diferentes para desconectarem. Ao romper as ligações de hidrogênio, a fita simples passa a absorver luz ultravioleta na faixa de 260 nm devido à transição eletrônica $\pi - \pi^*$ nas bases nitrogenadas, que ocorre por causa da mudança da configuração eletrônica decorrente da redução do empilhamento e pareamento [19, 20]. Então pode-se incidir radiação UV com intensidade

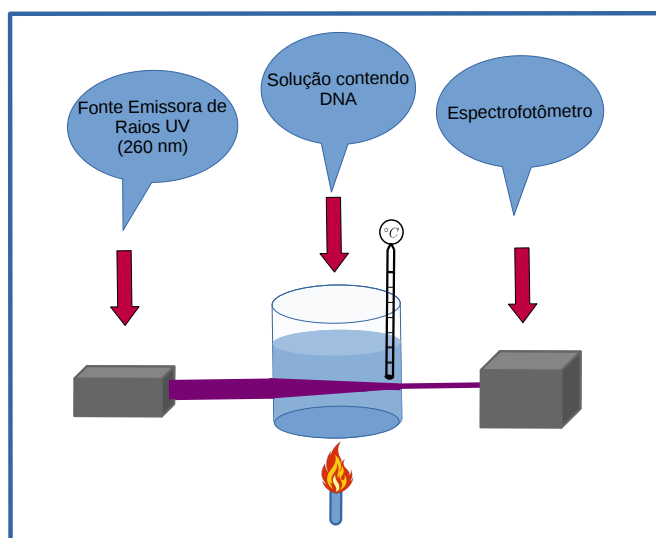


Figura 8

A figura mostra um esquema simplificado de experimento com UV em solução de DNA.

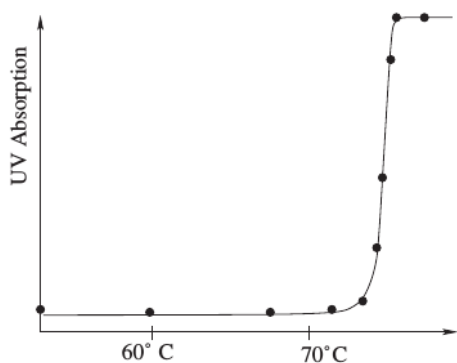


Figura 9

A figura mostra a curva de absorção UV *versus* temperatura para uma sequência C-G. Figura retirada de [21].

conhecida (I_0) na solução, aquecê-la vagarosamente e, com um espectrofotômetro, medir a intensidade (I) da radiação que passa pela solução, conforme o esquema mostrado na figura 8. Com isso, sabemos quanto de radiação foi absorvido pela amostra nas diferentes temperaturas. A diferença entre a intensidade incidente e a transmitida (em escala logarítmica) é o que chamamos de Absorbância (A), dada por

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

O aumento da absorbância com o acréscimo da temperatura obtida em um experimento de espectrometria UV para uma sequência C-G [21] é mostrado na figura 9.

Conforme comentamos anteriormente no texto e confirmado na figura 9, o aumento de temperatura de fato é um dos fatores que leva o rompimento da fita, portanto podemos con-

cluír que o estudo das propriedades termodinâmicas do DNA é essencial para compreender o comportamento do DNA. A compreensão dos processos de tradução, replicação e transcrição possui aplicações em técnicas de PCR (reação de polimerase em cadeia, utilizada para sequenciamento de DNA) [22], produção de sondas [23] e *gene array* [24]. É importante destacar que a temperatura de desnaturação (T_m) é definida como aquela em que metade dos nucleotídeos na solução estão desemparelhados. A temperatura de desnaturação (T_m) está relacionada com a concentração de fitas de DNA na solução e esta relação é diferente para o caso de fitas que denominamos auto-complementares e não auto-complementares. Uma é dita fita auto-complementar quando, independente da extremidade que partimos, a sequência de pares de base é a mesma. Já uma fita não auto-complementar é quando essa propriedade não ocorre. Por exemplo: uma sequência ATA é auto-complementar, enquanto que a sequência ATG é não auto-complementar.

Vamos definir f como a fração de duplexos dissociados, $[X]$ e $[Y]$ como as concentrações de cada fita separada, $[X.Y]$ como a concentração de duplexo e C_t como a concentração total de fitas. No caso das fitas auto-complementares temos:

$$f = \frac{[Y]}{[X.Y]} = \frac{[X]}{[X.Y]} \quad (2)$$

e

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

A constante de equilíbrio K_{eq} é definida como

$$K_{eq} = \frac{[\text{produtos}]}{[\text{reagentes}]} = \frac{[X][Y]}{[X.Y]}, \quad (4)$$

onde, por definição, reagente é todo componente presente antes da reação e produto é tudo que se forma após a reação ocorrer. Nesse caso específico, a reação é a desnaturação, portanto, antes da reação temos a fita dupla (reagente) e após a desnaturação temos as fitas separadas (produtos).

Como a temperatura T_m é definida como aquela em que metade das fitas estão desacopla-

das, temos $[x] = [y] = [X.Y]$. Assim, usando a equação (3) obtemos

$$[X] = [Y] = [X.Y] = \frac{[C_t]}{4}. \quad (5)$$

Substituindo (5) na equação (4) obtemos

$$K_{eq} = \frac{[X].[Y]}{[X.Y]} = \frac{\left(\frac{[C_t]}{4}\right)\left(\frac{[C_t]}{4}\right)}{\left(\frac{[C_t]}{4}\right)} = \frac{[C_t]}{4}. \quad (6)$$

A relação entre a constante de equilíbrio K_{eq} e a temperatura pode ser obtida utilizando-se do fato de que as diferenciais da energia livre de Gibbs $g(T, P)$ e da entalpia $h(s, P)$ são dadas por

$$dg(T, P) = -sdT + vdP \quad (7)$$

$$dh(s, P) = Tds + vdP. \quad (8)$$

Subtraindo a equação (7) da (8) obtemos

$$dg = dh - Tds - sdT. \quad (9)$$

Como T_m é uma constante, a equação (9) torna-se

$$dg = dh - T_m ds, \quad (10)$$

integrando a equação (10) obtemos

$$\Delta g = \Delta h - T_m \Delta s. \quad (11)$$

O químico neerlandês Jacobus Henricus van't Hoff estudou a pressão osmótica, em uma membrana, causada pela passagem de soluto de um meio menos para outro mais concentrado e para solucionar esse problema, Van't Hoff baseou-se na equação dos gases ideais. A ideia

é mapear o problema do DNA na resolução de Van't Hoff. Definimos $K = \frac{1}{v}$ e usando uma equação de estado similar aos gases ideais, temos

$$P(T, K) = P(K) = \eta T_m K \implies \left(\frac{\partial K}{\partial P} \right)_T = \frac{1}{\eta T_m}. \quad (12)$$

Da equação 7 podemos obter a seguinte relação

$$\left(\frac{\partial g}{\partial P} \right)_T = \frac{1}{K}. \quad (13)$$

Então

$$\left(\frac{\partial g}{\partial P} \right)_T = \left(\frac{\partial K}{\partial P} \right)_T \left(\frac{\partial g}{\partial K} \right)_T. \quad (14)$$

Substituindo as equações (12) e (13) em (14) obtemos

$$\frac{1}{K} = \frac{1}{\eta T_m} \left(\frac{\partial g}{\partial K} \right)_T \implies \left(\frac{\partial g}{\partial K} \right)_T = \frac{\eta T_m}{K} \quad (15)$$

$$\Delta g = \eta T_m \int_{K_i}^{K_f} \frac{dK}{K} \implies \Delta g = -\eta T_m \ln \left(\frac{K_i}{K_f} \right) \quad (16)$$

Como a constante de equilíbrio K_{eq} é definida pela razão das concentrações, K_i e K_f também são dados por uma “densidade”, portanto podemos escrever

$$\Delta g = -\eta T_m \ln (K_{eq}). \quad (17)$$

Substituindo (17) na equação (11) obtemos

$$\Delta h - T_m \Delta s = \eta T_m \ln (K_{eq}) \quad (18)$$

ou

$$\frac{1}{T_m} = -\frac{R}{\Delta H} \ln(C_t) + \frac{\Delta S - R \ln(4)}{\Delta H}. \quad (19)$$

No caso de fitas complementares, a principal diferença é que as fitas são iguais quando olhamos uma no sentido contrário da outra. Assim na equação das concentrações não há $[X]$ e

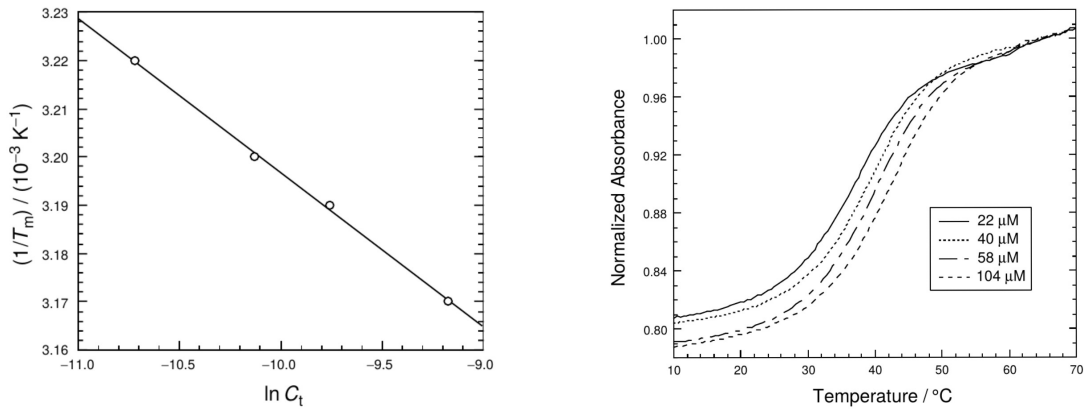


Figura 10

A figura da esquerda compara os dados experimentais com o previsto pela equação (19) e a da direita mostra a curva de absorção UV *versus* temperatura para quatro concentrações diferentes de fitas auto-complementares $d(CAAAAAAG)$ e $d(GTTTTTTC)$. Figura retirada de [20].

[Y]. Não vou refazer os cálculos por serem similares ao caso que acabamos de fazer, mas o resultado está expresso na equação (20).

$$\frac{1}{T_m} = -\frac{R}{\Delta H} \ln(C_t) + \frac{\Delta S}{\Delta H}. \quad (20)$$

As equações (19) e (20) relacionam a temperatura de desnaturação com a concentração da solução e isso de fato é comprovado experimentalmente, conforme mostrado na figura 10 da esquerda. Por outro lado, é de se esperar que a absorbância diminua com o aumento da concentração, pois à medida que temos mais fitas para “absorver” a radiação UV menos radiação irá passar pela solução, o que diminui a razão entre as intensidades incidente e transmitida. Essa conclusão também está de acordo com a equação (1) e também é observado experimentalmente, conforme a figura 10 da direita.

1.4 Modelos teóricos de DNA

Na tentativa de descrever as propriedades termodinâmicas do DNA, surgiram diversos modelos teóricos que são classificados em três grupos: macroscópico, microscópico e mesoscópico. De maneira geral, os modelos macroscópicos fornecem bons resultados para fenômenos que ocorrem na escala da molécula como um todo, já os microscópicos conseguem descrever

fenômenos locais e específicos e os mesoscópicos são intermediários aos dois primeiros. Os modelos microscópicos visam descrever os fenômenos locais e específicos de cada par de base. Podemos citar como exemplos de modelos microscópicos os cálculos de DFT (*Density Functional Theory*) [25] e dinâmica molecular [26]. Estes modelos possuem uma grande riqueza de detalhes na descrição dos fenômenos, todavia possuem alto custo computacional, sendo por isso inviável para o estudo de moléculas grandes.

Os modelos macroscópicos fornecem bons resultados para fenômenos que ocorrem na escala da cadeia toda. Podemos citar como exemplos os modelo WLC (*Worm-Like-Chain*) [27] e o de próximos vizinhos [28]. Este último consiste em atribuir incrementos de energia livre de Gibbs (ΔG) a cada par vizinho. Esse método funciona razoavelmente bem, além de ser relativamente simples, uma vez que ΔS e ΔH de cada ligação pode ser obtido a partir de um conjunto de temperaturas experimentais como as da figura 9. Assim, pode-se usar as equações (19) e (20) para obter T_m . O principal problema do modelo de próximos vizinhos é que quando obtemos os valores de ΔH e ΔS para uma determinada sequência, não conseguimos distinguir a contribuição da ligação de hidrogênio e do empilhamento [28]. Os modelos macroscópicos não são capazes de descrever fenômenos termodinâmicos localizados, como por exemplo o surgimento dos *hairpins* e *bulges* que aparecem ao longo da hélice [29].

Os primeiros modelos mesoscópicos surgiram na década de 1960, inspirados no modelo de Ising 1D [30]. Esse modelo não levou em conta as interações de empilhamento e seus resultados mostram que a transição ocorre de maneira muito suave ($\Delta T > 50$ K), o que difere muito dos dados experimentais. A cooperatividade é responsável por restringir a transição para poucos kelvins e é responsável pela estabilidade da dupla hélice. Um modelo levando em conta a energia de empilhamento foi proposto posteriormente, tornando os parâmetros como energia livre no local da abertura e cooperatividade dos pares mais realísticos, embora o modelo não considerou a entropia da cadeia [31]. Na década de 1970, Poland e Schegara [32] implementaram uma versão mais sofisticada do modelo de Ising, onde eles acrescentaram a entropia associada a cada "bolha" de comprimento n ($\Delta S_{LE} = ck_B \ln(n)$), sendo esse o primeiro modelo a considerar alguma orientação na desnaturação, pois o requisito para a existência de uma bolha é que suas extremidades sejam fechadas. Isso leva, na situação

de equilíbrio, uma interação eficaz de longo alcance nesse modelo originalmente 1D, o que, a princípio, permite uma transição termodinâmica no limite infinito para o comprimento do polímero. Em 1989, Peyrard e Bishop desenvolveram um modelo que possui solução analítica baseado em potenciais que descrevem as ligações entre os nucleotídeos, o qual iremos descrever com detalhes na próxima seção.

Ainda podemos estudar a dinâmica do DNA através da equação de Itô (Também conhecida por equação de Langevin) [33], dada por

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (21)$$

Na equação (21), o termo X_t pode ser relacionado, nesse caso, com o deslocamento, o termo $b(t, X_t)$ representa a parte determinística, como por exemplo a velocidade (podemos ter contribuição de força proporcional a velocidade) e $\sigma(t, X_t)dB_t$ é a contribuição estocástica, representando por exemplo a viscosidade, o atrito, etc.

Nessa abordagem, obtemos um conjunto de equações diferenciais acopladas que devem ser resolvidas como função do tempo [34–36]. Esse método de tratamento nos permite obter tanto informações localizadas quanto gerais sobre o polímero, como por exemplo a ocorrência de “bolhas” ou a própria desnaturação da molécula, como mostrado na figura 11.

O principal problema desse método é o custo computacional, uma vez que para um conjunto de N pares de base, temos $2N$ equações diferenciais não lineares para resolver, pois os potenciais que descrevem as ligações de hidrogênio são não lineares como por exemplo o potencial de Morse, além disso, o potencial de empilhamento também pode ser não linear, como o usado por Dauxois et al. [37].

1.5 O modelo Peyrard-Bishop bidimensional para o DNA

O modelo Peyrard-Bishop (PB) foi apresentado pela primeira vez em 1989 [38] e seu sucesso continua sendo bastante significativo nos dias atuais [16, 39, 40]. A principal vantagem deste modelo é descrever as interações entre os nucleotídeos por potenciais com bastante simplicidade e, além disso, permite estudar sequências heterogêneas, que é o caso de maior

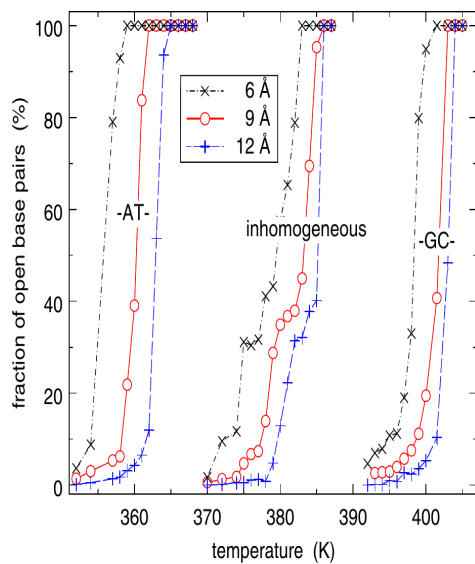


Figura 11

Desnaturação baseada na equação de Langevin usando 1793 pares de bases. As curvas em preto, vermelho e azul são obtidas usando três diferentes parâmetros para a separação média dos nucleotídeos. Figura retirada de [34]

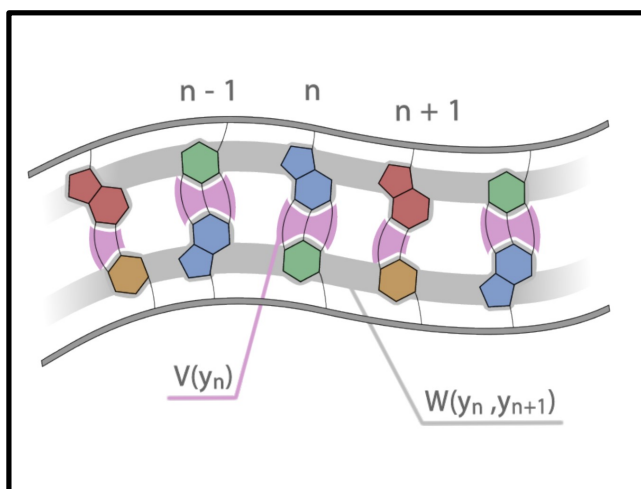


Figura 12

Esquema representativo para as interações consideradas no modelo Peyrard-Bishop do DNA.

relevância prática (ver figura 12). O fato de usar potenciais para descrever as interações entre os nucleotídeos permite adicionar termos que correspondem, por exemplo, as interações com solventes [16, 40–46]. Além disso, podemos usar diferentes potenciais para simular as ligações entre os componentes do DNA, como mostraremos adiante neste texto. Outra vantagem é que o modelo é facilmente estendido ao estudo de RNA [47], modificando-se basicamente as constantes dos potenciais. No caso de RNA, as constantes mudam porque os componentes são diferentes, por exemplo, o açúcar do DNA é uma desoxirribose, já no RNA é uma ribose. Originalmente, o modelo Peyrard-Bishop propôs o potencial harmônico para descrever o empilhamento e o potencial de Morse para representar a ligação entre as bases complementares (ligação entre as fitas) [38]. Mais tarde, surgiu uma modificação no potencial de empilhamento, substituindo-o por um potencial anarmônico, que melhorou o modelo Peyrard-Bishop original no sentido de que a curva de desnaturação teórica passou a concordar com curva obtida experimentalmente [37, 48].

Estamos interessados nas situações práticas onde um conjunto de sequências de DNA é colocado em uma solução que será aquecida até atingir a temperatura de desnaturação, portanto estamos trabalhando com um reservatório térmico. Neste caso, o ensemble estatístico mais adequado para descrever o sistema é o ensemble canônico e as grandezas termodinâmicas são obtidas a partir da função de partição canônica (Z), que pode ser compreendido como uma função da temperatura que normaliza a probabilidade do sistema encontrar-se com energia H . A expressão para obter Z é

$$Z = \Lambda \int \int dp_i dq_i \exp[-\beta H(q_i, p_i)], \quad (22)$$

onde Λ é uma constante que deixa Z adimensional e β é

$$\beta = \frac{1}{k_B T}. \quad (23)$$

k_B é a constante de Boltzmann.

Em nosso problema, os potenciais não dependem da velocidade e a energia mecânica é de

ordem quadrática na velocidade, portanto o hamiltoniano (H) é a soma da energia cinética (E) com a energia potencial de empilhamento, que originalmente foi adotada a energia potencial elástica (W) [38] e a energia de interação dos nucleotídeo que conectam uma fita à outra (V).

1.5.1 Sequências homogêneas

O modelo Peyrard Bishop foi inicialmente proposto para sequências homogêneas, isto é, todos os pares de base foram considerados como sendo do mesmo tipo, ou seja uma sequência somente composta por CG ou somente por AT. Matematicamente a homogeneidade é descrita nos potenciais pelo uso dos mesmos valores das constantes em todos os pares de base. Para introduzirmos o modelo Peyrard-Bishop, vamos definir algumas quantidades que facilitarão o entendimento e sua representação matemática do.

Definição 1. Sejam u e v as variáveis que representam as posições das bases relativas ao ponto de equilíbrio em cada fita do DNA e p a variável de momento. Dessa forma, p_n^u e v_n representam, respectivamente, o momento e a posição da n -ésima base relativa a posição de equilíbrio da variável r .

Definição 2. Seja $W(u_n, u_{n-1})$ a energia potencial da “mola” entre as bases n e $n - 1$ da fita u .

Definição 3. Seja $V(u_n, v_n)$ a energia potencial que representa a ligação de hidrogênio entre o n -ésimo par de base das fitas u e v .

De posse das definições acima, podemos escrever o hamiltoniano do nosso sistema, cuja expressão matemática é dada pela equação (24)

$$H = \sum_{n=1}^N \underbrace{\frac{p_{u_n}^2}{2m} + \frac{p_{v_n}^2}{2m}}_{E_n} + W(u_n, u_{n-1}, v_n, v_{n-1}) + V(u_n, v_n), \quad (24)$$

onde

$$W(u_n, u_{n-1}, v_n, v_{n-1}) = \frac{k}{2}(u_n - u_{n-1})^2 + \frac{k}{2}(v_n - v_{n-1})^2 \quad (25)$$

$$V(u_n, v_n) = D \{ \exp [-\alpha(u_n - v_n)] - 1 \}^2. \quad (26)$$

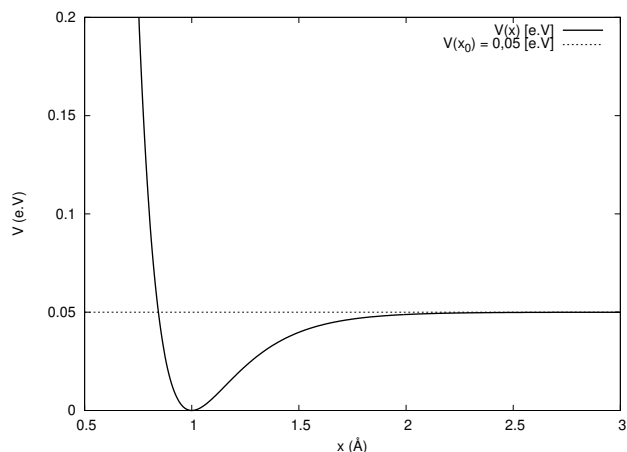


Figura 13

A figura mostra o potencial de Morse e sua assíntota horizontal.

Parâmetros: $D = 0,05 \text{ eV}$ e $a = 4,45 \text{ nm}^{-1}$

O potencial de Morse em uma variável é descrito por

$$V(x) = D [\exp[-a(x - x_0)] - 1]^2 \quad (27)$$

e é adequado para descrever as ligações de hidrogênio porque possui o seguinte comportamento: considere o ponto x_0 onde

$$F(x_0) = -\left. \frac{\partial V}{\partial x} \right|_{x=x_0} = 0.$$

A força derivada deste potencial é repulsiva para $x < x_0$, atrativa para $x > x_0$ e diminui a medida que $x > x_0$ afasta do ponto de equilíbrio (x_0) como mostra a figura 13.

A função de partição é dada pela integração no espaço de fase de $\exp(-\beta H)$, sendo que podemos integrar a parte da energia cinética da equação (24) e obter

$$Z = \frac{1}{h^{2N}} \underbrace{\int_{-\infty}^{+\infty} \cdots \prod_{n=1}^N dp_{u_n} dp_{v_n} \exp[-\beta E_n(p_{u_n}, p_{v_n})]}_{\Lambda'} \times \int_{-\infty}^{+\infty} \prod_{n=1}^N du_n dv_n \exp\{-\beta [W(u_n, u_{n-1}, v_n, v_{n-1}) + V(u_n, v_n)]\} \quad (28)$$

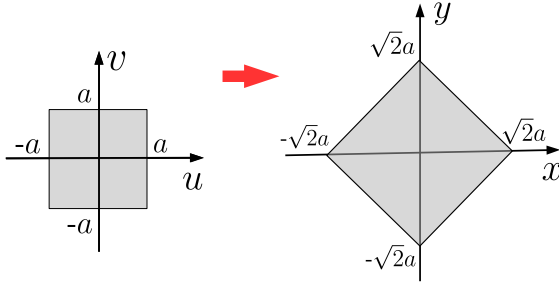


Figura 14

Região de integração nas variáveis (u, v) e (x, y) .

$$Z = \Lambda' \int_{-\infty}^{+\infty} \prod_{n=1}^N du_n dv_n \exp \{-\beta [W(u_n, u_{n-1}, v_n, v_{n-1}) + V(u_n, v_n)]\}, \quad (29)$$

onde h é a constante de Planck.

Faremos a seguinte mudança de variáveis:

$$x_n = \frac{u_n + v_n}{\sqrt{2}} \quad (30)$$

$$y_n = \frac{u_n - v_n}{\sqrt{2}} \quad (31)$$

A mudança de variáveis acima mapeia $\{u, v\} \rightarrow \{x, y\}$ e torna a região de integração mais complicada caso u, v não pertença domínio $[-\infty, +\infty] \times [-\infty, +\infty]$. Isso porque a mudança de variáveis proposta em pelas equações (30) e (31) gira em 45° o quadrado e duplica o comprimento de suas arestas. A figura 14 mostra o caso de um domínio finito para $\{u, v\} \in [-a, +a] \times [-a, +a]$.

Escrito nas novas variáveis, $\sum_n W_n$ e $\sum_n V$ tornam-se respectivamente

$$\sum_n W_n = \sum_n \left[\frac{k}{2} (x_n - x_{n-1})^2 + \frac{k}{2} (y_n - y_{n-1})^2 \right] \quad (32)$$

$$\sum_n V(y_n) = \sum_n D \{ \exp [-\alpha(y_n)] - 1 \}^2. \quad (33)$$

A função de partição pode ser reescrita como

$$\begin{aligned}
Z &= \overbrace{\Lambda' \int \prod_{n=1}^N dx_n \exp \left[-\frac{\beta k}{2} (x_n - x_{n-1})^2 \right]}{=\Gamma} \\
&\times \int \prod_{n=1}^N dy_n \exp \left\{ -\beta \left[\frac{k}{2} (y_n - y_{n-1})^2 + V(y_n) \right] \right\} \\
&= \Gamma \int \prod_{n=1}^N dy_n \exp \left\{ -\frac{\beta}{2} [k(y_n - y_{n-1})^2 + V(y_n) + V(y_{n-1})] \right\}. \quad (34)
\end{aligned}$$

A mudança de variáveis que utilizamos foi importante porque nos permitiu integrar a variável x na equação (34), uma vez que as variáveis x e y ficaram desacopladas. A separação média dos pares de base é dada pela equação

$$\langle y_k \rangle = \frac{1}{Z} \int \prod_{n=1}^N dy_n y_k \exp \left\{ -\frac{\beta}{2} [k(y_n - y_{n-1})^2 + V(y_n) + V(y_{n-1})] \right\}. \quad (35)$$

Para resolver as equações (34) e (35) utilizamos a técnica de integral de transferência (TIT), que consiste em expandir o núcleo (o integrando) da equação (34) em autofunções ortonormalizadas, ou seja, transformamos as equações (34) e (35) em equações de autovetor-autovalor. É importante observar que $\Gamma \rightarrow \infty$ à medida que $x_n \rightarrow \pm\infty$. Isso porque ao longo da reta $x_n = x_{n-1}$ o integrando é constante e vale um, portanto, ao integrar no intervalo infinito, haverá divergência. Vale também ressaltar que não é possível uma mudança de variáveis do tipo

$$s_n = \alpha(x_n - x_{n-1}) \quad (36)$$

Isso porque o Jacobiano seria nulo, contrariando uma condição necessária à mudança de variáveis.

Demonstração:

$$J = \begin{vmatrix} \alpha & -\alpha & 0 & \dots & 0 \\ 0 & \alpha & -\alpha & \dots & 0 \\ & & \vdots & & \\ 0 & \dots & 0 & \alpha & -\alpha \\ -\alpha & 0 & \dots & 0 & \alpha \end{vmatrix}_{N \times N} \quad (37)$$

Vamos usar a regra de Laplace para calcular o determinante acima. Escolhendo a primeira coluna como referência obtemos

$$J = (\alpha) \begin{vmatrix} \alpha & -\alpha & 0 & \dots & 0 \\ 0 & \alpha & -\alpha & \dots & 0 \\ & & \vdots & & \\ 0 & \dots & 0 & \alpha & -\alpha \\ 0 & 0 & \dots & 0 & \alpha \end{vmatrix}_{N-1 \times N-1} + (-\alpha)(-1)^{N+1} \begin{vmatrix} -\alpha & 0 & 0 & \dots & 0 \\ \alpha & -\alpha & 0 & \dots & 0 \\ & & \vdots & & \\ 0 & \dots & \alpha & -\alpha & 0 \\ 0 & 0 & \dots & \alpha & -\alpha \end{vmatrix}_{N-1 \times N-1}$$

Temos agora dois determinantes diagonais, cujos valores são obtidos pela multiplicação dos elementos da diagonal principal, o que dá:

$$J = (\alpha)(\alpha)^{N-1} + (-\alpha)(-1)^{N+1}(-\alpha)^{N-1} \quad (38)$$

$$J = (\alpha)^N + (-1)^{2N+1}(\alpha)^N = 0. \quad (39)$$

Todavia, ao calcular a separação média $\langle y \rangle$, Γ aparece no numerador e no denominador, por isso é cancelado e não interfere nos valores da separação média.

1.5.2 Técnica de Integral de Transferência (TIT)

Para resolver equações como (34) e (35), vamos usar a técnica de integral de transferência (TIT), a qual detalharemos nessa seção. Esse método nos permite resolver, de maneira exata, equações do tipo

$$I = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \prod_{n=1}^N dy_{n+1} K(y_n, y_{n+1}). \quad (40)$$

No caso do modelo PB, o cálculo da função de partição Z envolve integrais como as da equação (40) e para simplificar a notação, vamos definir

$$K(y_n, y_{n-1}) = \exp \left\{ -\frac{\beta}{2} [k(y_n - y_{n-1})^2 + V(y_n) + V(y_{n-1})] \right\} \quad (41)$$

O que torna o método apropriado para resolver equações como (40) é que a complexidade computacional independe do número de integrais. Em nosso caso do estudo de DNA, se estivermos trabalhando com sequências longas, ou seja, N grande, a integração direta fica inviável.

A teoria de equação integral nos permite obter um tratamento matemático às equações integrais do tipo

$$f(x) = g(x) + \lambda \int K(x, y)\phi(y)dy \quad (42)$$

Quando $g(x) = 0$, dizemos que a equação (42) é homogênea, e se, além disso, $f(x) \equiv \phi(x)$, a equação (42) é denominada equação de Fredholm do segundo tipo homogênea, cuja expressão matemática é dada por

$$\phi(x) = \lambda \int K(x, y)\phi(y)dy. \quad (43)$$

Nesse texto vamos focar apenas na equação de Fredholm do segundo tipo homogênea e para maiores detalhes acerca da teoria de Fredholm, indicamos as referências 49–51.

No nosso caso, temos que resolver a seguinte equação integral

$$\int K(y_n, y_{n+1})\phi_i(y_{n+1})dy_{n+1} = \lambda_i\phi_i(y_n), \quad (44)$$

onde $\phi_i(y_n)$ são as autofunções, λ_i são os autovalores, $K(y_n, y_{n+1})$ é o núcleo da equação e y_n e y_{n+1} são variáveis independentes. A equação (44) é equivalente à equação (43), portanto é uma equação integral de Fredholm homogênea do segundo tipo. Para introduzir esse método, devemos iniciar considerando que o núcleo de (40) satisfaz

$$K(y_n, y_{n+1}) = K(y_{n+1}, y_n) \quad (45)$$

e

$$\|K(y_n, y_{n+1})\| = \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dy_n dy_{n+1} \|K(y_n, y_{n+1})\|^2 \right)^{\frac{1}{2}} < +\infty. \quad (46)$$

A condição da equação (45) impõe que o núcleo deve ser simétrico e a condição da equação (46) é necessária para satisfazer o teorema espectral, que será a base do método de solução.

Dado as condições acima, $\exists \{\lambda_i\} > 0$ com $i \in \mathbb{N}$ um conjunto de autovalores e um conjunto completo ortogonal de autofunções $\{\phi_i(y_n)\}$ com as seguintes propriedades:

$$\int_{-\infty}^{+\infty} dy_n \phi_i^\dagger(y_n) \phi_j(y_n) = \delta_{i,j} \quad (47)$$

$$\sum_{i=1}^{+\infty} \phi_i^\dagger(y_n) \phi_i(y_{n+1}) = \delta(y_{n+1} - y_n) \quad (48)$$

$$K(y_n, y_{n+1}) = \sum_{i=1}^{+\infty} \lambda_i \phi_i^\dagger(y_n) \phi_i(y_{n+1}), \quad (49)$$

onde o símbolo \dagger aplicado ao autovetor ϕ indica o transposto conjugado do mesmo.

Substituímos a equação (49) em (40) obtemos

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} dy_1 \int_{-\infty}^{+\infty} dy_2 \dots \int_{-\infty}^{+\infty} dy_N \sum_{i=1}^{+\infty} \lambda_i \phi_i^\dagger(y_1) \phi_i(y_2) \times \sum_{n=1}^{+\infty} \lambda_n \phi_n^\dagger(y_2) \phi_n(y_3) \times \dots \\ &\times \sum_{m=1}^{+\infty} \lambda_m \phi_m^\dagger(y_N) \phi_m(y_1) \sum_{k=1}^{+\infty} \lambda_k \phi_k^\dagger(y_N) \phi_k(y_1) \end{aligned} \quad (50)$$

Na equação (50) usamos a condição de contorno periódica $y_{N+1} = y_1$. Ao realizar a integração na variável y_N na equação (50) obtemos

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} dy_1 \int_{-\infty}^{+\infty} dy_2 \dots \sum_{i=1}^{+\infty} \lambda_i \phi_i^\dagger(y_1) \phi_i(y_2) \times \sum_{n=1}^{+\infty} \lambda_n \phi_n^\dagger(y_2) \phi_n(y_3) \times \dots \\ &\times \sum_{m=1}^{+\infty} \sum_{k=1}^{+\infty} \lambda_k \lambda_m \phi_m^\dagger(y_{N-1}) \phi_k(y_1) \underbrace{\int_{-\infty}^{+\infty} dy_N \phi_m^\dagger(y_N) \phi_k(y_N)}_{\delta_{m,k}}. \end{aligned} \quad (51)$$

Agora podemos calcular a integral na variável y_{N-1} e obter

$$\begin{aligned}
I &= \int_{-\infty}^{+\infty} dy_1 \int_{-\infty}^{+\infty} dy_2 \dots \sum_{i=1}^{+\infty} \lambda_i \phi_i^\dagger(y_1) \phi_i(y_2) \times \sum_{n=1}^{+\infty} \lambda_n \phi_n^\dagger(y_2) \phi_n(y_3) \times \dots \\
&\times \sum_{l=1}^{+\infty} \sum_{m=1}^{+\infty} \lambda_l \lambda_m^2 \phi_l^\dagger(y_{N-2}) \phi_m(y_1) \underbrace{\int_{-\infty}^{+\infty} dy_{N-1} \phi_l^\dagger(y_{N-1}) \phi_m(y_{N-1})}_{\delta_{l,m}}. \quad (52)
\end{aligned}$$

Prosseguimos integrando nas outras variáveis de maneira análoga ao realizado anteriormente.

Feito isso obtemos

$$I = \sum_{i=1}^{+\infty} \lambda_i^N \int_{-\infty}^{+\infty} dy_1 \|\phi_i(y_1)\|^2. \quad (53)$$

Estamos supondo que todas as autofunções estão normalizadas e portanto obtemos

$$I = \sum_{i=1}^{+\infty} \lambda_i^N. \quad (54)$$

Sendo assim, duas condições necessárias para que (54) seja convergente é

$$\lim_{i \rightarrow +\infty} \lambda_i^N = 0 \quad (55)$$

e

$$0 \leq |\lambda_{i+1}^N| \leq |\lambda_i^N|. \quad (56)$$

Se associarmos o núcleo $K(y_n, y_{n-1})$ ao integrando da equação (34), podemos reescrevê-la como

$$Z = \Gamma \sum_{i=1}^{+\infty} \lambda_i^N \quad (57)$$

Para escrever as relações acima podemos utilizar a notação de Dirac, definindo

$$\langle i|j \rangle = \int_{-\infty}^{+\infty} dy \phi_i^\dagger(y) \phi_j(y). \quad (58)$$

Então

$$\langle y_m \rangle = \frac{1}{Z} \sum_{i=1}^{+\infty} \cdots \sum_{l=1}^{+\infty} \sum_{k=1}^{+\infty} \sum_{j=1}^{+\infty} \cdots \sum_{p=1}^{+\infty} \lambda_i \cdots \lambda_p \underbrace{\langle i|p \rangle}_{\delta_{i,p}} \cdots \langle l|y_m|k \rangle \underbrace{\langle k|j \rangle}_{\delta_{k,j}} \cdots, \quad (59)$$

o que resulta em

$$\langle y_m \rangle = \frac{1}{Z} \sum_{i=1}^{+\infty} \lambda_i^N \langle i|y_m|i \rangle. \quad (60)$$

É importante perceber que a condição de contorno periódica leva a uma invariância translacional, ou seja, $\langle y_m \rangle$ é independente do sítio m , então a equação (60) pode ser escrita como

$$\langle y \rangle = \frac{1}{Z} \sum_{i=1}^{+\infty} \lambda_i^N \langle i|y|i \rangle. \quad (61)$$

O cálculo dos autovalores e autofunções não dependem de N , portanto a técnica torna-se vantajosa para sequências grandes.

1.5.3 Resultados obtido pelo Modelo Peyrard Bishop e algumas de suas variações

Ao reproduzir a curva de desnaturação $\langle y \rangle \times T$ do DNA prevista pelo modelo Peyrard-Bishop, observamos que o perfil da desnaturação é bem menos abrupto que o obtido experimentalmente via espectroscopia UV. Então propostas para modificação do potencial surgiram em diversos trabalhos, sendo que uma delas está descrita em [48], onde os autores sugerem a substituição do potencial harmônico por um potencial anarmônico, cuja expressão matemática é

$$W(y_n, y_{n-1}) = \frac{k}{2} (1 + \rho \exp^{-\alpha(y_n + y_{n-1})}) (y_n + y_{n-1})^2. \quad (62)$$

Este potencial tem a propriedade de diminuir a constante elástica da ligação de empilhamento quando há o desparelhamento. Esta modificação mostrou-se eficaz e prevê um perfil de desnaturação semelhante ao que é obtido experimentalmente, como mostrado na figura 15.

O cálculo analítico de $\langle y \rangle$ obtido a partir da equação (61) é feito integrando as coordenadas espaciais no intervalo $]-\infty, +\infty[$ e quando realizamos a integração numérica, discretizamos a

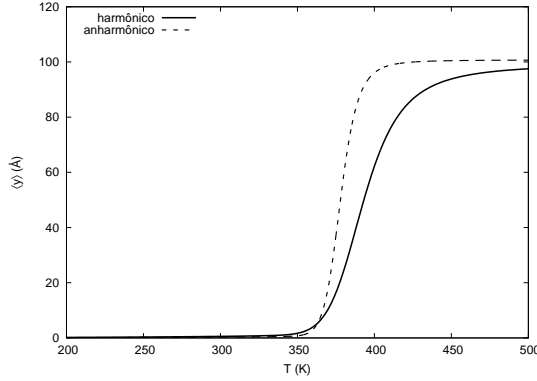


Figura 15

A figura mostra como o modelo Peyrard-Bishop prevê a dependência de $\langle y \rangle$ com temperatura T usando os potenciais harmônico e anarmônico para representar o empilhamento. Parâmetros: $D = 0.04$ eV, $k = 0.04$ eVÅ⁻¹, $a = 4.45$ Å⁻¹, $\alpha = 0.35$ Å⁻¹, $\rho = 0.5$ e $\theta = 0.01$ rad.

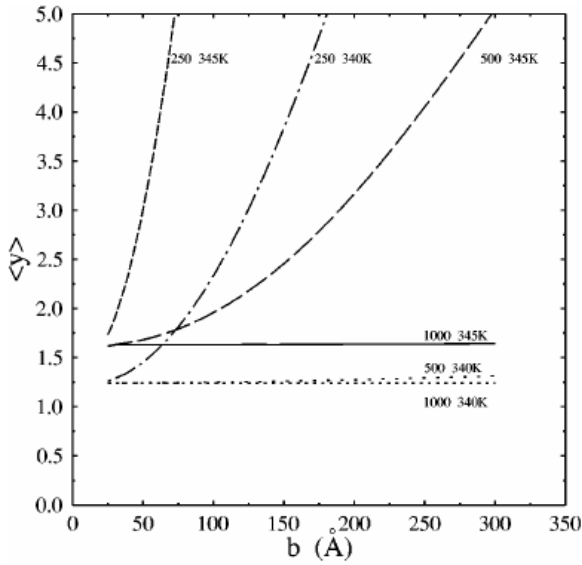


Figura 16

A figura mostra a dependência de $\langle y \rangle$ com o limite superior de integração b . A conclusão é que, embora fatores como temperatura e número de bases N interferem no valor de $\langle y \rangle$, se $b \rightarrow +\infty$, $\langle y \rangle$ parece divergir. Figura retirada da referência [52].

equação (61) e truncamos os limites de integração em $[a, b]$. Embora a proposta feita em [48] nos permite obter uma melhor previsão da temperatura de desnaturação que aquela obtida em [38], o modelo apresenta uma divergência de $\langle y \rangle$ quando $b \rightarrow \infty$ [52], como mostra a figura 16. A divergência de $\langle y \rangle$ devido ao intervalo de integração introduz como consequência direta o fato de não podermos usar o intervalo infinito e com isso a região de integração será um quadrado girado de 45° (veja a figura 14), o que dificulta bastante a integração.

Uma das maneiras de contornar a divergência é introduzir um pequeno ângulo de torção [41] no potencial de empilhamento, que passa ter a forma

$$W_n = \frac{k}{2} (y_n^2 - 2y_n y_{n-1} \cos(\theta) + y_{n-1}^2). \quad (63)$$

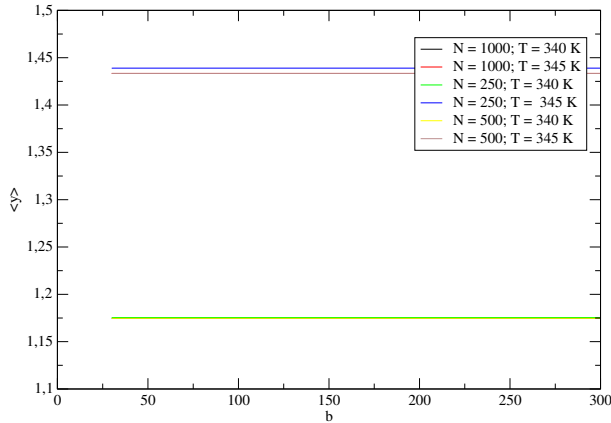


Figura 17

A figura mostra a separação média em função do limite superior da integração (b) para sequências de diferentes tamanhos (N) e com três valores de temperatura (T). Podemos observar que a inclusão de um ângulo de torção $\theta = 0.01$ (rad) elimina a dependência de $\langle y \rangle$ com b ($\times 10^{-1}$ nm) em todos os casos. Foram usados o potencial de Morse e o potencial anarmônico para o empilhamento.

Ao introduzir tal ângulo, $\langle y \rangle$ não diverge mais ao aumentar o intervalo de integração independente do número de pares de bases, como mostrado na figura 17. É importante destacar que tamanho da sequência do DNA pode alterar a temperatura de desnaturação, como mostrado em [40].

Podemos facilmente entender o porquê do ângulo eliminar a divergência se expandirmos o termo $\cos(\theta)$ até segunda ordem. Fazendo isso obtemos:

$$\cos(\theta) \approx 1 - \frac{\theta^2}{2} \quad (64)$$

substituindo a equação (64) em (63) obtemos

$$W_n = \frac{k}{2} [(y_n - y_{n-1})^2 + y_n y_{n-1} \theta^2]. \quad (65)$$

Então o integrando da equação (34) torna-se:

$$Z \approx \Gamma \int \prod_{n=1}^N dy_n \exp \left\{ -\frac{\beta}{2} [k(y_n - y_{n-1})^2 + V(y_n) + V(y_{n-1})] \right\} \exp \left(-\frac{\beta k}{2} y_n y_{n-1} \theta^2 \right) \quad (66)$$

O termo $\exp \left(-\frac{\beta k}{2} y_n y_{n-1} \theta^2 \right)$ faz com que o integrando da equação (66) aproxime de zero quando $\{y_n, y_{n-1}\} \rightarrow +\infty$, conforme a figura 18. A integração nas variáveis $\{y_n, y_{n-1}\}$ significa obter o volume abaixo da superfície $K(y_n, y_{n-1})$ e acima do plano $\{x, y, 0\}$. Se desconsiderarmos a torção ($\theta = 0$), o volume diverge, como podemos observar na figura 18 da esquerda, diferente de quando incluímos a torção, conforme mostra a figura 18 da direita.

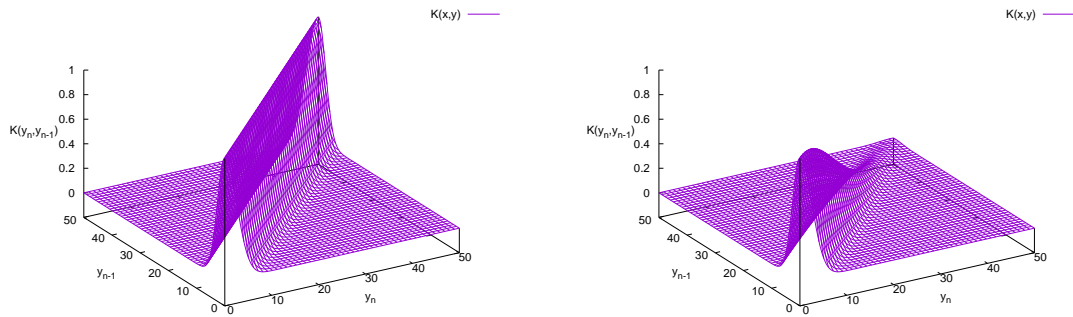


Figura 18

A figura da esquerda mostra o núcleo da função partição do modelo PB original, já a da direita mostra o integrando modificado pela equação (63).

Para analisar o comportamento da função de partição calculada com TIT, basicamente devemos entender o comportamento dos autovalores, já que $Z = \sum_{i=1}^{+\infty} \lambda_i^N$. Como vamos discretizar o núcleo para fazer a integração numérica, o número de autovalores irá depender da quantidade de passos utilizado na integração, que chamaremos de L . Dessa forma, a função de partição é calculada por $Z = \sum_{i=1}^L \lambda_i^N$.

Calculamos os primeiros autovalores do modelo PB em função do limite superior de integração, mantendo fixo as demais constantes. O resultado está na figura 19. A condição expressa na equação (55) não é satisfeita quando há degenerescência como na figura (19) da esquerda, pois nesse caso o termo geral não decresce, o que é uma condição necessária para a convergência da série. Então a função de partição irá divergir, assim como $\langle y \rangle$.

1.5.4 Modelo Peyrard-Bishop para sequência heterogênea

Para sequências heterogêneas, a intensidade das ligações varia com o tipo de bases envolvidas. As bases pirimídicas (timina e citosina) fazem duas ligações de hidrogênio, já as purinas (guanina e adenina) se conectam por um conjunto de três ligações de hidrogênio. Além disso, a ligação de empilhamento também é dependente do tipo de base. As diferentes intensidades se manifestam nos potenciais através das constantes, por exemplo, o parâmetro D do potencial de Morse descrito na equação (26), página 28, possui maior valor para pares de purinas do que para as pirimidinas. Embora o estudo de sequências homogêneas, isto é, sequências

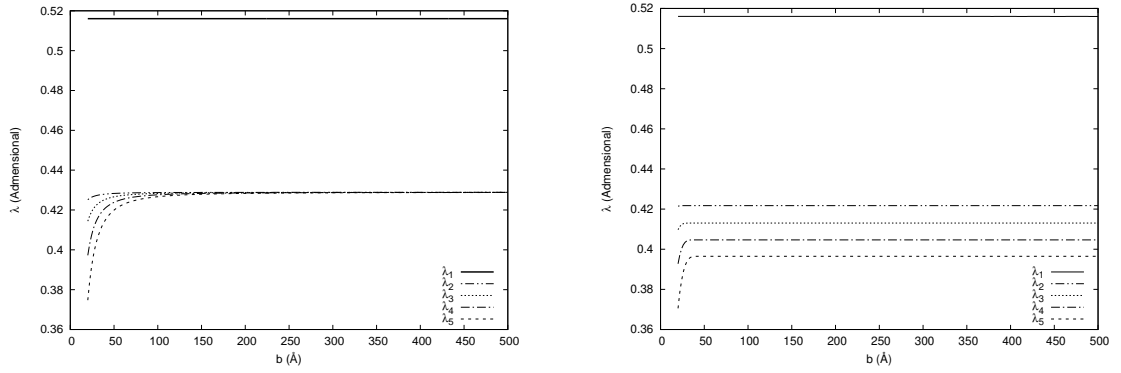


Figura 19

Os cinco primeiros autovalores em função do limite superior de integração (b) para $T = 350$ K. A figura da esquerda foi calculada a partir do modelo PB original e a da direita usando um ângulo ($\theta = 0.1$).

composta apenas de A-T ou somente C-G, seja importante para entender as propriedades do modelo (os cálculos são mais simples), na realidade raramente encontramos sequências de DNA homogêneo em aplicações práticas. Então, para que o modelo possa ter aplicação em problemas reais, é necessário que ele seja estendido à sequências heterogêneas e nesse caso, o integrando da função de partição é mais complicado e as integrações ficam mais complexas para resolver, pois, nesse caso, as constantes dos potenciais dependem do tipo de par de base. Assim a equação (41), página 33, torna-se

$$K^{(n,n+1)}(y_n, y_{n-1}) = \exp \left\{ -\frac{\beta}{2} [W_n(y_n, y_{n-1})^2 + V_n(y_n) + V_{n+1}(y_{n-1})] \right\}. \quad (67)$$

Como os potenciais dependem do sítio n , a equação (61) não pode mais ser obtida a partir da equação (67), uma vez que muitas simplificações feitas para obter a equação (61) já não se aplicam mais. Todavia, Zhang et al. [52] propuseram expandir a equação (67) em autofunções do núcleo de sequências homogênea. Assim a equação (67) fica

$$K^{(n,n+1)}(y_n, y_{n-1}) = \sum_{j,k=1}^{\infty} C_{j,k}^{(n,n+1)} \phi_j(y_n) \phi_k(y_{n-1}) \quad (68)$$

e $C_{j,k}^{(i,i+1)}$ é obtido fazendo

$$C_{j,k}^{(n,n+1)} = \int \int dy_n dy_{n-1} K^{(n,n+1)}(y_n, y_{n-1}) \phi_j(y_n) \phi_k(y_{n-1}). \quad (69)$$

A função de partição passa a ser um produto de matrizes

$$Z = Tr (C^{(1,2)} C^{(2,3)} \dots C^{(N,1)}), \quad (70)$$

onde as matrizes C são

$$C_{j,k}^{(n,n+1)} = [C_{j,k}^{(n,n+1)}]. \quad (71)$$

De maneira análoga, o cálculo de $\langle y_m \rangle$ fica

$$\langle y_m \rangle = \frac{1}{Z} Tr (C^{(1,2)} C^{(2,3)} \dots C^{(n-1,n)} Y C^{(n,n+1)} \dots C^{(N,1)}) \quad (72)$$

e os elementos da matriz Y são obtidos fazendo

$$Y_{jk} = \int \phi_j(y_n) y_n \phi_k(y_n). \quad (73)$$

Os elementos das matrizes C dependem apenas dos potenciais, que por sua vez dependem dos tipos de bases envolvidas nas ligações de primeiros vizinhos. Como temos 4 tipos de bases diferentes e portanto 4 possíveis vizinhos, o total de matrizes C diferentes são dezesseis, mas considerações de simetria reduzem este número para dez que são

$$\begin{aligned} & C^{CG,GC} & C^{TA,AT} \\ C^{AT,AT} &= (C^{TA,TA})^\dagger & C^{CG,AT} &= (C^{TA,GC})^\dagger \\ C^{AT,CG} &= (C^{GC,TA})^\dagger & C^{AT,GC} &= (C^{CG,TA})^\dagger \\ C^{GC,AT} &= (C^{TA,GC})^\dagger & C^{CG,CG} &= (C^{GC,GC})^\dagger, \end{aligned}$$

onde \dagger denota a matriz adjunta. É importante observar que as matrizes C não dependem do número de bases (N) e nem do ordenamento na sequência, e por isso só precisam ser

calculadas uma única vez. Ao calcular $\langle y_m \rangle$ de sequências diferentes, o que muda em relação as matrizes C são a quantidade e o ordenamento na equação (72).

O método para obter as matrizes C a partir da expansão de uma sequência homogênea consiste no seguinte: suponhamos que nossa sequência homogênea seja formada por CG. Então, neste caso, a matriz $C^{CG,GC}$ será diagonal e os elementos são os autovalores λ_i da equação (34). Vamos escrever as outras matrizes como a soma de uma matriz diagonal Λ e uma outra matriz não-diagonal $\Delta^{(n,n-1)}$

$$C^{(n,n-1)} = \Lambda + \Delta^{(n,n-1)}. \quad (74)$$

As equações (70) e (72) tomam as seguintes formas

$$Z = Tr [(\Lambda + \Delta^{(1,2)})(\Lambda + \Delta^{(2,3)}) \dots (\Lambda + \Delta^{(N,1)})]. \quad (75)$$

Realizando o produto matricial na equação (75) obtemos uma soma de $N + 1$ termos em ordem de Λ^ω

$$Z = \sum_{\omega=0}^N Z_\omega(\Lambda) = \sum_{\omega=0}^N Tr [\mathcal{O}(\Lambda^\omega)], \quad (76)$$

onde $\mathcal{O}(\Lambda^\omega)$ são os termos da expansão que contém ω multiplicações da matriz Λ . No caso de sequências homogêneas CG, as matrizes $\Delta^{(n,n-1)}$ seriam nulas e

$$Z = \sum_{\omega=0}^N Tr [\mathcal{O}(\Lambda^\omega)] = \sum_{i=1}^L \Lambda_i^N, \text{ neste caso, recuperamos a equação (57)}. \quad (77)$$

Na figura 20 mostramos um exemplo da expansão da equação (77), onde podemos notar como a curva muda e, em particular, o valor de ω que maximiza Z_ω (ω_{\max}) depende da estrutura da sequência.

Embora Z tenha forte dependência da temperatura, observou-se que os valores de ω que maximizam as curvas $Z(\Lambda_\omega)$ pouco são influenciados pela temperatura, como mostra a figura 21 [53], além disso, podemos perceber na figura que a grandeza ω_{\max} possui informações sobre o tipo da sequência.

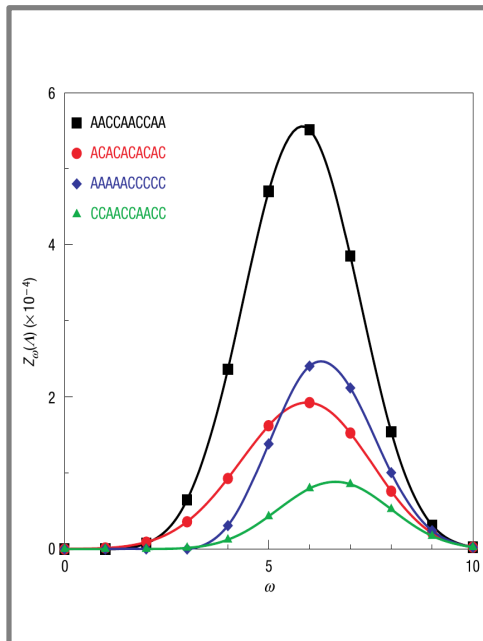


Figura 20

Componentes $Z_{\omega(\Lambda)}$, equação (77), para vários tipos de sequência. Note como o máximo da curva aparece para ω depende da ordem das bases. Figura retirada da referência [53].

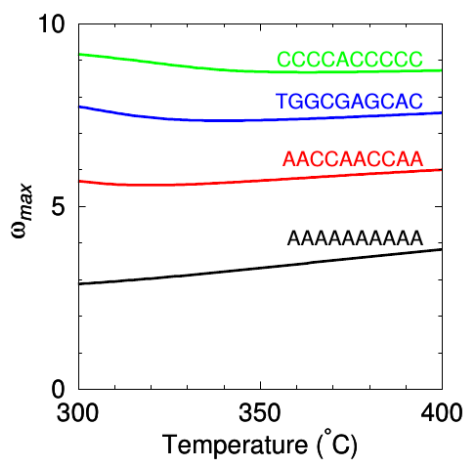


Figura 21

Notamos que os valores de ω_{max} praticamente não sofrem influência da temperatura, todavia o tipo de bases que compõem a sequência os modifica consideravelmente. Figura retirada da referência [53].

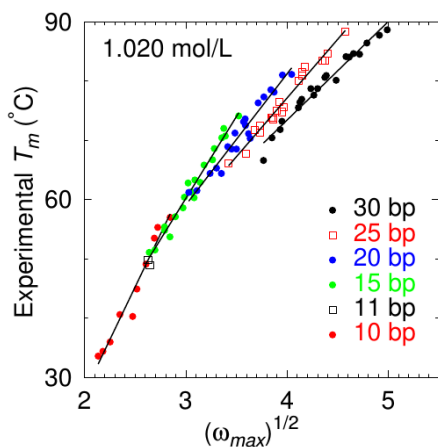


Figura 22

A figura mostra a correlação das temperaturas medidas experimentalmente, T_m , com ω_{\max} . Figura retirada da referência [53].

A figura 22 mostra a regressão linear da temperatura de um conjunto de dados experimentais [53] em função de $\sqrt{\omega_{\max}}$. A reta possui a forma

$$T_m = \beta_0(N) + \beta_1(N)\tau, \quad (78)$$

onde

$$\sqrt{\omega_{\max}} = \tau. \quad (79)$$

Devido a relação linear entre T_m e τ , este parâmetro pode ser usado como um medidor de “equivalência termodinâmica” e por isso recebeu o nome “índice de desnaturação”. Para qualquer outra sequência de DNA, basta calcular o seu respectivo ω_{\max} e aplicar a equação (78).

A partir da grandeza τ , Weber et al. [53] inferiu uma método que permite otimizar os parâmetros do modelo PB a um custo computacional relativamente baixo.

O método consiste em variar os parâmetros do hamiltoniano até encontrar o conjunto de parâmetros que resulte na menor diferença entre as temperaturas de desnaturação teóricas e experimentais encontradas na literatura. Isto é feito minimizando a quantidade

$$\chi = \sum_{i=1}^N (T'_i - T_i)^2. \quad (80)$$

Uma vez encontrado o “melhor” conjunto de parâmetros, podemos prever a temperatura de desnaturação de uma sequência qualquer com a melhor precisão possível para o modelo. Além disso, o método de minimização nos permite obter informações intramoleculares, uma vez que os parâmetros dos potenciais estão relacionados com a natureza das interações entre os componentes do DNA. Para exemplificar isto, em [47] os autores mostraram que as ligações de hidrogênio entre AU são mais fortes que entre AT, fato que concorda com os dados experimentais descritos na referência [54].

2 Modelo Peyrard-Bishop tridimensional

Os modelos bidimensionais, embora façam previsões que concordam com os dados experimentais, possui, intrinsecamente, inúmeras diferenças quando os comparamos com a geometria tridimensional da molécula de DNA. A diferença mais fundamental é que, nos casos unidimensionais, não se leva em conta a torção natural da dupla hélice e especificamente no modelo Peyrard-Bishop, a distância do empilhamento também é negligenciada. Sendo assim, modelos mais sofisticados e realistas que o bidimensional foram propostos com o intuito de obter informações mais precisas e condizentes com os dados obtidos através das experiências que utilizam técnicas de difração de raios-X [1] ou NMR (Ressonância magnética nuclear) [55].

Uma versão do modelo Peyrard-Bishop tridimensional foi proposto por Cocco e Monas-son [56], baseado na geometria da figura 23. Nesse trabalho, os autores descrevem o DNA como uma estrutura helicoidal com três graus de liberdade: radial, angular e longitudinal. O movimento radial ocorre de maneira simétrica, ou seja, os nucleotídeos complementares deslocam-se da mesma forma em relação ao centro de massa do par e, além disso, o movimento só é permitido no plano perpendicular ao eixo central da dupla fita, o que elimina qualquer curvatura da molécula. Na direção longitudinal, considera a torção no intervalo $(0, \pi)$ e limita-se o deslocamento máximo ao valor de uma constante L . Essa última consideração possibilita escrever as variáveis longitudinais em termos da constante e das variáveis radiais e angulares, o que, devido a escolha dos potenciais, elimina a integração na direção z , conforme veremos mais adiante. As condições de contorno usadas foram as seguintes: o primeiro par de base do duplexo possui coordenadas $(r_1, \phi_1) \equiv (R, 0)$, ou seja, essa "ponta" é fixa; Já a extremidade oposta possui coordenadas $(r_N, \phi_N) \equiv (R, \phi_N)$, portanto é possível ter movimento angular (é importante ressaltar que, devido à simetria do movimento radial entre as os nucleotídeos complementares, as coordenadas de apenas um dos nucleotídeos é suficiente para descrever o par).

Para descrever as interações, os autores usaram os potenciais de morse $V_m(r_n)$, harmônico

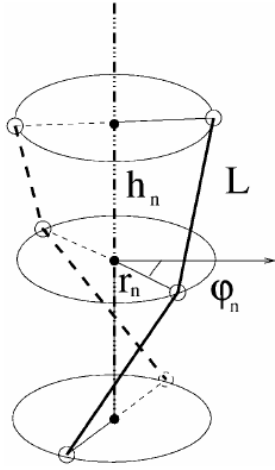


Figura 23

A figura mostra um esquema proposto para o modelo tridimensional para a estrutura da dupla hélice do DNA. Figura retirada da referência [56].

$V_b(r_n, r_{n-1}, \theta_n)$, anarmônico $V_s(r_n, r_{n-1})$ e de torção $V_\Gamma(\theta_n)$, cujas expressões são

$$V_m = D (\exp [a (r_n - R)] - 1)^2 \quad (81)$$

$$V_b = K \left(\sqrt{L^2 - r_n^2 - r_{n-1}^2 + 2r_n r_{n-1} \cos(\theta_n)} - H \right)^2 \quad (82)$$

$$V_s = E (r_n - r_{n-1})^2 \exp [-b(r_n + r_{n-1} - 2R)] \quad (83)$$

$$V_\Gamma = -\Gamma \theta_n, \quad (84)$$

onde $\theta_n = \phi_n - \phi_{n-1}$. A função de partição é obtida da seguinte forma:

$$Z = \int_{-\infty}^{+\infty} \prod_{n=1}^N dr_n d\phi_n \overbrace{\sqrt{r_n r_{n+1}} \exp \left\{ -\beta \left[\frac{V_m(r_n)}{2} + \frac{V_m(r_{n-1})}{2} + V_s(r_n, r_{n-1}) \right] \right\}}^{=X(r_n, r_{n-1})} \times \exp \{ -\beta [V_b(r_n, r_{n-1}, \theta_n) + V_\Gamma(\theta_n)] \} \chi(\theta_n), \quad (85)$$

onde $\chi(\theta_n)$ é definida como

$$\chi(\theta_n) = \begin{cases} 1 & \text{se } \phi_n - \phi_{n-1} \leq \pi \\ 0 & \text{se } \phi_n - \phi_{n-1} > \pi \end{cases} \quad (86)$$

A ideia agora é implementar o formalismo da técnica de integral de transferência (TIT) e para isso, os autores definiram a seguinte operação:

$$\langle r_n, \phi_n | T | r_{n-1}, \phi_{n-1} \rangle = X(r_n, r_{n-1}) \exp \{ -\beta [V_b(r_n, r_{n-1}, \theta_n) + V_\Gamma(\theta_n)] \} \chi(\theta_n) \quad (87)$$

substituindo a equação (88) em (85) obtemos:

$$Z = \int_{-\infty}^{+\infty} \prod_{n=1}^N dr_n d\phi_n \langle r_n, \phi_n | T | r_{n-1}, \phi_{n-1} \rangle \quad (88)$$

Usando as condições de contorno $\phi_1 = 0$ e $r_1 = r_N = R$ a equação (88) pode ser reescrita da seguinte forma:

$$\begin{aligned} Z &= \int_{-\infty}^{+\infty} d\phi_N dr_{N-1} d\phi_{N-1} \cdots dr_2 d\phi_2 \langle R, \phi_N | T | r_{N-1}, \phi_{N-1} \rangle \langle r_{N-1}, \phi_{N-1} | T | r_{N-2}, \phi_{N-2} \rangle \\ &\quad \vdots \\ &\quad \times \langle r_3, \phi_3 | T | r_2, \phi_2 \rangle \langle r_2, \phi_2 | T | R, 0 \rangle \\ &= \langle R, \phi_N | T^N | R, 0 \rangle \end{aligned} \quad (89)$$

A parte angular da matriz de transferência T possui invariância translacional, podemos diagonalizá-la através de uma transformada de Fourier definindo a seguinte equação:

$$T_k(r_n, r_{n-1}) = X(r_n, r_{n-1}) Y_k(r_n, r_{n-1}) \quad (90)$$

com

$$\begin{aligned} Y_k(r_n, r_{n-1}) &= \int_{-\infty}^{+\infty} d\phi_n d\phi_{n-1} \exp[-\beta V_b(r_n, r_{n-1}, \phi_n, \phi_{n-1}) + V_\Gamma(\phi_n - \phi_{n-1})] \\ &\quad \times \exp[-ik(\phi_n + \phi_{n-1})] \chi(\theta_n) \\ &= \int_0^\pi d\theta_n \exp[-\beta V_b(r_n, r_{n-1}, \theta_n) + V_\Gamma(\theta_n)] \exp(-ik\theta_n) \end{aligned} \quad (91)$$

Se $k = 0$, a equação (91) é idêntica a parte angular da equação (85). Então somente os modos com $k = 0$ devem contribuir com o cálculo de Z . Denotando os autovalores e autovetores do operador T_0 respectivamente por $\lambda_q^{(\Gamma)}$ e $\psi_q(r)^{(\Gamma)}$, com $\lambda_q^{(\Gamma)} \geq \lambda_{q-1}^{(\Gamma)}$, podemos escrever a densidade de energia livre quando $N \rightarrow \infty$ como

$$f^{(\Gamma)} = -k_B T \ln(\lambda_0^{(\Gamma)}). \quad (92)$$

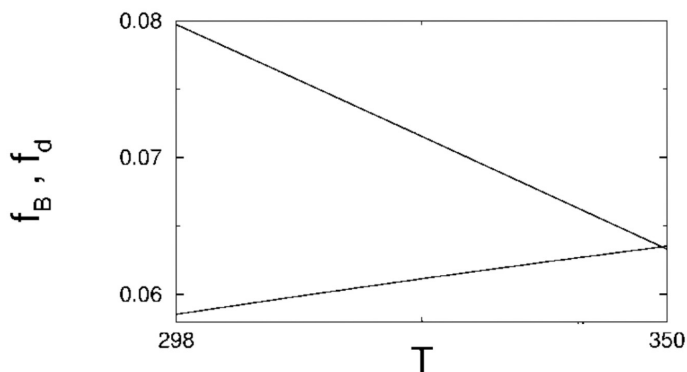


Figura 24

Figura retirada de [56]. A figura mostra a convergência entre as energia livre dos estados aberto (f_a) e ligado (f_B) em função da temperatura. Nesse caso temos $T_m = 350$ k.

A diferença ΔG entre a energia livre f^F de um estado aberto ($q = 1$) e um estado ligado ($q = 0$) nos dá energia livre de desnaturação para cada temperatura (T). Assim, quando $\Delta G \rightarrow 0$, obtemos T_m (veja a figura 24).

O tratamento matemático usando técnica de integral de transferência (TIT) possui a vantagem do baixo custo computacional ao tratar moléculas grandes, mas o estudo dos modelos mesoscópicos não estão restritos à TIT. Uma possibilidade é a dinâmica de Langevin [36, 57–59]. Nessa descrição, acrescenta-se à equação de movimento determinística um termo estocástico. De forma genérica, a equação que rege a dinâmica de Langevin para o DNA é dada por

$$\zeta \frac{d\vec{r}_i}{dt} = -\vec{\nabla} H_{r_i}(\{\vec{r}_i\}) + \vec{\xi}_i(t). \quad (93)$$

O termo $\vec{\xi}_i(t)$ representa a contribuição estocástica na equação de movimento (93) e a esse termo, atribui-se interações com solventes, viscosidade, etc. A dificuldade do tratamento via equação (93) é devido ao tempo computacional elevado, pois no caso tridimensional, para cada par de base, teremos doze equações diferenciais acopladas para resolver. Outra possibilidade é o estudo através de dinâmica molecular usando apenas forças determinísticas [60, 61]. Da mesma maneira que a dinâmica de Langevin, o maior problema desse tratamento é o custo computacional elevado em caso de moléculas com elevado número de nucleotídeos. A principal vantagem dos tratamentos que utilizam Dinâmica Molecular é “resolução” de detalhes,

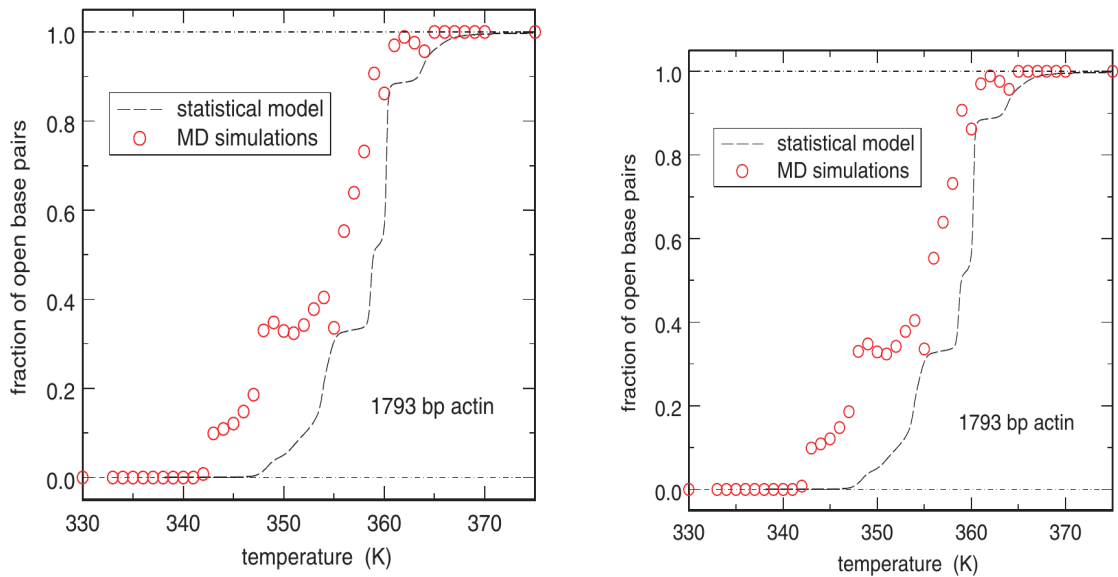


Figura 25

A figura da esquerda mostra a evolução temporal do número de pares de bases abertas em função do tempo para duas temperaturas fixas $T = 335$ K e $T = 340$ K e a da direita traz uma comparação dos cálculos usando dinâmica molecular e modelo estatístico. Figuras retiradas de [36].

o que muitas vezes não pode ser obtido pelo tratamento estatístico, como por exemplo, a evolução das grandezas físicas com o tempo de forma direta, como mostrado na figura 25. No caso do estudo via TIT, os potenciais devem ser mais restritos na sua forma funcional como por exemplo serem simétricos, mas vamos detalhar isso mais adiante ao discutirmos a TIT.

2.1 Desenvolvimento Matemático do cálculo de Z e $\langle r \rangle$ para a versão tridimensional

A função de partição clássica de um sistema tridimensional em coordenadas cartesianas pode, em geral, ser escrita como um produto de duas partes: Uma dos momentos e outra das coordenadas espaciais. Isso porque no hamiltoniano, os momentos ficam desacoplados das variáveis

espaciais, conforme a equação (94).

$$H(\vec{P}_n, \vec{r}_n) = \frac{\vec{P}_n \cdot \vec{P}_n}{2m} + F(\vec{r}_n, \vec{r}_{n-1}), \quad (94)$$

com $F(\vec{r}_n, \vec{r}_{n-1}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ a função que representa as energias potenciais do sistema. Denominaremos Z_T para referirmos à função de partição completa, cujo cálculo é dado pela equação (95).

$$Z_T = \frac{1}{h^{3N}} \int_{\mathbb{R}^3} \prod_{n=1}^N d^3 r_n d^3 P_n \exp[-\beta H(\vec{P}_n, \vec{r}_n)], \quad (95)$$

onde h é a constante de Planck, cuja finalidade é deixar Z adimensional. A equação (95) pode ser reescrita como

$$\begin{aligned} Z_T &= \frac{1}{h^{3N}} \int_{\mathbb{R}^3} \prod_{n=1}^N d^3 P_n \exp \left[-\beta \frac{\vec{P}_n \cdot \vec{P}_n}{2m} \right] \\ &\times \int_{\mathbb{R}^3} \prod_{n=1}^N d^3 r_n \exp \{ -\beta F(\vec{r}_n, \vec{r}_{n-1}) \}. \end{aligned} \quad (96)$$

Para simplificar a notação, definiremos Z_P para designar a partes dos momentos e Z para a contribuição das coordenadas no calculo da função de partição. Assim,

$$Z_T = \frac{1}{h^{3N}} (Z_P \cdot Z). \quad (97)$$

A integral dos momentos Z_P pode ser separada da integração espacial, uma vez que não possui dependência de nenhuma coordenada. Sendo assim, a contribuição Z_P é dada pela equação (98).

$$\begin{aligned} Z_P &= \int_{\mathbb{R}^3} \prod_{n=1}^N d^3 P_n \exp \left[-\beta \frac{\vec{P}_n \cdot \vec{P}_n}{2m} \right] \\ &= \prod_{n=1}^N \int_{\mathbb{R}^3} d^3 P_n \exp \left[-\beta \frac{|\vec{P}_n|^2}{2m} \right]. \end{aligned} \quad (98)$$

Em coordenadas cartesianas temos $|\vec{P}_n|^2 = p_x^2 + p_y^2 + p_z^2$, onde omitimos o índice n de cada componente por simplificação de notação. Então podemos reescrever a equação (98) como

$$Z_P = \prod_{n=1}^N \left(\int_{-\infty}^{+\infty} dP_x \exp \left[-\beta \frac{p_x^2}{2m} \right] \right)^3. \quad (99)$$

Resolver a equação (99) consiste em integrar a função Gaussiana e, para isso, usaremos a conhecida relação matemática para executar a integração:

$$\int_{-\infty}^{+\infty} dx \exp(-\alpha x^2) = \sqrt{\frac{\pi}{\alpha}}. \quad (100)$$

Na equação (99), a constante α é equivalente à $\frac{\beta}{2m}$, logo temos:

$$\begin{aligned} Z_P &= \prod_{n=1}^N (2\pi m k_B T)^{\frac{3}{2}} \\ &= (2\pi m k_B T)^{\frac{3N}{2}}. \end{aligned} \quad (101)$$

Podemos reescrever a equação (96) como

$$Z_T = \Gamma \int_{\mathbb{R}^3} \prod_{n=1}^N d^3 r_n \exp\{-\beta F(\vec{r}_n, \vec{r}_{n-1})\} \quad (102)$$

ou

$$Z_T = \Gamma Z, \quad (103)$$

onde

$$\Gamma = \left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3N}{2}}. \quad (104)$$

Para calcularmos $\langle r_k \rangle$, utilizamos

$$\langle r_k \rangle = \frac{1}{Z} \int_{\mathbb{R}^3} \prod_{n=1}^N d^3 r_n r_k \exp\{-\beta F(\vec{r}_n, \vec{r}_{n-1})\}, \quad (105)$$

com $k \in \mathbb{N}$ e $1 \leq k \leq N$.

Como vimos anteriormente, há um leque extenso de modelos que fornecem informações de grandezas físicas variadas e que em todos eles há pontos favoráveis e desfavoráveis. A escolha deve ser baseada no tipo de fenômeno que se deseja estudar, pois em regra geral, aqueles que oferecem mais detalhes exigem um alto custo computacional e vice-versa. Para se ter uma ideia do tempo gasto, o cálculo direto pela quadratura de Gauss Legendre de uma sequência de dois pares de base já foi necessário calcular em paralelo para cada temperatura e demorou cerca de um dia, já usando TIT (sem ser em paralelo), o tempo fica próximo de dez minutos.

A partir do modelo Peyrard-Bishop bidimensional podemos obter bons resultados para fenômeno como a desnaturação térmica, todavia ele tem como ponto de partida um modelo planar, o que é muito diferente da realidade da dupla hélice. As versões tridimensionais são bem mais realistas e também oferecem excelentes previsões para a temperatura de desnaturação, embora inúmeras aproximações são impostas afim de simplificar as equações a ponto de ser possível obter as soluções.

3 Modelo Peyrard-Bishop tridimensional com três graus de liberdade

O objetivo do nosso trabalho foi propôr um modelo mesoscópico capaz de descrever a dupla hélice de maneira mais realística que os modelos unidimensionais usuais e que ao mesmo tempo tivesse um tratamento numérico viável para moléculas grandes. Então propomos simplificações que nos permitiu integrar analiticamente as variáveis longitudinais z e angulares θ , o que reduziu o problema a uma versão unidimensional. Isso nos possibilitou utilizar a TIT para calcular a função de partição (Z) e a separação média $\langle r \rangle$ entre cada par de base.

Nesta secção vamos apresentar nosso modelo, discutir o desenvolvimento matemático utilizado e os resultados qualitativos obtidos para dupla hélice homogênea.

A estrutura geométrica que utilizamos foi baseada no modelo tridimensional proposto por Cocco e Monasson [56] que discutimos com mais detalhes anteriormente na secção 2. Nosso objetivo é considerar os três graus de liberdade para cada par de nucleotídeo, conforme a figura 26, e obter $\langle r \rangle$ e a função de partição usando TIT. Para isso, é necessário integrar as variáveis z e θ analiticamente, uma vez que, para usar TIT, devemos ficar com apenas uma variável. O modelo tridimensional introduz diferenças significativas ao modelo unidimensional, sendo que a mais direta é a possibilidade de movimento nas três dimensões. Outra importante diferença é que consideramos uma distância de equilíbrio diferente de zero entre os nucleotídeos, ou seja, no modelo 3D a dupla fita tem representação realística para seu comprimento e espessura à temperaturas muito baixas.

Em nosso modelo, consideramos dois tipos de interação entre os nucleotídeos: o empilhamento e a interação entre as fitas, os quais são representados por potenciais que descrevem o comportamento esperado de cada tipo de interação, seguindo a definição

Definição 4. Seja $F(\vec{r}_n, \vec{r}_{n-1}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ a parte da hamiltoniana que representa esses potenciais.

Assim como no modelo PB-1D [38], usamos o potencial harmônico $W(\vec{r}_n, \vec{r}_{n-1})$ para descrever o empilhamento e o potencial de Morse $V(\vec{r}_n)$ para simular a ligação de hidrogênio

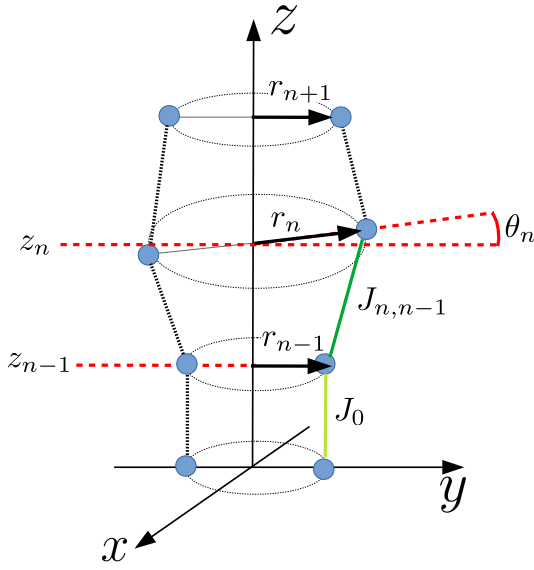


Figura 26

Figura esquemática do modelo Peyrard Bishop 3D. O parâmetro J_0 é a distância de equilíbrio no empilhamento, $J_{n,n-1}$ e θ são, respectivamente, a distância geral e a torção entre os pares de nucleotídeos $n, n - 1$.

entre nucleotídeos complementares. Então $F(\vec{r}_n, \vec{r}_{n-1})$ toma a seguinte forma:

$$F(\vec{r}_n, \vec{r}_{n-1}) = W(\vec{r}_n, \vec{r}_{n-1}) + V(\vec{r}_n). \quad (106)$$

Pela geometria da figura 26, nota-se que o sistema de coordenadas cilíndricas é o mais conveniente, uma vez que a ligação entre as fitas no B-DNA é praticamente na direção radial. Na versão tridimensional em coordenadas cilíndricas, a função potencial $F(\vec{r}_n, \vec{r}_{n-1})$ fica com a seguinte forma:

$$F(\vec{r}_n, \vec{r}_{n-1}) = F(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1}) \quad (107)$$

ou de forma mais explícita,

$$F(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1}) = W(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1}) + V(\vec{r}_n). \quad (108)$$

3.1 Modelo 3D aproximado

Para que as equações (102) e (105) sejam válidas (do ponto de vista puramente matemático), basta que a função $\exp[-\beta F(\vec{r}_n, \vec{r}_{n-1})]$ seja integrável, o que representa um vasto conjunto de funções, uma vez que a condição é pouco restritiva. Todavia para aplicar o método TIT precisamos integrar analiticamente em duas das três variáveis, o que limita muito nosso conjunto de funções $F(\vec{r}_n, \vec{r}_{n-1})$. Mesmo usando o potencial harmônico para o empilhamento, as

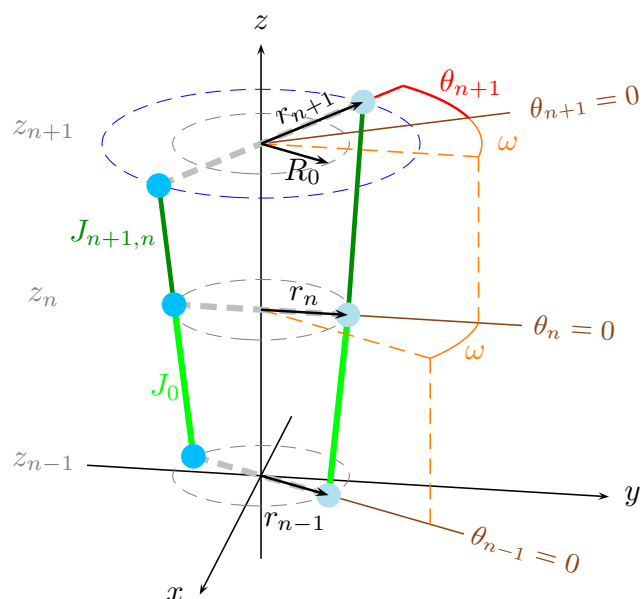


Figura 27

Figura esquemática do modelo 3D. Os pares de bases $n - 1$ e n (bolas azuis) são mostrados em suas posições de equilíbrios, e o par de base $n + 1$ está rodado por θ_{n+1} e deslocado de sua posição de equilíbrio para r_{n+1} . O ângulo de torção ω (linhas laranjadas) define a origem do deslocamento angular entre pares de bases consecutivos $\theta = 0$ (linha marrom). A linha verde mostra a interação de empilhamento entre bases vizinhas, em particular a distancia $J_{n+1,n}$ são mostradas pelas linhas verde escuro e J_0 é a distancia de equilíbrio no empilhamento, mostrado na cor verde claro. Os círculos cinza mostram o raio da distância de equilíbrio R_0 e o círculo azul mostra a distância de separação r_{n+1} entre as fitas.

integrais não são analíticas em nenhuma das variáveis e isso nos dá duas opções: a primeira é resolver numericamente as $3N$ integrais, o que é computacionalmente muito custoso para N grande. A segunda opção é buscar simplificações que nos permita integrar analiticamente duas das três variáveis. Como estamos interessados em desenvolver um modelo que nos permita calcular a temperatura de desnaturação de uma sequência de tamanho qualquer, inclusive grande, devemos escolher a segunda opção.

A representação geométrica do nosso modelo está representada na figura 27. Antes de introduzirmos as simplificações específicas do nosso modelo, vale ressaltar que os modelos mesoscópicos tridimensionais encontrados na literatura [56, 62, 63] frequentemente consideram apenas o movimento simétrico entre os nucleotídeo que formam par complementar. Essa simplificação reduz pela metade o número de variáveis. A primeira aproximação que fizemos foi considerar que a interação entre os pares de base complementares ocorre somente na direção radial. Essa aproximação se reflete no potencial que descreve a ligação de hi-

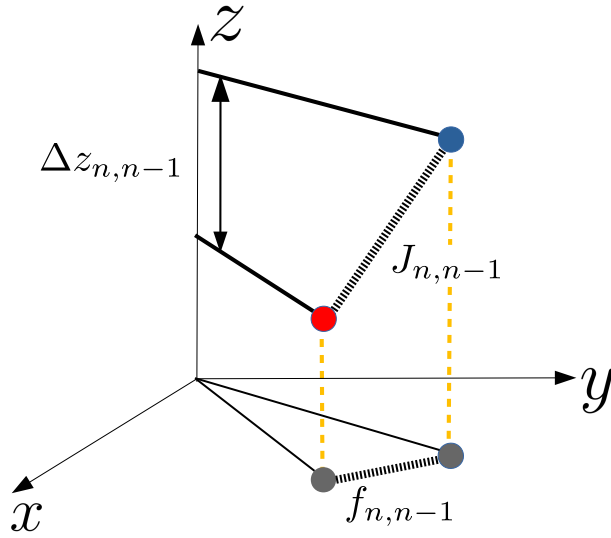


Figura 28

Definição da distância $J_{n,n-1}$ e sua projeção $f_{n,n-1}$ no plano polar.

drogênio (usamos o potencial de Morse) eliminando a dependência das variáveis θ e z , ou seja, $V(\vec{r}_n) = V(r_n)$. Isso introduz uma enorme simplificação, pois, como veremos adiante, teremos que desacoplar as variáveis (r, θ, z) e caso a função $V(\vec{r}_n)$ dependa de (r, θ, z) , a tarefa fica praticamente impossível. Essa aproximação é coerente, principalmente para o *B-DNA*, pois suas ligações de hidrogênio são praticamente perpendiculares ao eixo z [64].

O potencial harmônico, que descreve o empilhamento, é dado pela equação (109)

$$W(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1}) = \frac{k}{2} (J_{n,n-1} - J_0)^2, \quad (109)$$

onde a distância de empilhamento entre dois nucleotídeos vizinhos na mesma fita é

$$J_{n,n-1} = \sqrt{(z_n - z_{n-1})^2 + f_{n,n-1}^2}. \quad (110)$$

De maneira mais explícita, o potencial de empilhamento é dado por

$$W(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1}) = \frac{k}{2} \left(\sqrt{(z_n - z_{n-1})^2 + f_{n,n-1}^2} - J_0 \right)^2, \quad (111)$$

e

$$f_{n,n-1} = \sqrt{r_n^2 + r_{n-1}^2 - 2r_n r_{n-1} \cos(\omega + \theta_n - \theta_{n-1})}. \quad (112)$$

A função $f_{n,n-1}$ é a projeção de $J_{n,n-1}$ no plano xy , conforme a figura 28. É importante

destacar que em um modelo planar a projeção de $J_{n,n-1}$ é nula. Usando coordenadas polares e as equações (111) e (112), as equações (102) e (105) possuem, respectivamente, as seguintes formas

$$Z_T = \Gamma \int_0^\infty \int_{-\Theta}^\Theta \int_{(n-1)h_0-\zeta}^{(n-1)h_0+\zeta} \prod_{n=1}^N dz_n d\theta_n dr_n r_n \exp\{-\beta F(\vec{r}_n, \vec{r}_{n-1})\} \quad (113)$$

e

$$\langle r_k \rangle = \frac{1}{Z} \int_0^\infty \int_{-\Theta}^\Theta \int_{(n-1)h_0-\zeta}^{(n-1)h_0+\zeta} \prod_{n=1}^N dz_n d\theta_n dr_n r_n r_k \exp\{-\beta F(\vec{r}_n, \vec{r}_{n-1})\}. \quad (114)$$

Nosso modelo permite introduzir aproximações na equação (114) para torná-las integráveis nas variáveis θ e z .

Nossa próxima aproximação considera o deslocamento $z_n - z_{n-1}$ ao longo do eixo z muito pequeno, assim podemos usar a seguinte aproximação:

$$(z_n - z_{n-1})^2 \approx J_0^2. \quad (115)$$

Em nossos cálculos numéricos, usamos $J_0 = 7.0 \text{ \AA}$ baseado no valor usado na referência [56] e $\zeta = 0.1 \text{ \AA}$, escolhido empiricamente por não haver referências para esse valor nos modelos mesoscópicos tridimensionais.

Com essa aproximação, a função $F(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1})$ perde a dependência da variável z , portanto $F(r_n, r_{n-1}, \theta_n, \theta_{n-1}, z_n, z_{n-1}) = F(r_n, r_{n-1}, \theta_n, \theta_{n-1})$. Dessa forma podemos facilmente integrar as equações (113) e (114) na variável z , resultando respectivamente em

$$Z_T = \Gamma (4\zeta)^N \int_0^\infty \int_{-\Theta}^\Theta \prod_{n=1}^N dr_n d\theta_n r_n \exp\{-\beta F(r_n, r_{n-1}, \theta_n, \theta_{n-1})\} \quad (116)$$

e

$$\langle r_k \rangle = \frac{(4\zeta)^N}{Z} \int_0^\infty \int_{-\Theta}^\Theta \prod_{n=1}^N dr_n d\theta_n r_n r_k \exp\{-\beta F(r_n, r_{n-1}, \theta_n, \theta_{n-1})\}. \quad (117)$$

É importante observar que essa aproximação elimina de $\langle r_k \rangle$ a dependência de ζ , situação que não ocorre com Z .

A próxima aproximação que faremos será baseada no fato de que $f_{n,n-1} < J_0$. Usando essa condição, faremos uma expansão em série de Taylor da função $W(r_n, r_{n-1})$. Antes de fazer isso, reescrevemos o potencial de harmônico da seguinte forma:

$$\begin{aligned} W_{n,n-1} &= \frac{k}{2} \left(\sqrt{J_0^2 + f_{n,n-1}^2} - J_0 \right)^2 \\ &= \frac{k}{2} J_0^2 \left(\sqrt{1 + \frac{f_{n,n-1}^2}{J_0^2}} - 1 \right)^2. \end{aligned} \quad (118)$$

Para facilitar a notação, vamos usar as seguintes definições:

$$x^2 = \frac{f_{n,n-1}^2}{J_0^2} \quad (119)$$

$$g(x) = \left(\sqrt{1 + x^2} - 1 \right)^2. \quad (120)$$

Em termos da função $g(x)$ que definimos anteriormente, $W_{n,n-1}$ possui a seguinte expressão:

$$W_{n,n-1} = \sum_{n=1}^N \frac{k}{2} J_0^2 g(x). \quad (121)$$

O próximo passo é expandir a função $g(x)$ em torno de $x_0 = 0$.

$$g(x) = \sum_{i=2}^{\infty} \frac{g^{(i)}(0)}{i!} x^{2i}, \quad (122)$$

onde $g^{(i)}(0)$ indica a i -ésima derivada de $g(x)$ em $x_0 = 0$. Então a equação (121) torna-se

$$W_{n,n-1} = \frac{k J_0^2}{2} \sum_{i=2}^{\infty} \frac{g^{(i)}(0)}{i!} x^{2i}. \quad (123)$$

Reescrevendo o potencial de empilhamento $W_{n,n-1}$ em termos de $f_{n,n-1}$ obtemos

$$W_{n,n-1} = \frac{k J_0^2}{2} \sum_{i=2}^{\infty} \frac{g^{(i)}(0)}{J_0^{2i} i!} f_{n,n-1}^{2i}. \quad (124)$$

Nosso próximo passo será considerar pequenas torções, o que nos permite aproximar $\cos(\omega + \theta_n - \theta_{n-1})$ por

$$\cos(\omega + \theta_n - \theta_{n-1}) \approx 1 - \frac{(\omega + \theta_n - \theta_{n-1})^2}{2}. \quad (125)$$

Com a expansão acima, as variáveis θ_n e θ_{n-1} continuam acopladas (lembrando que nosso objetivo é promover o desacoplamento para realizar a integração), por isso vamos considerar $\theta_n - \theta_{n-1} < \omega$. Assim podemos escrever a equação (125) como

$$\cos(\omega + \theta_n - \theta_{n-1}) \approx 1 - \frac{\omega^2}{2}. \quad (126)$$

Usando a aproximação (126), a função $f_{n,n-1}$ da equação (112) torna-se

$$f_{n,n-1} \approx \sqrt{(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}}. \quad (127)$$

Substituindo (127) em (124) obtemos

$$W_{n,n-1} = \frac{kJ_0^2}{2} \sum_{i=2}^{\infty} \frac{g^{(i)}(0)}{J_0^{2i} i!} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^i. \quad (128)$$

Com tais aproximações, obtemos um hamiltoniano 1D, que pode ser expresso como

$$H(r_n, r_{n-1}) = \frac{kJ_0^2}{2} \sum_{i=2}^{\infty} \frac{g^{(i)}(0)}{J_0^{2i} i!} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^i + V(r_n) \quad (129)$$

Com o potencial expandido, podemos reescrever as equações (116) e (117) como

$$Z_T = \Gamma(16\zeta\Theta)^N \times \int_0^\infty \prod_{n=1}^N dr_n r_n \exp \left\{ -\beta \left[\frac{kJ_0^2}{2} \sum_{i=2}^{\infty} \frac{g^{(i)}(0)}{J_0^{2i} i!} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^i + V(r_n) \right] \right\} \quad (130)$$

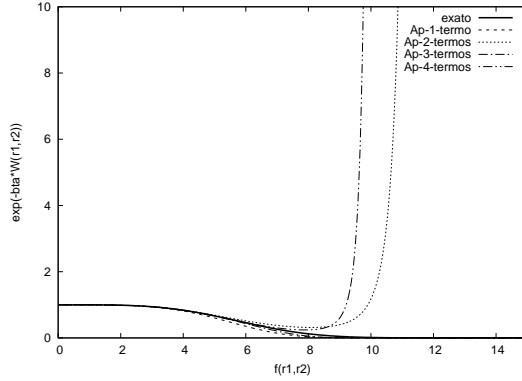


Figura 29

Exponencial do potencial elástico com $N = 2$ em termos de $f(r_1, r_2)$ nm. Usamos $\omega + \theta_n - \theta_{n-1} = 0.05$ rad e $z_n - z_{n-1} = 0$ nm.

e

$$\langle r_k \rangle = \frac{(16\zeta\Theta)^N}{Z_T} \times \int_0^\infty \prod_{n=1}^N dr_n r_n r_k \exp \left\{ -\beta \left[\frac{kJ_0^2}{2} \sum_{i=2}^\infty \frac{g^{(i)}(0)}{J_0^{2i} i!} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^i + V(r_n) \right] \right\}. \quad (131)$$

As aproximações feitas em (115) e (126) são razoáveis para temperaturas próximas à temperatura de desnaturação. A primeira delas é justificada pelo fato das ligações covalentes presentes no empilhamento ser muito mais fortes que as interações de hidrogênio, então a desnaturação ocorre à uma temperatura muito menor que aquela necessária para alterar significativamente a distância entre os nucleotídeos vizinhos na mesma fita. Outro fenômeno observado experimentalmente é que, à medida que a temperatura aproxima inferiormente de T_m , o duplexo torna-se menos “torcido”, ou seja, aproxima-se da estrutura planar, o que justifica a aproximação em pequenos ângulos.

A expansão de $W_{n,n-1}$ implica desvios das equações (130) e (131) em relação às respectivas equações (113) e (114). Para avaliar tais diferenças, comparamos a exponencial da função $W_{n,n-1}$ exata e expandida, conforme a figura 29. Na figura 29 notamos que a aproximação com $f(r_1, r_2) < 7$ já é boa para um termo e se a expansão for de ordem ímpar, mesmo para $f(r_1, r_2)$ grande, não há erros muito significativos, o que nos permite ter confiança em aproximar o potencial sem cometer grandes desvios.

Os núcleos das equações (130) e (131) só possuem dependência da variável radial (r_n) e,

além disso, podem ser simetrizados. Isso é fundamental para que possamos usar a técnica de integral de transferência (TIT) e conseqüentemente reduzir significativamente o custo computacional no caso de N grande.

3.2 Análise numérica

Nesta seção apresentaremos os resultados obtidos com o nosso modelo e também uma análise comparativa entre a hamiltoniana aproximada e a completa, onde integramos as três variáveis espaciais para o caso $N = 2$ com o objetivo de obter uma noção do quão boas são nossas aproximações. Como dissemos anteriormente, o cálculo sem as aproximações possui um alto custo computacional quando o número de pares de base é grande, já que para cada par de nucleotídeo temos três integrais espaciais para resolver, por isso fizemos o caso da menor estrutura possível ($N = 2$). Para analisar as aproximações com as diferentes técnicas de integração, calculamos Z e $\langle r \rangle$ para cada uma das aproximações (aproximação em z equação (115) e em θ equação (118)) usando TIT e integração usando a quadratura de gauss-legendre.

A expansão que resultou na equação (124) exige que $f_n(r_n, r_{n-1}) < J_0$, portanto temos que introduzir tal condição nas integrações que fizemos. Após as aproximações de pequenos ângulos e pequenos desvios em z , tal condição é expressa por

$$(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1} < J_0^2, \quad (132)$$

A condição imposta pela inequação (132) modifica o domínio do integrando e isso é introduzido através dos limites da integração. Sendo assim, temos que analisar qual região no quadrado $[r_n \times r_{n-1}]$ devemos considerar. Para o caso onde $\omega = 0$, temos $|r_n - r_{n-1}| < J_0$ e o domínio para $K(r_n, r_{n-1})$ não é mais um quadrado, conforme mostrado na figura 30.

É importante notar que, caso não limitemos o domínio, a função de partição será superestimada, já que o núcleo é uma função positiva. O termo $(r_n - r_{n-1})^4$ cresce rapidamente à medida que r_n se afasta de r_{n-1} , fazendo com que $\exp[-\beta(r_n - r_{n-1})^4] \rightarrow 0$.

Para abreviar, definimos um conjunto de nomenclatura para referirmos às diferentes situações de aproximação e integração: C que indica potencial completo e integração de todas as seis variáveis pela quadratura gauss-legendre, R significa potencial com as aproximações das equações (115,126) e mesma integração usado em CU, A indica aproximações (124, 115, 126) e mesmo método de integração dos anteriores, porém as variáveis angulares e longitudinais (z) são integradas analiticamente e T representa as aproximações (124, 115, 126) usando

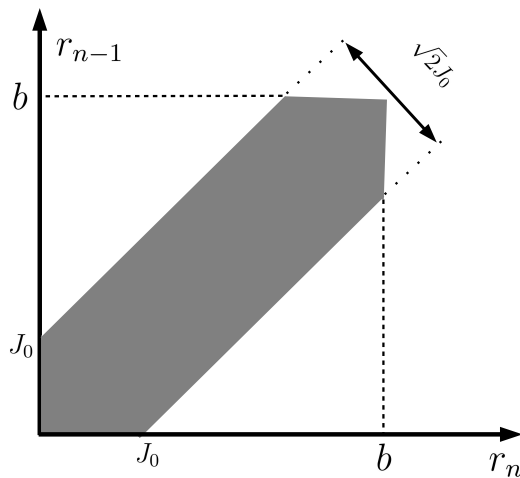


Figura 30

Região de integração da variável radial sem a torção com a condição $J_0 > f_{n,n-1}$.

método	aproximação	cálculo numérico
CU, CL	nenhuma	integral sextupla
RU, RL	115,126	integral sextupla
AU, AL	115,124,126	integral dupla
TU, TL	115,124,126	diagonalização

Tabela 1

Resumo da avaliação numérica usada para cálculo da função de partição e $\langle r \rangle$, sua identificação (ID), restrições e métodos de cálculo.

TIT para a integração numérica. As siglas L e U significam, respectivamente, região de integração da coordenada r limitada (conforme a figura 30) e ilimitada. A tabela 1 apresenta um resumo das nomenclaturas que adotamos.

3.3 Comparação de Z e $\langle r \rangle$ usando as hamiltonianas exata e aproximada.

Nesta seção apresentaremos uma série de testes que realizamos afim de verificar possíveis desvios nos valores de Z e $\langle r \rangle$ ocasionado pelas aproximações que fizemos na hamiltoniana. Iniciamos com os cálculos de Z e $\langle r \rangle$ em função do limite superior de integração da variável

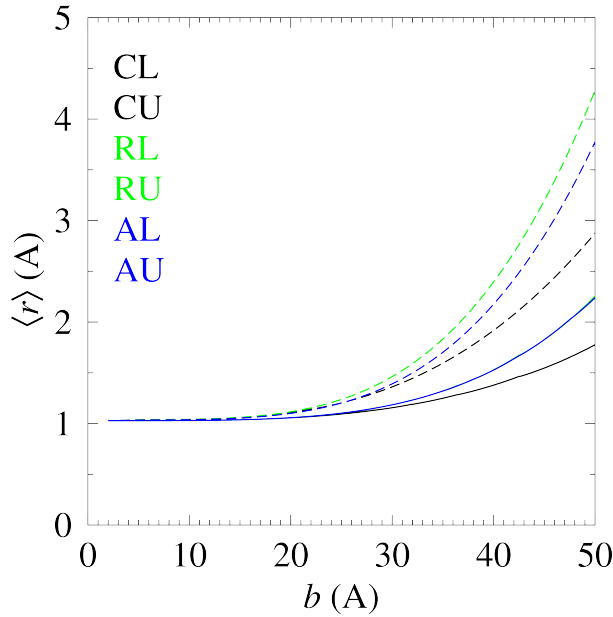


Figura 31

A figura mostra desvios médios $\langle r \rangle$ em função do limite superior de integração b . Parâmetros: $\Theta = 0.01$ rad, $\omega = 0.00$ rad e $T = 250$ K

radial b . Esse teste possui grande importância, uma vez que o modelo Peyrard-Bishop bidimensional apresenta um comportamento anômalo de divergência, isto é, à medida que b aumenta, os valores de $\langle r \rangle$ crescem, mesmo mantendo fixo a temperatura, conforme foi mostrado na figura 16 na página 37 e relatado em [52].

Usando $\omega = 0$ rad, calculamos a separação média entre os pares com as diferentes aproximações e técnicas de integração. Nesse caso, estamos considerando a torção sem preferência de orientação, ou seja, $\langle \Delta\theta \rangle = 0$. Em todos os casos observamos divergência, ou seja, sua origem não é nas aproximações, conforme mostra a figura 31. Então, somente com os potenciais harmônico e Morse, a divergência em função de b também está presente nessa versão para o caso $\omega = 0$ rad, mesmo que consideremos movimento na direção angular. Matematicamente é fácil entender o motivo pelo qual a divergência aparece mesmo quando consideramos a possibilidade de torção em torno de um plano formado pela dupla fita, isto é, $\omega = 0$ rad. Ao integrarmos a variável angular passaremos em algum momento por $\theta = 0$ rad e quando isso ocorre as outras integrais divergem.

Como dissemos anteriormente, a torção ω surge naturalmente no modelo tridimensional, mas vale ressaltar que ω representa uma torção fixa, isto é, uma torção de equilíbrio onde o par de nucleotídeo poderá girar em seu entorno. Todavia nas aproximações que fizemos, o ângulo ω deve ser pequeno, assim como a variação angular (Θ) em torno de ω . Refizemos os

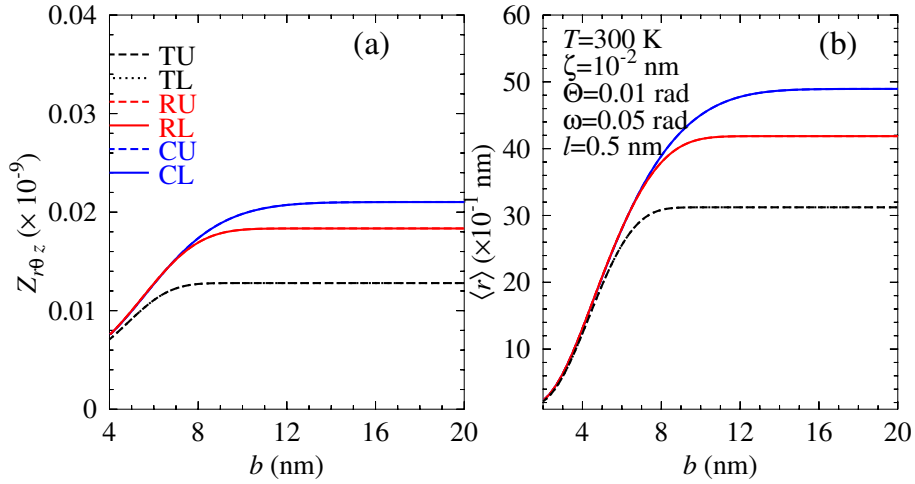


Figura 32

A figura mostra a função de partição e os desvios médios em função do limite superior de integração b . ω é um ângulo de torção, ζ é o intervalo de integração ao longo do eixo z e Θ é o intervalo de integração da variável angular. Parâmetros: $\zeta = 0.01 \text{ nm}$, $\Theta = 0.01 \text{ rad}$, $\omega = 0.05 \text{ rad}$ e $T = 300 \text{ K}$.

cálculos que originaram a figura 31 considerando $\omega = 0.05 \text{ rad}$ e $\Theta = 0.01 \text{ rad}$. Nesse caso estamos eliminando a possibilidade da fitar atingir a configuração planar, já que $\omega - 2\Theta > 0 \text{ rad}$. Assim, como proposto por Weber et al. [53], esperávamos eliminar a dependência de b da função de partição e do raio médio com essas considerações sobre os ângulos. O resultado segue na figura 32, que mostra que de fato a torção elimina a divergência causada pelo aumento de b , mesmo para ângulos pequenos. Isso é fundamental para o modelo, pois tanto Z quanto $\langle r \rangle$ deve perder a dependência da variável radial para r muito grande, uma vez que, para distâncias grandes entre os nucleotídeos complementares, o potencial de morse possui força de interação que tende a zero. Sendo assim, podemos dizer que nosso modelo já apresenta uma vantagem em relação aos PB bidimensionais, pois elimina a divergência de maneira intrínseca.

O próximo passo foi analisar o efeito da aproximação na variável angular. Nesse caso, integramos z , r e θ , porém variamos o limite de integração da variável angular (Θ) em torno do ângulo de torção ω (ver figura 33). A função partição calculada a partir da hamiltoniana aproximada difere pouco daquela calculada sem aproximação para ϕ_0 pequenos, mostrando que a aproximação é razoável. Para $\langle r \rangle$, quando calculamos usando a aproximação, a dependência angular desaparece como era de se esperar, já que a dependência de Θ é cancelada. Para o cálculo exato, a dependência modifica significativamente a separação média, mas quando se

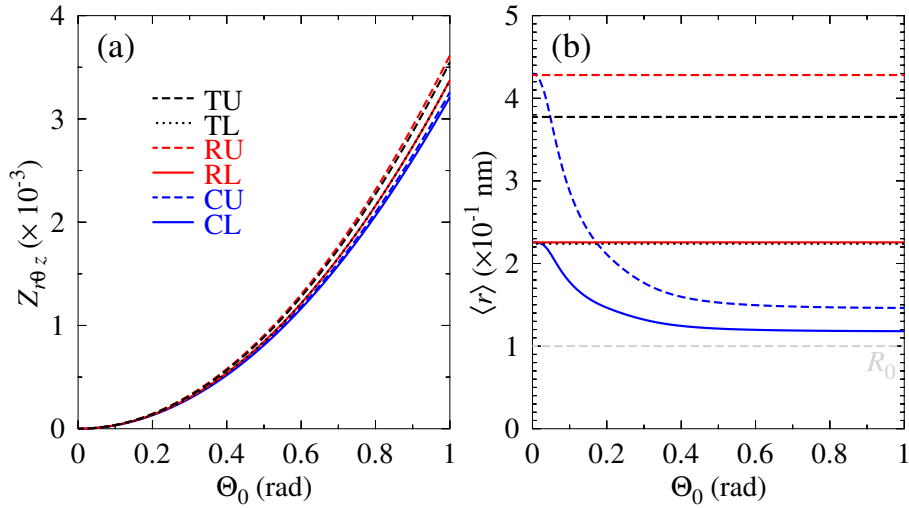


Figura 33

A figura mostra a função de partição e os desvios médios em função do limite superior de integração Θ_0 . Parâmetros: $b = 20$ nm, $\zeta = 0.1$ rad, $\omega = 0.05$ rad e $T = 250$ K.

limita a região de integração, em todos os casos $\langle r \rangle$ tende ao mesmo valor quando $\Theta \rightarrow 0$. Já quando não limitamos a região de integração, a aproximação representa na equação (128) introduz um pequeno desvio, mesmo quando $\Theta \rightarrow 0$. Por outro lado, devemos ressaltar que a figura 33 foi calculada utilizando apenas um termo na expansão de (128). No caso onde não há aproximação do potencial, o aumento de Θ faz com que o valor de $\langle r \rangle$ aproxime de R_0 , como era de se esperar, uma vez que a torção dificulta a abertura da fita. Vale ressaltar que quando consideramos Θ grande, a integração contabilizará também os valores de Θ pequenos, portanto justifica o aumento da função de partição quando Θ assume valores maiores. Todavia a contribuição para $\langle r \rangle$ quando a dupla hélice estiver em uma configuração muito torcida é muito baixa, já que a temperatura considerada não é suficiente para desnaturar o polímero com tal conformação.

Analisamos também o efeito da aproximação feita na variável z , conforme mostra a figura 34. Nesse caso, integramos as variáveis radial, angular e longitudinal (z) e variamos o limite de integração (ζ) ao longo do eixo z . O resultado nos permite concluir que, para pequenos deslocamentos ao longo de z , a função de partição calculada com o hamiltoniano aproximado fica em excelente acordo com o cálculo exato. No caso de $\langle r \rangle$, não há dependência de ζ , pois os termos dependentes de ζ são idênticos no numerador e no denominador.

Analisamos o comportamento da função de partição Z e $\langle r \rangle$ como função da tempera-

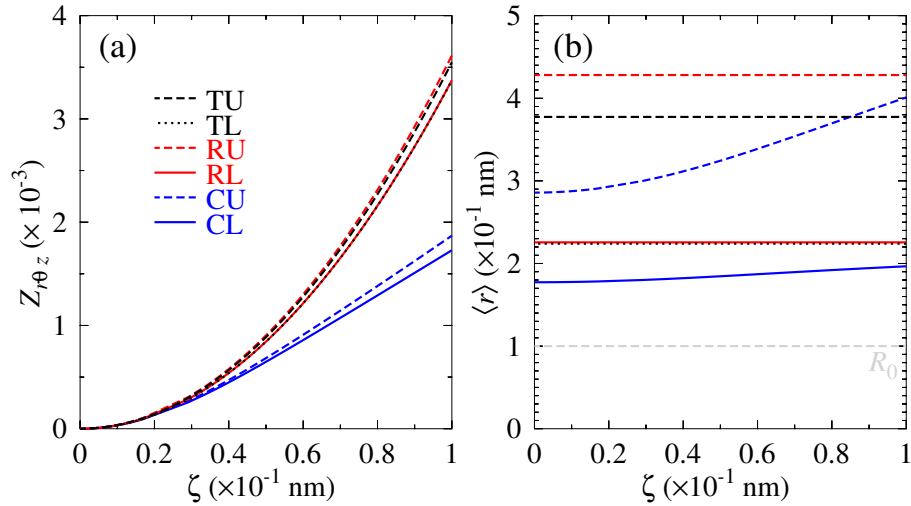


Figura 34

A figura mostra a função de partição e os desvios médios em função do limite superior de integração ζ da variável z . Parâmetros: $b = 20 \text{ nm}$, $\Theta = 0.01 \text{ rad}$, $\omega = 0.05 \text{ rad}$ e $T = 250 \text{ K}$.

tura (T). Esse resultado nos permite comparar o comportamento da curva de desnaturação da molécula homogênea. O resultado segue na figura 35. Os valores de Z (figura 35a) aproximados e exato passam a diferenciar mais significativamente à medida que a temperatura aumenta, embora mantêm o comportamento de crescimento semelhante. Por outro lado, a figura (figura 35b) mostra que $\langle r \rangle$ também passa a diferenciar nos casos exatos e aproximados à medida que T aproxima da temperatura de desnaturação. No caso do modelo aproximado TU, vemos que a temperatura de desnaturação é superior aos demais casos.

Com os dados que analisarmos até aqui, concluímos que a aproximação da hamiltoniana produz resultados razoáveis quando comparados com o cálculo exato se considerarmos as premissas de ângulos e deslocamentos ζ pequenos. Esse estudo de comparação nos permitiu comparar o efeito das aproximações utilizadas e nos deu uma certa confiança para aplicar o modelo aproximado em sequências maiores, mas é importante frisar que o fato da aproximação funcionar bem com dois pares de base não implica necessariamente que o caso geral de uma sequência com N pares de nucleotídeos, esse comportamento irá permanecer.

3.3.1 Variação dos Parâmetros dos potenciais

Os potenciais possuem parâmetros que devem ser escolhidos afim de minimizar a diferença entre os dados experimentais e os teóricos previstos pelo modelo. A técnica para se obter tal

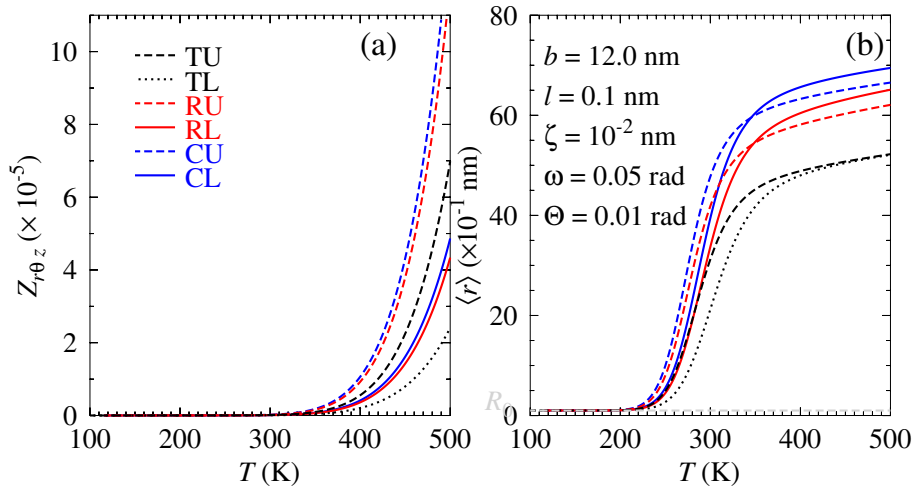


Figura 35

A figura (a) mostra o comportamento $\langle r \rangle$ e da função partição em função da temperatura para uma sequência com dois pares de bases usando o modelo aproximado.

minimização é a equivalência termodinâmica, mostrada na figura 22 na página 44. Esse passo é fundamental para que possamos obter resultados quantitativos, já que os parâmetros podem mudar de acordo com o modelo. Para termos uma ideia disso, o modelo Peyrard Bishop bidimensional considerou $D = 0.33$ eV [38], Barbi et al. [62] usou $D = 0.04$ eV [62] e o modelo tridimensional proposto por Cocco e Monasson [56] considerou $D = 0.16$ eV. Isso ocorre porque cada modelo possui aproximações particulares que modificam o formato da força de interação entre os componentes da dupla hélice. O modelo Peyrard Bishop bidimensional, por exemplo, trabalha com uma força de empilhamento cuja direção é exclusivamente perpendicular à molécula (mesma direção das ligações de hidrogênio), já os modelos tridimensionais [56, 62] consideram a força de empilhamento na direção entre os nucleotídeos vizinhos, ou seja, a direção perpendicular (considerada em [38]) é apenas uma das componentes. Evidentemente nosso modelo também possui aproximações que possivelmente requerem conjuntos de parâmetros diferentes dos outros modelos, por isso, para termos resultados quantitativos, necessitamos obter esse conjunto de parâmetros otimizados, que pode ser obtido pela técnica da equivalência termodinâmica. Nosso próximo passo será fazer esse trabalho de otimização dos parâmetros e vale ressaltar que isso somente poderá ser feito se tivermos o modelo homogêneo finalizado, daí uma parte importante do nosso trabalho.

Como nosso estudo, por enquanto, é qualitativo, escolhemos parâmetros que nos permitiu

obter a temperatura de desnaturação próxima a obtida com outros modelos (lembrando que estamos tratando uma molécula homogênea). Após obter alguns resultados razoáveis com o conjunto de parâmetros escolhido, variamos algumas dessas constantes afim de analisar o comportamento de $\langle r \rangle$ e Z . É importante salientar que o que fizemos não foi a otimização desses parâmetros, mas observar como cada um deles, separadamente, modificam os valores de $\langle r \rangle$ e Z .

É importante destacar que devemos ser cuidadosos ao usar condições de contorno periódicas (CCP) em nosso modelo. Isso porque partimos da premissa de uma dupla fita retilínea, que não pode se deslocar do eixo z e por isso há apenas dois casos onde podemos aplicar CCP sem inconsistência física: o primeiro é para dois pares de base, que devido à simetria do potencial harmônico podemos escrever

$$W_{1,2} + W_{2,1} = 2W_{1,N}. \quad (133)$$

No caso $N = 3$ isso não é possível, pois na soma das três contribuições $W_{1,N} + W_{2,3} + W_{3,1}$, o último termo corresponde à uma ligação a mais que não pode ser substituída por um fator multiplicativo (no caso $N = 2$ o fator é 2) e nem é desprezível, já que corresponde à 1/3 de toda a energia de empilhamento. Além disso, o termo $W(r_3, r_1)$ não possui simetria em relação aos outros, pois

$$W_{1,3} = \frac{k}{2} \left[\sqrt{(z_1 - z_3)^2 + f^2(r_1, r_3, \theta_1, \theta_3)} - 2J_0 \right]^2. \quad (134)$$

Evidentemente a distância de equilíbrio muda no caso $W_{1,3}$, pois não estamos pegando “vizinhos”. A segunda situação possível é considerar CCP é para sequências muito grandes ($N \rightarrow \infty$). Nesse caso a contribuição da ligação de empilhamento entre o último e o primeiro par de base é dada por

$$W_{1,N} = \frac{k}{2} \left[\sqrt{(z_1 - z_N)^2 + f^2(r_1, r_N, \theta_1, \theta_N)} - (N - 1)J_0 \right]^2. \quad (135)$$

Para N grande, o termo $(z_1 - z_N)^2 \approx [(N - 1)J_0]^2 \gg f^2(r_1, r_N, \theta_1, \theta_N)$, portanto podemos

reescrever a equação (135) como

$$\begin{aligned} W_{1,N} &\approx \frac{k}{2} \left\{ \sqrt{[(N-1)J_0]^2 - (N-1)J_0} \right\}^2 \\ &\approx \frac{k}{2} \{(N-1)J_0 - (N-1)J_0\}^2 = 0, \end{aligned} \quad (136)$$

ou seja, quando usamos sequências grandes, a contribuição do empilhamento entre o primeiro e o último par de base praticamente não altera a energia total do sistema. Todavia, se usarmos $W(r_1, r_N)$ como na equação (135), não teremos a simetria necessária para usar TIT. Então, substituímos (135) por

$$W_{1,N} = \frac{k}{2} \left[\sqrt{(z_1 - z_N)^2 + f^2(r_1, r_N, \theta_1, \theta_N)} - J_0 \right]^2. \quad (137)$$

Quanto estamos errando? Basta olhar a contribuição total do empilhamento na hamiltoniana, dado por

$$\sum_{n=1}^N W_{n,n-1} = \frac{k}{2} \overbrace{[W_{2,1} + \dots + W_{N,N-1}]}^{\text{contribuição exata}} + \frac{k}{2} \overbrace{W_{1,N}}^{\text{termo extra}}, \quad (138)$$

sendo assim, a contribuição do termo extra é $1/(N-1)$ da energia total de empilhamento e torna desprezível em sequências grandes. Todos nossos testes comparativos foram feitos com $N = 2$, portanto não há problemas relacionado à CCP. Dado a importância da inclusão do termo J_0 no modelo, calculamos $\langle r \rangle$ e $Z_{r,\theta,z}$ com diferentes valores de $\langle r \rangle$ e Z . O resultado segue na figura 36. Notamos que $Z_{r,\theta,z}$ não possui forte dependência de J_0 , tanto CU quanto TU sofrem apenas um leve aumento à medida que J_0 assume valores maiores. Além disso, a diferença entre os valores da função de partição é pequena se compararmos aos valores de $Z_{r,\theta,z}$ e se mantém praticamente constante à medida que J_0 aumenta.

Já $\langle r \rangle$ é mais sensível à mudanças do parâmetro J_0 , pois, como podemos ver na figura 36, os valores de $\langle r \rangle$ aumenta significativamente quando J_0 aumenta. Quando J_0 tende a zero tanto CU quanto TU tendem ao mesmo valor de $\langle r \rangle$. Esse comportamento no caso CU é devido ao fato que o modelo tridimensional se aproxima do modelo Peyrard-Bishop modificado,

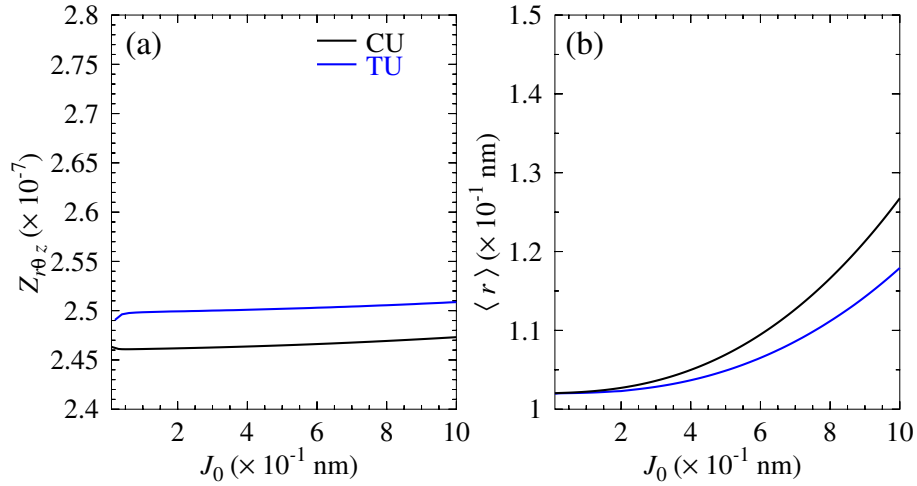


Figura 36

Valores de (a) Z e (b) $\langle r \rangle$ em função de J_0 para uma sequência com dois pares de bases, calculados com os potenciais exato e aproximado. Parâmetros utilizados: $a = 4.45 \times 10^{-1}$ nm, $D = 0.2$ eV, $k = 4.00$ eV/nm² e $T = 200$ K.

equação 65,(lembrando que assumimos ângulos muito pequenos), pois

$$\lim_{J_0 \rightarrow 0} \frac{k}{2} \left(\sqrt{(z_n - z_{n-1})^2 + f_{n,n-1}^2} - J_0 \right)^2 = \frac{k}{2} [r_n^2 + r_{n-1}^2 - 2r_{n-1}r_n \cos(\theta)]^2. \quad (139)$$

Como o modelo Peyrard-Bishop bidimensional modificado usa $D = 0,04$ eV e nós usamos $D = 0,20$ eV, a temperatura $T = 200$ K é muito baixa para que $\langle r \rangle$ desvie significativamente de $R_0 = 0,1$ nm (lembrando que valores grandes para D requerem temperaturas maiores para abrir a dupla fita).

Os parâmetros que se relacionam com a intensidade das ligações entre os nucleotídeos também influenciam os valores de Z e $\langle r \rangle$. Um maior valor para k é equivalente ao aumento da rigidez de cada fita do DNA, o que, em nosso modelo, significa $f_{n,n-1}$ e $z_n - z_{n-1}$ menores. À medida que k aumenta, esperamos uma diminuição dos valores de Z , já que o integrando diminui. Com relação à $\langle r \rangle$, esperamos que também diminua. Isso porque com o aumento da rigidez, torna-se mais difícil deslocar um único nucleotídeo em meio a seus dois vizinhos de empilhamento, portanto, privilegia a situação de abertura simultânea de todos os nucleotídeos, conforme mostra a figura 37. O aumento do valor de k diminui Z e $\langle r \rangle$, como podemos observar na figura 37. Esse aumento da rigidez eleva a temperatura de desnaturação, já que para $T < T_m$ o valor de $\langle r \rangle$ diminui à medida que k aumenta. Por outro lado, para k pequeno,

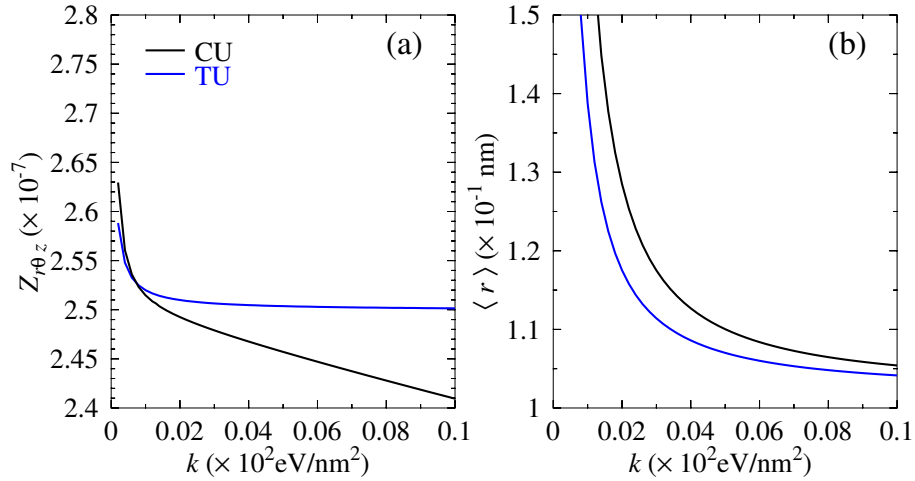


Figura 37

As figuras mostram os valores de (a) Z e (b) $\langle r \rangle$ em função de k para uma sequência com dois pares de bases, calculados a partir dos modelos exato e aproximado. Parâmetros utilizados: $a = 4.45 \times 10^{-1}$ nm, $D = 0.2$ eV, $J_0 = 7.0 \times 10^{-1}$ nm e $T = 200$ K.

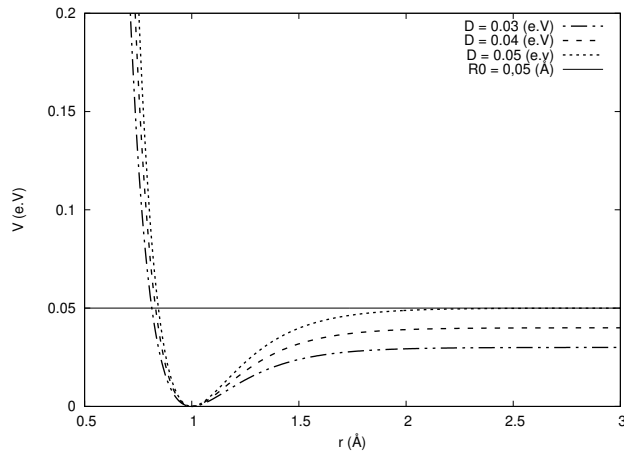


Figura 38

Potencial de morse usando três valores diferentes para o parâmetro D e a assíntota horizontal para o caso onde $D = 0,05$ eV. Parâmetro utilizado: $a = 4.45 \times 10^{-1}$ nm.

os valores de $\langle r \rangle$ diferenciam muito quando calculado com os modelos aproximado e exato, enquanto que com a função de partição ocorre o contrário.

Outro parâmetro que variamos foi o D do potencial de Morse, equação (27). Ele nos fornece a “profundidade” do poço, conforme mostra a figura 38 Na prática, a profundidade do poço indica a intensidade da ligação, já que para romper completamente a ligação é necessário uma quantidade de energia que supere a assíntota horizontal que é exatamente o valor de D , como podemos constatar na figura 38. Sendo assim, esperamos que maiores valores de D aumente a temperatura de desnaturação. No caso da molécula de DNA, sequências de nucleotídeos C-G necessitam de uma temperatura maior que sequências A-T, devido a ligação tripla no primeiro caso e dupla no segundo, portanto o valor do parâmetro D para sequências

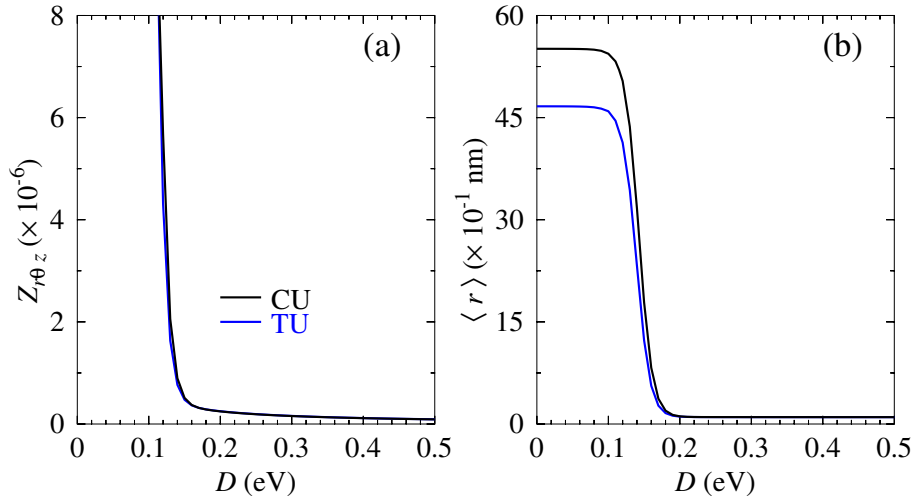


Figura 39

As figuras mostram os valores de (a) Z e (b) $\langle r \rangle$ em função de D para uma sequência com dois pares de bases, calculados a partir dos modelos exato e aproximado. Parâmetros utilizados: $k = 0.04 \times 10^2 \text{ eV/nm}^2$, $a = 4.45 \times 10^{-1} \text{ nm}$, $J_0 = 7.0 \times 10^{-1} \text{ nm}$ e $T = 200 \text{ K}$.

C-G precisa ser maior que no caso de A-T. Esperamos que nosso modelo reproduza tal efeito, por isso calculamos a função partição e a separação média entre os pares de base para diferentes valores de D , cujo resultado está na figura 39. Conforme esperado, o aumento do valor de D eleva a temperatura de desnaturação, no caso da figura 39, para $T = 200 \text{ K}$, quando $D > 0,1 \text{ eV}$ temos uma redução da separação média entre as bases. A função de partição também diminui seu valor à medida em que D aumenta e, em termos matemáticos, o aumento de D implica em um acréscimo no valor do potencial para r_n deslocado de R_0 , portanto o termo $\exp[-\beta V(r_n)]$ decresce e diminui o valor da integral, o que explica a redução da função de partição.

O outro parâmetro presente no potencial de Morse é o a , que está relacionado com o alcance da interação. Veja a figura 40. Como podemos observar na figura 40, quanto maior o valor de a maior é a espessura do poço, portanto maior o alcance da interação. Por outro lado, a curva de $V(r)$ é mais suave quando o parâmetro a assume valores maiores, portanto a inclinação da reta tangente torna-se menor. A intensidade da força de interação que esse potencial representa é dada por $F = -\frac{dV(r)}{dr}$, logo haverá diminuição de $|\vec{F}|$ e por isso esperamos um aumento de $\langle r \rangle$. Com relação à função de partição, o aumento de a faz com que $V(r)$ se aproxime da assíntota mais rapidamente e conseqüentemente a exponencial negativa

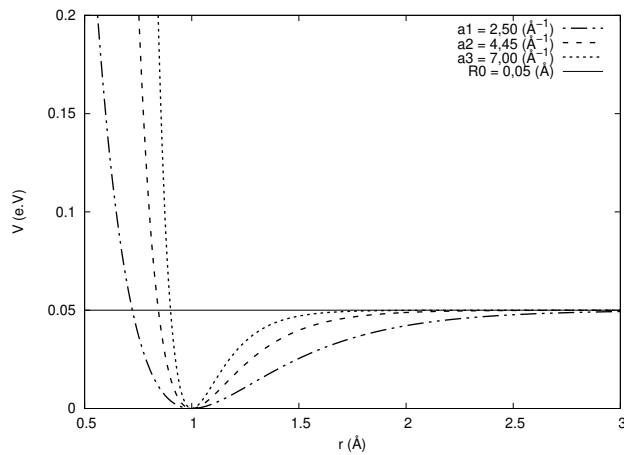


Figura 40

Potencial de morse usando três valores diferentes para o parâmetro a e a assíntota horizontal para o caso onde $D = 0,05$ eV.

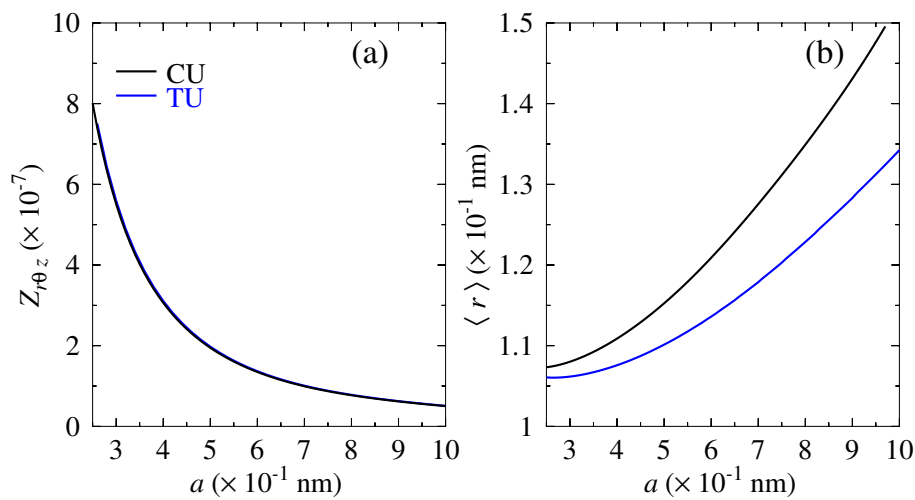


Figura 41

As figuras mostram os valores de (a) Z e (b) $\langle r \rangle$ em função de a para uma sequência com dois pares de bases, calculados a partir dos modelos exato e aproximado. Parâmetros utilizados: $k = 0.04 \times 10^2$ eV/nm², $D = 0.2$ eV, $J_0 = 7.0 \times 10^{-1}$ nm e $T = 200$ K.

de $V(x)$ terá valores menores para o domínio $r \in (0, b)$, o que deve diminuir o valor de Z . Calculamos a separação média e a função de partição para diferentes valores do parâmetro a , conforme mostrado na figura 41

As figuras 41a e 41b mostram que o comportamento de $Z_{r,\theta,z}$ e $\langle r \rangle$ estão em acordo com o que descrevemos anteriormente. Também podemos observar que, quanto maior o valor do parâmetro a , maior é a diferença entre os $\langle r \rangle$ calculados usando os modelos TU e CU, já a função de partição praticamente possui praticamente o mesmo valor nos dois casos.

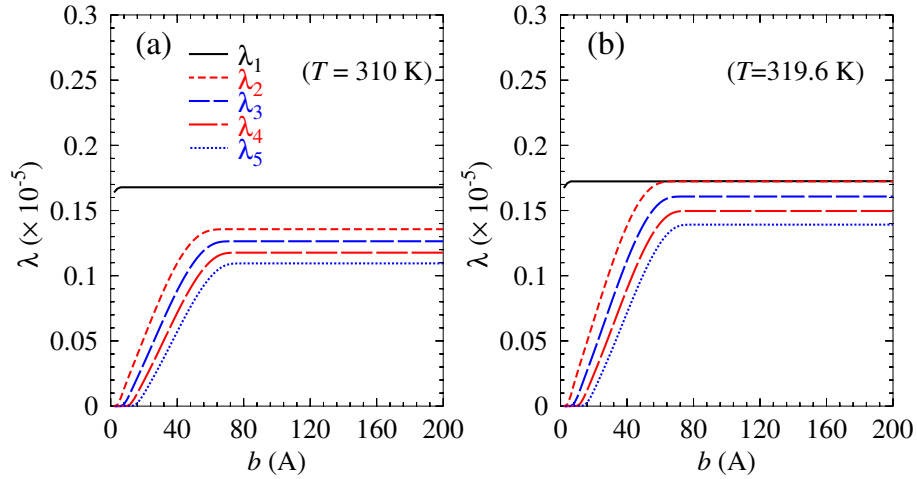


Figura 42

Cálculos dos autovalores em função do limite superior de integração b . Parâmetros: $D = 0.2$ eV, $a = 44.5$ nm⁻¹, $k = 4$ eV/nm², $J_0 = 0.7$ nm.

3.3.2 Autovalores

O cálculo da função partição usando a técnica de integral de transferência é feito a partir dos autovalores, conforme a equação (57), página 35. Então, de maneira geral, todas as propriedades físicas dependem diretamente deles, por isso realizamos um estudo onde buscamos entender como se comportam. Conforme relatado em [52], e discutido anteriormente na subsecção 3.3, o tamanho do intervalo de integração interfere substancialmente na obtenção da temperatura de desnaturação e os autovalores do modelo PB são degenerados quando $b \rightarrow \infty$, mas ao incluir uma torção, tal comportamento é eliminado. No nosso modelo, a torção ocorre naturalmente, por isso esperávamos não obter a degenerescência dos autovalores que, conforme a equação (57), é uma condição necessária para a convergência da função de partição (Z).

À temperatura de $T = 310$ K todos os autovalores são distintos e, exceto o maior deles (λ_1), todos os outros possuem comportamento similar, conforme a figura 42a. Já à temperatura de $T = 319,6$ K, os dois primeiros são degenerados para valores grandes de b e os demais distintos, 42b. Isso tem uma grande importância na escolha do tratamento de sequências longas e para entender isso precisamos voltar à equação 57 que calcula Z .

$$Z = \Gamma \sum_{i=1}^{+\infty} \lambda_i^N. \quad (140)$$

Como $0 < \lambda_i < 1$, em caso de sequências longas ($N \rightarrow \infty$), frequentemente nos deparamos na literatura ([38, 52]) com a seguinte aproximação:

$$\lim_{N \rightarrow \infty} Z = \Gamma \lambda_1^N. \quad (141)$$

Para analisar melhor a aproximação acima, vamos recorrer aos seguintes teoremas da álgebra linear [65]:

Teorema 1. *Sejam V um espaço vetorial de dimensão n sobre K e $T : V \rightarrow V$ um operador linear. T é diagonalizável se, e somente se, existe uma base B de V formada por autovetores de T .*

Teorema 2. *Sejam V um espaço vetorial sobre K e $T : V \rightarrow V$ um operador linear, com $\lambda_1, \dots, \lambda_k$ autovalores distintos de T e v_1, \dots, v_k os autovetores associados, respectivamente. Então, v_1, \dots, v_k são linearmente independentes.*

Se a matriz for de ordem M , o teorema 1 exige que haja M autovetores $\{|\phi_1\rangle, \dots, |\phi_M\rangle\}$ L.I para que possamos escrever a matriz na base $|\phi\rangle$ e o teorema 2 nos garante que se a matriz possuir todos os autovalores distintos então ela terá os M autovetores L.I, portanto será diagonalizável. Todavia, se a matriz possuir autovalores $\{\lambda_j\}$ com multiplicidade maior que 1, pelo teorema 1, obrigatoriamente os autovetores associados aos λ_j deverão ser L.I e precisam completar o conjunto dos autovetores para formar a base, caso contrário a matriz não será diagonalizável.

Quando temos todos os autovetores degenerados temos $\{\lambda_1 = \lambda_2 = \dots = \lambda_M\}$, então, para que a matriz seja diagonalizável, devemos ter $\{|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_M\rangle\}$ linearmente independentes (L.I) Então as equações que calculam Z e $\langle r \rangle$ tornam-se respectivamente

$$Z = \Gamma \sum_{i=1}^{+\infty} \lambda_i^N = \Gamma M \lambda_1^N \quad (142)$$

e

$$\langle r \rangle = \frac{\sum_{i=1}^{+\infty} \lambda_i^N \langle \phi_i | r | \phi_i \rangle}{\sum_{i=1}^{+\infty} \lambda_i^N} = \frac{\sum_{i=1}^{+\infty} \langle \phi_i | r | \phi_i \rangle}{M} \quad (143)$$

Outra situação relevante é quando $\lambda_1 > \lambda_2 = \lambda_3 \cdots \lambda_M$. Nesse caso, para que a matriz seja diagonalizável, o conjunto $\{|\phi_2\rangle, |\phi_3\rangle, \dots, |\phi_M\rangle\}$ deve ser L.I, portanto $\{|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_M\rangle\}$ formará uma base para o integrando e as equações que fornece o cálculo de Z e $\langle r \rangle$ tornam-se respectivamente

$$Z = \Gamma \sum_{i=1}^{+\infty} \lambda_i^N = \Gamma [\lambda_1^N + (M-1)\lambda_2^N] \quad (144)$$

e

$$\begin{aligned} \langle r \rangle &= \frac{\sum_{i=1}^M \lambda_i^N \langle \phi_i | r | \phi_i \rangle}{\sum_{i=1}^M \lambda_i^N} = \frac{\lambda_1^N \langle \phi_1 | r | \phi_1 \rangle}{\lambda_1^N + (M-1)\lambda_2^N} + \frac{\lambda_2^N \sum_{i=2}^M \langle \phi_i | r | \phi_i \rangle}{\lambda_1^N + (M-1)\lambda_2^N} \\ \langle r \rangle &= \left(\frac{\lambda_1}{\lambda_1}\right)^N \frac{\langle \phi_1 | r | \phi_1 \rangle}{1 + (M-1)\left(\frac{\lambda_2}{\lambda_1}\right)^N} + \left(\frac{\lambda_2}{\lambda_1}\right)^N \left[\frac{\sum_{i=2}^M \langle \phi_i | r | \phi_i \rangle}{1 + (M-1)\left(\frac{\lambda_2}{\lambda_1}\right)^N} \right] \end{aligned} \quad (145)$$

No caso onde N e M são muito grandes (tendem a infinito) temos

$$\begin{aligned} \langle r \rangle &\approx \frac{\langle \phi_1 | r | \phi_1 \rangle}{1 + (M-1)\left(\frac{\lambda_2}{\lambda_1}\right)^N} \\ \langle r \rangle &\approx \frac{\langle \phi_1 | r | \phi_1 \rangle}{1 + M\left(\frac{\lambda_2}{\lambda_1}\right)^N} \end{aligned} \quad (146)$$

O caso descrito acima é o que acontece no modelo Peyrard Bishop original (veja figura 19, página 40, da esquerda) quando não incluímos a torção ω .

Outra situação que será muito importante em nosso estudo é quando ocorrer

$\lambda_1 = \lambda_2 > \lambda_3 > \dots \lambda_M$. Nesse caso teremos

$$Z = \Gamma \sum_{i=1}^{+M} \lambda_i^N = \Gamma \left[2\lambda_1^N + \sum_{i=3}^M \lambda_i^N \right] = \Gamma \lambda_1^N \left[2 + \sum_{i=3}^M \left(\frac{\lambda_i}{\lambda_1}\right)^N \right] \approx 2\Gamma \lambda_1^N \quad (147)$$

A equação da separação média pode ser simplificada da seguinte maneira

$$\begin{aligned}
\langle r \rangle &= \frac{\lambda_1^N [\langle \phi_1 | r | \phi_1 \rangle + \langle \phi_2 | r | \phi_2 \rangle]}{2\lambda_1^N} + \frac{\sum_{i=3}^M \lambda_i^N \langle \phi_i | r | \phi_i \rangle}{2\lambda_1^N} \\
\langle r \rangle &= \frac{\langle \phi_1 | r | \phi_1 \rangle + \langle \phi_2 | r | \phi_2 \rangle}{2} + \frac{1}{2} \sum_{i=3}^M \left(\frac{\lambda_i}{\lambda_1} \right)^N \langle \phi_i | r | \phi_i \rangle \\
\langle r \rangle &\approx \frac{\langle \phi_1 | r | \phi_1 \rangle + \langle \phi_2 | r | \phi_2 \rangle}{2}
\end{aligned} \tag{148}$$

Quando os autovalores comportam-se como na figura 42a podemos calcular Z e $\langle r \rangle$, respectivamente, com as equações

$$Z = \Gamma \lambda_1^N \tag{149}$$

e

$$\langle r \rangle = \langle \phi_1 | r | \phi_1 \rangle. \tag{150}$$

Todavia uma situação como a da figura 42b não pode ser tratada com as equações (149) e (150), já que $\lambda_1 = \lambda_2$. Nesse caso o cálculo de Z e $\langle r \rangle$ deve ser feito, respectivamente, usando as equações (147) e (148).

A pergunta natural agora é: qual é o conjunto de temperaturas tal que $\lambda_i = \lambda_j$? Para responder tal questionamento, calculamos os autovalores em função da temperatura, conforme mostrado na figura 43. A figura 43 mostra que $\lambda_n \geq \lambda_{n+1}$ e que a igualdade é verificada apenas à uma única temperatura e o caso particular $\lambda_1 = \lambda_2$ ocorre quando $T = 319,6$ K. Sendo assim, devemos separar o domínio T para usarmos as equações (149), (150), (147) e (148) corretamente. Como estamos interessados na temperatura de desnaturação, o mais importante é analisar o que acontece quando $T \leq T_m$. Além disso, no caso de sequências grandes, apenas o anti-cruzamento de λ_1 com λ_2 é importante, já que os outros autovalores são menores, e portanto, $\lambda_1^N \gg \lambda_i^N$ se $i = 2, 3, \dots, M$. Sendo assim, podemos considerar apenas a temperatura referente ao primeiro anti cruzamento como aquela onde as funções Z e $\langle r \rangle$ assumem valores modificados devido ao domínio T .

Em resumo, quando estamos trabalhando com sequências grandes e queremos calcular Z

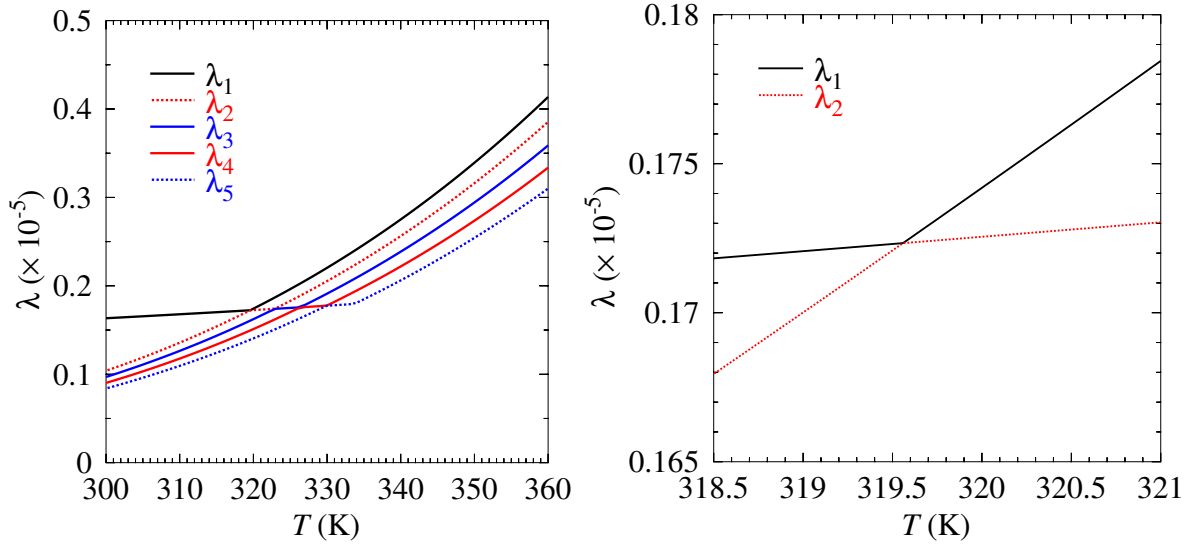


Figura 43

Cálculos dos autovalores em função da temperatura. Parâmetros: $D = 0.2$ eV, $b = 20$ nm, $a = 44.5$ nm⁻¹, $k = 4$ eV/nm², $J_0 = 0.7$ nm.

em todo o domínio T devemos usar

$$Z = \begin{cases} \Gamma \lambda_1^N, & \text{se } \lambda_1 > \lambda_2 > \dots > \lambda_M \\ 2\Gamma \lambda_1^N, & \text{se } \lambda_1 = \lambda_2 > \lambda_3 > \dots > \lambda_M \end{cases} \quad (151)$$

e para calcular $\langle r \rangle$ devemos usar

$$\langle r \rangle = \begin{cases} \langle \phi_1 | r | \phi_1 \rangle, & \text{se } \lambda_1 > \lambda_2 > \dots > \lambda_M \\ \frac{1}{2} (\langle \phi_1 | r | \phi_1 \rangle + \langle \phi_2 | r | \phi_2 \rangle), & \text{se } \lambda_1 = \lambda_2 > \lambda_3 > \dots > \lambda_M. \end{cases} \quad (152)$$

A equação (152) mostra que pode haver uma descontinuidade de $\langle r \rangle$ na temperatura onde $\lambda_1 = \lambda_2$ e para isso, basta que $\langle \phi_1 | r | \phi_1 \rangle \neq \langle \phi_2 | r | \phi_2 \rangle$.

Para o cálculo com sequências pequenas, não podemos usar o argumento

$$\left(\frac{\lambda_2}{\lambda_1} \right)^N \rightarrow 0,$$

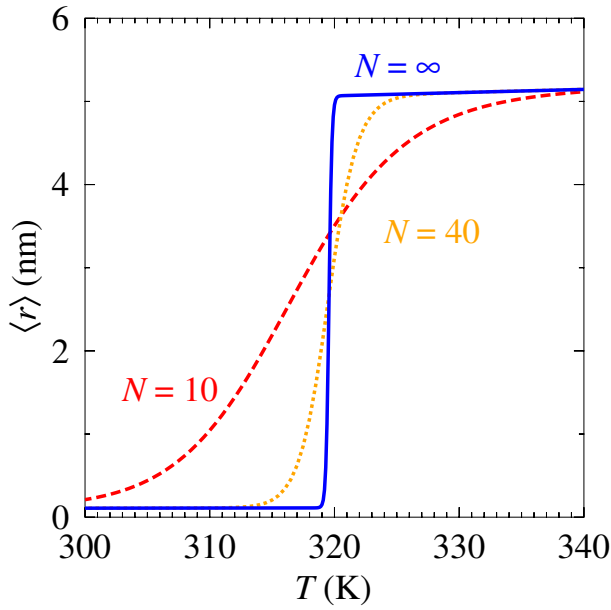


Figura 44

A figura (a) mostra o comportamento $\langle r \rangle$ em função da temperatura para sequências com diferentes números de pares de bases N usando o modelo aproximado. Parâmetros: $b = 20.0$ nm, $\zeta = 10^{-2}$ nm, $\omega = 0.05$ rad e $\Theta = 0.01$ rad.

portanto as equações (151) e (152) não se aplicam. Nesse caso, devemos usar

$$Z = \Gamma \sum_{i=1}^M \lambda_i^N \quad (153)$$

$$\langle r \rangle = \frac{\sum_{i=1}^M \lambda_i^N \langle \phi_i | r | \phi_i \rangle}{Z} \quad (154)$$

3.3.3 Estudo numérico de sequências grandes usando o Modelo 3D aproximado

Nosso próximo passo foi aplicar o modelo aproximado para sequências maiores, uma vez que a TIT nos permite calcular $\langle r \rangle$ para uma sequência de tamanho qualquer com praticamente o mesmo custo computacional. Calculamos $\langle r \rangle$ em função da temperatura (T) para sequências de tamanho $N = 10$, $N = 40$ e $N = 400$ (para efeito das aproximações consideramos infinito) com os mesmos parâmetros de integração. O resultado está apresentado na figura 44. A curva de desnaturação apresenta transição mais abrupta para sequências maiores, aproximando-se de uma descontinuidade à medida que N aumenta, já para sequências pequenas a curva de desnaturação é bastante suave, além de apresentar T_m menor que nos casos onde N é grande. É importante notar que a temperatura de desnaturação T_m de sequências grandes ($N \rightarrow \infty$) corresponde à temperatura onde ocorre o anti cruzamento λ_1 com λ_2 , portanto a descontinuidade prevista pela equação (152) de fato ocorreu com o modelo.

4 Conclusão

Nosso modelo tridimensional foi desenvolvido utilizando várias aproximações que nos permitiram integrar analiticamente as variáveis θ e z . Com isso, obtemos um Hamiltoniano 1D expresso na equação (129) dependente da torção ω , muito importante para eliminar a degenerescência dos autovalores, fonte da divergência de $\langle r \rangle$ existente no modelo PB unidimensional [52]. Mostramos, analiticamente, que o anti cruzamento dos dois maiores autovalores é responsável pela transição abrupta em sequência longas, fato que se confirmou nos testes numéricos.

Testamos tais aproximações para uma sequência de dois pares de base e os resultados qualitativos são satisfatórios, pois os cálculos apontam para uma concordância entre os modelos aproximado e completo em regime de pequenos ângulos. Quando aplicamos o modelo em sequências grandes, uma transição semelhante à de primeira ordem é observada (figura 44), sendo muito mais abrupta do que nos casos dos outros modelos PB (figura 15) e sem a necessidade de utilizar potenciais anarmônicos adicionais. O fato de restringirmos ω a valores pequenos nos permite concluir que o uso do modelo será útil para temperaturas próximas à temperatura de desnaturação, já que nessa situação a hélice encontra-se muito próxima de uma estrutura planar. Acreditamos que seu uso principal será substituir hamiltonianos do tipo PB em cálculos de temperatura de desnaturação, uma vez que pode-se utilizar a técnica de integral de transferência já implementada nos outros modelos do tipo PB. Nosso modelo helicoidal é baseado em uma estrutura geométrica que leva em conta parâmetros como distâncias de equilíbrio nas ligações de hidrogênio e no empilhamento, semelhante à definição usada em dinâmica molecular [66].

O próximo passo desse trabalho será utilizar o modelo helicoidal homogêneo como base para a generalização e utilização em duplexos heterogêneos. A técnica para tal generalização é conhecida e é aplicada nos modelos PB 1D [52], portanto, entendemos que cabe fazer apenas pequenas adaptações para a nossa versão. Além disso, para tornar o modelo aplicável em problemas reais, precisamos otimizar o conjunto de parâmetros dos potenciais e acreditamos ser possível fazer isso com o uso da técnica da equivalência termodinâmica [53].

Referências

- [1] Rosalind E Franklin e Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.
- [2] Brenda Maddox. *Rosalind Franklin: The dark lady of DNA*. HarperCollins New York, 2002.
- [3] J. D. Watson e F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [4] Richard E Dickerson. [5] dna structure from a to z. In *Methods in enzymology*, volume 211, pages 67–111. Elsevier, 1992.
- [5] Clarisse Gravina Ricci. A, b e z-dna: caracterização conformacional e importância biológica. 2007.
- [6] Richard E Dickerson, Horace R Drew, Benjamin N Conner, Richard M Wing, Albert V Fratini, e Mary L Kopka. The anatomy of a-, b-, and z-dna. *Science*, 216(4545):475–485, 1982.
- [7] K Itakura, N Katagiri, CP Bahl, RH Wightman, e SA Narang. Improved triester approach for the synthesis of pentadecathymidylic acid. *Journal of the American Chemical Society*, 97(25):7327–7332, 1975.
- [8] Andrew H-J Wang, Gary J Quigley, Francis J Kolpak, James L Crawford, Jacques H Van Boom, Gijs van der Marel, e Alexander Rich. Molecular structure of a left-handed double helical dna fragment at atomic resolution. *Nature*, 282(5740):680–686, 1979.
- [9] Alexander Rich e Shuguang Zhang. Z-dna: the long road to biological function. *Nature Reviews Genetics*, 4(7):566–572, 2003.
- [10] Rosalind E Franklin e Raymond G Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740, 1953.

- [11] Thomas J Thamann, Richard C Lord, Andrew HJ Wang, e Alexander Rich. The high salt form of poly (dg-dc)· poly (dg-dc) is left-handed z-dna: Raman spectra of crystals and solutions. *Nucleic acids research*, 9(20):5443–5458, 1981.
- [12] Carl Schildkraut e Shneior Lifson. Dependence of the melting temperature of DNA on salt concentration. *Biopoly.*, 3(2):195–208, 1965.
- [13] Aimee J Marko, Simon S Smith, Eric D Siggia, Juan C Bustamante, e James Mussell. Entropic elasticity of lambda-phage dna. *Science (New York, NY)*, 1994.
- [14] Michelle D Wang, Hong Yin, Robert Landick, Jeff Gelles, e Steven M Block. Stretching dna with optical tweezers. *Biophysical journal*, 72(3):1335, 1997.
- [15] Carlos Bustamante, Zev Bryant, e Steven B Smith. Ten years of tension: single-molecule dna mechanics. *Nature*, 421(6921):423–427, 2003.
- [16] Amar Singh e Navin Singh. Effect of salt concentration on the stability of heterogeneous DNA. *Phys. A (Amsterdam, Neth.)*, 419:328–334, 2015.
- [17] Thomas T Perkins. Ångström-precision optical traps and applications. 2014.
- [18] Jean-Louis Mergny e Laurent Lacroix. Analysis of thermal melting curves. *Oligonucleotides*, 13(6):515–537, 2003.
- [19] V. A. Bloomfield. Statistical thermodynamics of helix coil transitions in biopolymers. *Am. J. Phys.*, 67(12):1212–1215, 1999.
- [20] Kathleen P Howard. Thermodynamics of DNA duplex formation: A biophysical chemistry laboratory experiment. *J. Chem. Educ.*, 77(11):1469, 2000.
- [21] Ross B Inman. Transitions of DNA homopolymers. *J. Mol. Biol.*, 9(3):624–637, 1964.
- [22] Tobias Mann, Richard Humbert, Michael Dorschner, John Stamatoyannopoulos, e William Stafford Noble. A thermodynamic approach to PCR primer design. *Nucleic Acids Research*, 37(13):e95–e95, 2009.

- [23] N. Haslam, N. Whiteford, Gerald Weber, Adam Prügel-Bennett, Jonathan W. Essex, e C. Neylon. Optimal probe length varies for targets with high sequence variation: Implications for probe library design for resequencing highly variable genes. *PLoS ONE*, 3:e2500, 2008. doi: <http://dx.doi.org/10.1371/journal.pone.0002500>.
- [24] Jef Hooyberghs, Marco Baiesi, Alessandro Ferrantini, e Enrico Carlon. Breakdown of thermodynamic equilibrium for DNA hybridization in microarrays. *Physical Review E*, 81(1):012901, 2010.
- [25] Célia Fonseca Guerra, F Matthias Bickelhaupt, Jaap G Snijders, e Evert Jan Baerends. Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment. *J. Am. Chem. Soc.*, 122(17):4117–4128, 2000.
- [26] Angelo Di Garbo. Anharmonic longitudinal motion of bases and dynamics of nonlinear excitation in DNA. *Biophysical chemistry*, 208:76–83, 2016.
- [27] Marco Zoli. Flexibility of short DNA helices under mechanical stretching. *Physical Chemistry Chemical Physics*, 18(26):17666–17677, 2016.
- [28] K. J. Breslauer, R Frank, H Blocker, e L. A. Marky. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA*, 83(11):3746–3750, 1986.
- [29] S Wolfsheimer e A K Hartmann. Minimum-free-energy distribution of RNA secondary structures: Entropic and thermodynamic properties of rare events. *Physical Review E*, 82(2):021902, 2010.
- [30] BH Zimm e JK Bragg. Theory of the one-dimensional phase transition in polypeptide chains. *The Journal of Chemical Physics*, 28(6):1246–1247, 1958.
- [31] John SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95(4):1460–1465, 1998. URL <http://www.pnas.org/cgi/content/abstract/95/4/1460>.
- [32] D. Poland e H. A. Scheraga. Occurrence of a phase transition in nucleic acids models. *J. Chem. Phys.*, 45(5):1464–1469, 1966.

- [33] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.
- [34] Sahin Buyukdagli, Michael Sanrey, e Marc Joyeux. Towards more realistic dynamical models for DNA secondary structure. *Chem. Phys. Lett.*, 419:434–438, 2006.
- [35] T. Das e S. Chakraborty. A generalized Langevin formalism of complete dna melting transition. *Europhys. Lett.*, 83:48003, 2008.
- [36] Ana-Maria Florescu e Marc Joyeux. Thermal and mechanical denaturation properties of a dna model with three sites per nucleotide. *The Journal of Chemical Physics*, 135(8):085105, 2011.
- [37] T. Dauxois, M. Peyrard, e A. R. Bishop. Entropy-driven DNA denaturation. *Phys. Rev. E*, 47(1):R44–R47, 1993. doi: 10.1103/PhysRevE.47.R44.
- [38] M. Peyrard e 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.
- [39] M. Peyrard. Nonlinear dynamics and statistical physics of DNA. *Nonlinearity*, 17:R1, 2004.
- [40] Amar Singh e Navin Singh. Pulling DNA: The effect of chain length on the mechanical stability of DNA chain. In *Macromolecular Symposia*, volume 357, pages 64–69. Wiley Online Library, 2015.
- [41] Gerald Weber. Sharp DNA denaturation due to solvent interaction. *Europhys. Lett.*, 73(5):806–811, 2006. doi: 10.1209/epl/i2005-10466-6.
- [42] CB Tabi, A Dang Koko, R Oumarou Doko, HP Ekobena Fouda, e TC Kofané. Modulated charge patterns and noise effect in a twisted DNA model with solvent interaction. *Physica A*, 442:498–509, 2016.
- [43] Arghya Maity, Amar Singh, e Navin Singh. Differential stability of DNA based on salt concentration. *European Biophysics Journal*, 46(1):33–40, 2017.

- [44] R. Tapia-Rojo, D. Prada-Gracia, J. J. Mazo, e F. Falo. Mesoscopic model for free-energy-landscape analysis of DNA sequences. *Phys. Rev. E*, 86(2):021908, 2012.
- [45] Marco Zoli. Twisting and bending stress in DNA minicircles. *Soft Matter*, 10(24):4304–4311, 2014.
- [46] W. Hidayat, A. Sulaiman, S. Viridi, e F. P. Zen. The viscous and external forces effect on the thermal denaturation of Peyrard-Bishop model. *Journal of Physical Chemistry & Biophysics*, 5(5):1, 2015.
- [47] Gerald Weber. Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures. *Nucleic Acids Res.*, 41:e30, 2013. doi: 10.1093/nar/gks964. URL <http://nar.oxfordjournals.org/content/41/1/e30>.
- [48] T. Dauxois e M. Peyrard. Entropy-driven transition in a one-dimensional system. *Phys. Rev. E*, 51(5):4027–4040, 1995.
- [49] Michael Stone e Paul Goldbart. *Mathematics for physics: a guided tour for graduate students*. Cambridge University Press, 2009.
- [50] Francesco Giacomo Tricomi. *Integral equations*, volume 5. Courier Corporation, 1957.
- [51] George B Arfken, Hans J Weber, e Frank E Harris. *Mathematical methods for physicists: A comprehensive guide*. Academic press, 2011.
- [52] Yong-Li Zhang, Wei-Mou Zheng, Ji-Xing Liu, e Y. Z. Chen. Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E*, 56(6):7100–7115, 1997.
- [53] Gerald Weber, Niall Haslam, Nava Whiteford, Adam Prügel-Bennett, Jonathan W. Essex, e Cameron Neylon. Thermal equivalence of DNA duplexes without melting temperature calculation. *Nat. Phys.*, 2:55–59, 2006. doi: 10.1038/nphys189.
- [54] M. Swart, C. F. Guerra, e F. M. Bickelhaupt. Hydrogen bonds of RNA are stronger than

- those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine. *J. Am. Chem. Soc.*, 126(51):16718–16719, 2004.
- [55] Jennifer L Wilcox e Philip C Bevilacqua. pK_a shifting in double-stranded RNA is highly dependent upon nearest neighbors and bulge positioning. *Biochem.*, 52(42):7470–7476, 2013.
- [56] Simona Cocco e Remi Monasson. Statistical mechanics of torque induced denaturation of DNA. *Phys. Rev. Lett.*, 83:5178–81, 1999.
- [57] Anil Kumar Dasanna, Nicolas Destainville, John Palmeri, e Manoel Manghi. Slow closure of denaturation bubbles in dna: Twist matters. *Physical Review E*, 87(5):052703, 2013.
- [58] F De los Santos, Omar Al Hammal, e Miguel A Muñoz. Simplified langevin approach to the peyrard-bishop-dauxois model of dna. *Physical Review E*, 77(3):032901, 2008.
- [59] Marc Joyeux e Sahin Buyukdagli. Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous DNA thermal denaturation. *Phys. Rev. E*, 72:051902, 2005.
- [60] Dirk Hennig e Juan FR Archilla. Stretching and relaxation dynamics in double stranded DNA. *Physica A*, 331(3-4):579–601, 2004.
- [61] D Lacitignola, G Saccomandi, e I Sgura. Parametric resonance in a mesoscopic discrete DNA model. *Acta Applicandae Mathematicae*, 132(1):391–404, 2014.
- [62] Maria Barbi, Simona Cocco, e Michel Peyrard. Helicoidal model for DNA opening. *Phys. Lett. A*, 253:358–369, 1999.
- [63] M. Barbi, S. Lepri, Michel Peyrard, e Nikos Theodorakopoulos. Thermal denaturation of a helicoidal DNA model. *Phys. Rev. E*, 68:061909, 2003.
- [64] Wilma K. Olson, Manju Bansal, Stephen K. Burley, Richard E. Dickerson, Mark Gershtein, Stephen C. Harvey Udo Heinemann, Xiang-Jun Lu, Stephen Neidle, Zippora Shak-

- ked Heinz Sklenar, Masashi Suzuki, Chang-Shung Tung, Eric Westhof Cynthia Wolberger, e Helen M. Berman. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, 313:229–237, 2001.
- [65] Abramo Hefez e Cecília de Souza Fernandez. Introdução à álgebra linear. *Rio de Janeiro: SBM*, 2012.
- [66] K. Drukker, G. Wu, e G. C. Schatz. Model simulations of DNA denaturation dynamics. *J. Chem. Phys.*, 114(1):579–590, 2001.
- [67] Mateus Rodrigues Leal e Gerald Weber. Sharp DNA denaturation in a helicoidal mesoscopic model. *Chem. Phys. Lett.*, 755:137781, 2020. doi: 10.1016/j.cplett.2020.137781. URL <https://www.sciencedirect.com/science/article/pii/S00092614203069651>.
- [68] Arghya Maity e Navin Singh. Melting of DNA in confined geometries. *Eur. Biophys. J.*, 49(7):561–569, 2020.
- [69] Enrique Maciá. Base-pairs' correlated oscillation effects on the charge transfer in double-helix B-DNA molecules. *Materials*, 13(22):5119, 2020.
- [70] Enrique Maciá. Synchronized oscillations in double-helix B-DNA molecules with mirror-symmetric codons. 2021. Preprints.org.

A Apêndice

O programa que fizemos para calcular $\langle r \rangle$ em função da temperatura usando integração pela quadratura de gauss legendre possui o seguinte código em C++:

```
//linha de comando
//compila
//g++ -O3 -o ym-int-num ym-int-num.cpp

#include <math.h>
#include <iostream>
#include <fstream>
#include <string>
#include <stdio.h>
#include <stdlib.h>

// parametros que definem as constantes da funcao a ser integrada
long double ev=1.6021895e-19;
long double Kb=1.380658e-23/ev;

long double D      = 0.4;
long double a      = 4.45;
long double k      = 0.04;
long double L0     = 3.5;
long double R0     = 1.0;

void ler_tabela_gauss_legendre (const int N, long double M[])
{
    char nome_arquivo[100];
    sprintf(nome_arquivo, "gauss-legendre/gauss-legendre-%d.dat", N);
    std::cout << "nome_do_arquivo=" << nome_arquivo << std::endl;
    std::ifstream arquivo;
    arquivo.open(nome_arquivo);
    for(int i=0; i< N*2; i+=2)
    {
        arquivo >> M[i] >> M[i+1];
    }
    arquivo.close();
}

main(int argc, char *argv[])
{
    // Define o numero de passos de integracao das variaves z theta r
    const int Nz = 20;
    const int Nt = 20;
    const int Nr = 500;

    //limites de integracao

    long double r20 = 0.0001;
    long double r2m = 50;
    long double r10 = 0.0001;
    long double r1m = 50;
```

```

long double the20 = -0.01;
long double the2m = 0.01;
long double the10 = -0.01;
long double the1m = 0.01;

long double z10 = -0.05;
long double z1m = 0.05;
long double z20 = 3.35;
long double z2m = 3.45;

//constantes das variaveis dos limites de integracao

long double c1z1 = ((z1m-z10)/2);
long double c2z1 = ((z1m+z10)/2);
long double c1z2 = ((z2m-z20)/2);
long double c2z2 = ((z2m+z20)/2);

long double c1t1 = ((the1m-the10)/2);
long double c2t1 = ((the1m+the10)/2);
long double c1t2 = ((the2m-the20)/2);
long double c2t2 = ((the2m+the20)/2);

long double c1r2 = ((r2m-r20)/2);
long double c2r1 = ((r1m+r10)/2);
long double c1r1 = ((r1m-r10)/2);
long double c2r2 = ((r2m+r20)/2);

long double M1[2*Nz], g11[2][Nz];
ler_tabela_gauss_legendre(Nz,M1);
    for(int i=0,j=0; i< Nz; i++, j+=2)
    {
        g11[0][i]=M1[j];
        g11[1][i]=M1[j+1];
    }

long double M2[2*Nt], g12[2][Nt];
ler_tabela_gauss_legendre(Nt,M2);
    for(int i=0,j=0; i< Nt; i++, j+=2)
    {
        g12[0][i]=M2[j];
        g12[1][i]=M2[j+1];
    }

long double M3[2*Nr], g13[2][Nr];
ler_tabela_gauss_legendre(Nr,M3);
    for(int i=0,j=0; i< Nr; i++, j+=2)
    {
        g13[0][i]=M3[j];
        g13[1][i]=M3[j+1];
    }

double T0 = atof(argv[1]);
double Tm = atof(argv[2]);
double pT = 5;

int passos=(Tm-T0)/pT;

```

```

std::ofstream saida;
char nome_saida[100];
sprintf(nome_saida, "rm-%g-%g.dat", T0, Tm);
saida.open(nome_saida);

for (long double T = T0; T <= Tm; T += pT)
{
long double betta = 1/(Kb*T);
long double I6 = 0;
long double J6 = 0;

for (int nr2 = 0; nr2 < Nr; nr2++)
{
long double tr2 = g13[0][nr2];
long double Ar2 = g13[1][nr2];
long double r2 = c1r2*tr2 + c2r2;
long double g6 = D*pow((exp(-a*(r2-R0))-1), 2);

long double I5 = 0;
long double J5 = 0;
for (int nr1 = 0; nr1 < Nr; nr1++)
{
long double tr1 = g13[0][nr1];
long double Ar1 = g13[1][nr1];
long double r1 = c1r1*tr1 + c2r1;

long double g5 = D*pow((exp(-a*(r1-R0))-1), 2);

long double I4 = 0, J4 = 0;
for (int nthe2=0; nthe2 < Nt; nthe2++)
{
long double tthe2 = g12[0][nthe2];
long double Athe2 = g12[1][nthe2];
long double the2 = c1t2*tthe2 + c2t2;
long double I3 = 0, J3 = 0;
for (int nthe1 = 0; nthe1 < Nt; nthe1++)
{
long double tthe1 = g12[0][nthe1];
long double Athe1 = g12[1][nthe1];
long double the1 = c1t1*tthe1 + c2t1;
long double coss = cos(the1-the2);
long double g4 = pow(r1, 2)+pow(r2, 2) - 2*r1*r2*coss;

long double I2 = 0, J2 = 0;
for (int nz2 = 0; nz2 < Nz; nz2++)
{
long double tz2 = g11[0][nz2];
long double Az2 = g11[1][nz2];
long double z2 = c1z2*tz2 + c2z2;

long double I1=0, J1=0;
for (int nz1 = 0; nz1 < Nz; nz1++)
{
long double tz1 = g11[0][nz1];
long double Az1 = g11[1][nz1];
long double z1 = c1z1*tz1 + c2z1;

```

```

    long double g3 = sqrt(pow(z1-z2,2)+g4);
    long double f1 = r1*r2*exp(-beta*(k*pow(g3-L0,2)+ (g5 + g6)/2))*exp(-
;

    I1 = I1 + Az1*f1;
    J1 = J1 + r1*Az1*f1;

    I2 = I2 + Az2*I1;
    J2 = J2 + Az2*J1;
    }
    I3 = I3 + Athe1*I2;
    J3 = J3 + Athe1*J2;
    }
    I4 = I4 + Athe2*I3;
    J4 = J4 + Athe2*J3;
    }
    I5 = I5 + Ar1*I4;
    J5 = J5 + Ar1*J4;
    }

    I6 = I6 + Ar2*I5;
    J6 = J6 + Ar2*J5;
    }
    long double rm = J6/I6;

    std::cout << rm << "\n" << T << std::endl;

saida << T << "\n" << rm << std::endl;

}

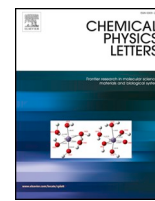
}

```

Artigo associado

Publicado em Junho de 2020 [67], e citado por Maity e Singh [68] e Maciá [69, 70].

Mateus Rodrigues Leal e Gerald Weber. Sharp DNA denaturation in a helicoidal mesoscopic model. *Chem. Phys. Lett.*, 755:137781, 2020. doi: 10.1016/j.cplett.2020.137781. URL <https://www.sciencedirect.com/science/article/pii/S00092614203069651>



Research paper

Sharp DNA denaturation in a helicoidal mesoscopic model

Mateus Rodrigues Leal, Gerald Weber*

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



HIGHLIGHTS

- A new 1D DNA Hamiltonian with twist angle dependence was obtained from a 3D model.
- We show the validity of the new Hamiltonian in the regime of small angles.
- Resulting melting transition are found to be very sharp, even first-order like in the limit of long sequences, without the need of additional non-linear potentials.
- Model parameters are compatible with other microscopic models and hydrogen bond parameters are of the same order of magnitude than those from ab-initio calculations.

ARTICLE INFO

Keywords:

DNA stability
Mesoscopic models
Peyrard Bishop model

Abstract: The Peyrard-Bishop DNA model describes the molecular interactions with simple potentials which allow efficient calculations of melting temperatures. However, it is based on a Hamiltonian that does not consider the helical twist or any other relevant molecular dimensions. Here, we start from a more realistic 3D model and work out several approximations to arrive at a new non-linear 1D Hamiltonian with a twist angle dependence. Our approximations were numerically compared to full 3D calculations, and established its validity in the regime of small angles. For long DNA sequences we obtain sharp, first-order-like melting, transitions.

1. Introduction

Statistical physics models of DNA using interaction potentials, instead of statistical weights, made a debut with Peyrard and Bishop [1]. This model introduces several simplifications that leaves only a single degree of freedom to integrate, transversal to the helical axis, and for this reason it is commonly referred to as a 1D model [2–4]. This model allows the calculation of the average base pair displacement, representative of the melting transition, and it was shown that there is an increasing strand opening as temperatures increase. However, this strand opening occurs only gradually with increasing temperature, which has motivated the search for additional potentials that could result in much sharper transitions [2,5,6]. The simplicity of the PB model provides a computational efficiency that outcompetes atomistic simulations for certain applications, such as describing melting in DNA [7]. Evidently, the increased efficiency comes at the expense of lack details describing the intramolecular interactions.

In recent years, our group used the mesoscopic Peyrard-Bishop (PB) model for calculating melting temperatures in numerous nucleic acids systems, for instance for deoxyinosine [8], GU mismatches in RNA [9] and DNA-RNA hybrids [10]. Many of our findings correlate well with

existing structural data from NMR and X-ray measurements providing a good level of validation for this theoretical approach. However, as discussed in some of our previous publications [11], the missing helicity and the unusual definition of intramolecular distances of the original PB model [1] makes it difficult to compare the results with microscopic models, especially to those of molecular dynamics (MD) and coarse-grained models [12]. This is of special interest for collisional MD that integrates the PB model to study melting transitions [13]. The use of an analytical 1D helicoidal Hamiltonian, preferably set in a similar framework as MD models [14], and benefiting from the efficient transfer integral (TI) technique for calculating the strand separation would be desirable as it may overcome some of the interpretative shortcomings of the PB model.

There have been several proposals for helicoidal Hamiltonians within the framework of TI partition function calculation of the PB model [15–17]. These models add the helical twist angle, but also add constraints that make it difficult to integrate analytically the partition function. The common approach is to fix beforehand the distance $z_{n+1} - z_n$ between consecutive base pairs, see Fig. 1, known as the helical rise distance h [18]. By fixing h , the radial distance r and the twist angle θ both need to be integrated numerically [15,19,20]. Unlike

* Corresponding author at: Departamento de Física, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil.
E-mail address: gweber@ufmg.br (G. Weber).

<https://doi.org/10.1016/j.cplett.2020.137781>

Received 26 May 2020; Received in revised form 3 July 2020; Accepted 7 July 2020

Available online 15 July 2020

0009-2614/ © 2020 Elsevier B.V. All rights reserved.

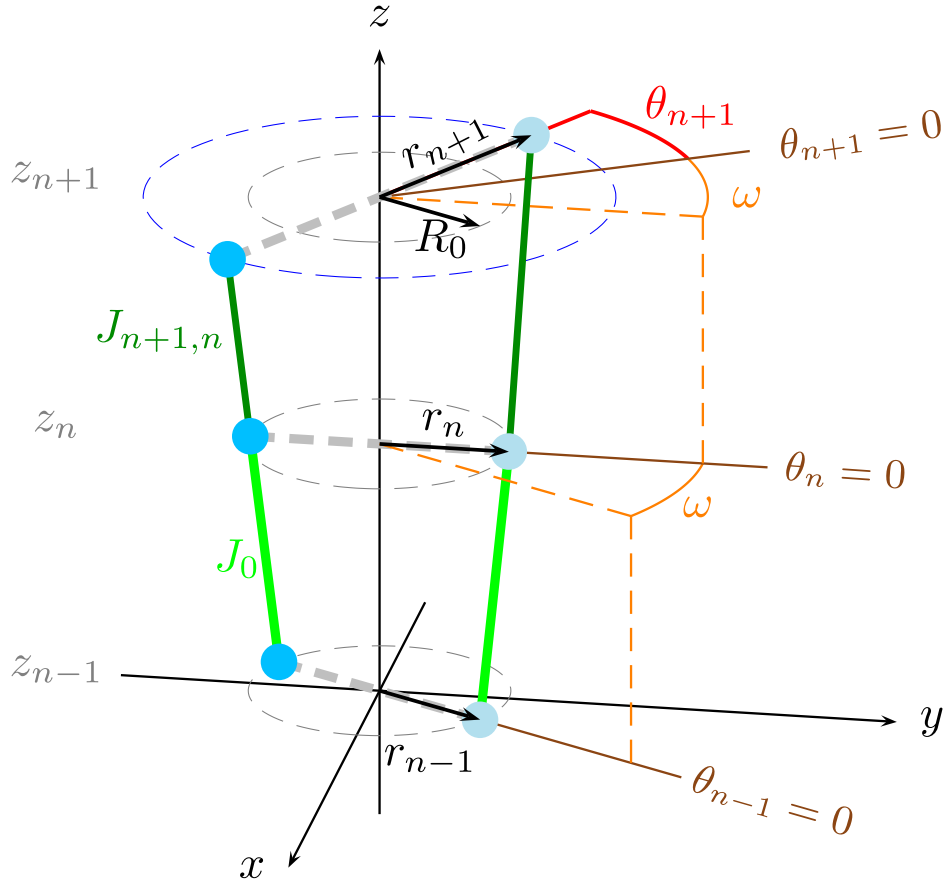


Fig. 1. Schematic diagram of the 3D model. Base pairs $n - 1$ and n are shown at their equilibrium positions, and the base pair $n + 1$ is rotated by θ_{n+1} and stretched from its equilibrium position to r_{n+1} . The twist angles ω (orange lines) define the offset between consecutive base pairs angle origins $\theta = 0$ (brown lines). Green lines show the stacking interaction between neighbouring bases, in particular the distances $J_{n+1,n}$ are shown as a dark green lines and J_0 are the equilibrium stacking distance shown as light green lines. The grey circles show the radius of the equilibrium distance and the blue circle shows the stretched distance r_{n+1} . Hydrogen bond distances are shown as thick dashed grey lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the PB model, the integration of the helicoidal Hamiltonian, restricted in this way, does not result in an analytical 1D Hamiltonian. Other helicoidal models based on the PB Hamiltonian that do not calculate melting transitions or do not use the transfer integral method, such as Refs. 21–26, are not covered here.

Our aim is to adapt a 3D Hamiltonian, with distances and angles as shown in Fig. 1, and obtain an analytical expression for a 1D Hamiltonian with included twist angle dependence. A constant twist angle ω defines the origin of the angles θ at each base pair. The setup of Fig. 1 follows closely that of B-DNA, which from crystallographic measurements it is known to have a tilt angle of -0.1° and roll of 0.6° per base pair step [18]. In other words, for short DNA sequences there is no appreciable bending and the configuration shown in Fig. 1 is justified.

We integrate the 3D partition function by carefully introducing approximations and restrictions and arrive at a new 1D Hamiltonian with radial distance and twist angle dependence. For small twist angles, the resulting melting transitions are sharp first-order-like with strong discontinuity of the strand opening for very long DNA sequences. On the other hand, as soon as the twist angles are increased these transitions rapidly lose their strength. To evaluate the impact of the approximations, we numerically integrated the configurational part of the partition function of the 3D Hamiltonian. We repeat these numerical integrations also applying similar restrictions that were used for the 1D Hamiltonian. These numerical tests show that the results from the helicoidal 1D Hamiltonian are qualitatively similar to the full 3D model within the regime of small angles.

2. Model

The configurational part of the classical partition of a oligonucleotide duplex composed of N base-pairs is written in terms of the polar cylindrical coordinates z , r and θ

$$Z_{r\theta z} = \Gamma^N \int \prod_{n=1}^N dz_n d\theta_n r_n dr_n \exp[-\beta U_{n,n-1}(z_n, \theta_n, r_n, z_{n-1}, \theta_{n-1}, r_{n-1})] \quad (1)$$

where $\beta = 1/(k_B T)$, k_B is the Boltzmann constant and T the absolute temperature. Γ is a density factor, which is taken here as a reciprocal unit of volume, such that $Z_{r\theta z}$ becomes adimensional. U is the configurational part of model Hamiltonian and is a function of the (z, θ, r) positions of two consecutive base pairs. The customary periodic boundary condition, where the last base-pair interacts with the first, is represented by the potential $U_{1,N}$. The average radius $\langle r_k \rangle$, representing the intra-strand opening, can be calculated as follows

$$\langle r_k \rangle = \frac{\Gamma^N}{Z_{r\theta z}} \int \prod_{n=1}^N r_n dz_n d\theta_n r_n dr_n \exp[-\beta U_{n,n-1}(z_n, \theta_n, r_n, z_{n-1}, \theta_{n-1}, r_{n-1})] \quad (2)$$

For the case where all model parameters are the same at each site k , we have $\langle r \rangle = \langle r_1 \rangle = \dots = \langle r_N \rangle$.

The origins of the θ angles of consecutive base pairs are offset by a fixed twist angle ω between base pair steps, see Fig. 1, which allows the use of a single integration limit for all angle variables, that is,

$\theta_n \in [-\Theta, \Theta]$. For r_n the integrations are taken within the limits $r_n \in [0, b]$. For z we integrate within a region $\pm \zeta$ around the rise distance h_0 such that the limit is taken as

$$z_n \in [(n-1)h_0 - \zeta, (n-1)h_0 + \zeta] \quad (3)$$

the partition function is then written with explicit integration limits as

$$Z_{r\theta z} = \Gamma^N \int_0^b \int_{-\Theta}^{\Theta} \int_{(n-1)h_0 - \zeta}^{(n-1)h_0 + \zeta} \prod_{n=1}^N dz_n d\theta_n r_n dr_n \times \exp[-\beta U_{n,n-1}(z_n, \theta_n, r_n, z_{n-1}, \theta_{n-1}, r_{n-1})] \quad (4)$$

where each integration symbol implies N -uple integrals. The interaction potential U is divided into stacking interactions $W_{n,n-1}$ and base-pair interactions V_n ,

$$U_{n,n-1}(z_n, \theta_n, r_n, z_{n-1}, \theta_{n-1}, r_{n-1}) = V_n(r_n) + W_{n,n-1}(z_n, \theta_n, r_n, z_{n-1}, \theta_{n-1}, r_{n-1}) \quad (5)$$

In terms of the polar cylindrical coordinates and considering the 3D scheme shown in Fig. 1, the base-pair interaction potential is solely a function of r_n , that is $V_n(r_n)$ and it brings no difficulty for the integration of Eq. (4). Here we will use the Morse potential

$$V_n(r_n) = D[e^{-a(r-R_0)} - 1]^2 \quad (6)$$

where D is the depth of the potential, a the width and R_0 an equilibrium distance. The stacking interaction W however depends on all coordinates and links to consecutive base-pairs n and $n-1$, which is the main point of difficulty for a full algebraic integration of the partition function, Eq. (4). Therefore, our efforts will centre on the handling of the 3D stacking potential and, unlike the base-pair potential, the specific form of this potential is a crucial aspect of the theoretical method. Here, we use the harmonic potential as stacking interaction potential, this is the form used in the original PB model [1] which has provided good results for parametrization in heterogeneous DNA [11]. The potential between neighbouring bases n and $n-1$ is given by

$$W_{n,n-1} = \frac{k}{2}(J_{n,n-1} - J_0)^2, \quad (7)$$

where $J_{n,n-1}$, shown as a green line in Fig. 1, is the distance between two bases belonging to the same strand. J_0 is the equilibrium distance and k the elastic constant. In polar cylindrical coordinates z , r and θ , shown in Fig. 1, the distance $J_{n,n-1}$ is written as

$$J_{n,n-1} = \sqrt{\Delta z_{n,n-1}^2 + f_{n,n-1}^2} \quad (8)$$

where $f_{n,n-1}$ is the xy -projection

$$f_{n,n-1} = \sqrt{r_n^2 + r_{n-1}^2 - 2r_n r_{n-1} \cos(\omega + \theta_n - \theta_{n-1})} \quad (9)$$

The equilibrium distance between the two consecutive base-pairs along the z -axis is h_0 , corresponding to the rise distance and ω is the structural twist angle [18]. For simplicity, we will assume that both bases at the n th site are at the same distance in regard to the z axis, that is, they move symmetrically with respect to the helical z axis. While this may seem overly restrictive, we have shown that for the classical partition function in the PB model this means that the elastic constant is simply the average of the elastic parameters of each strand [27]. Therefore, the elastic constant k is the equivalent constant of the two springs to each side of the duplex strand.

We now expand Eq. (8) to first order of $\Delta z_{n,n-1}^2$

$$J_{n,n-1} \approx \Delta z_{n,n-1} \left[1 + \frac{1}{2} \frac{f_{n,n-1}^2}{\Delta z_{n,n-1}^2} \right]. \quad (10)$$

To higher orders of $\Delta z_{n,n-1}^2$ the remaining equations become quite complicated. Therefore, for the sake of the discussion, we will present here only the simpler development following the first order expansion without loss of generality, and show the more complicated expansion to second order in supplementary equations (S1–S3). We now use the

additional restriction

$$\Delta z_{n,n-1} \approx J_0 \quad (11)$$

which is similar as used by other authors [15,19,20]. However, the crucial difference here is that we apply it after the expansion of Eq. (10), as it enables us to carry out the remaining integrations and arrive at an analytical form for the 1D Hamiltonian, which is the aim of this work. After integration in z , and the partition function simplifies to

$$Z_{r\theta z} = \Gamma(2\zeta)^N \int_0^b \int_{-\Theta}^{\Theta} \prod_{n=1}^N dr_n d\theta_n r_n e^{-\beta V(r_n)} \exp\left(-\beta k \frac{f_{n,n-1}^4}{8J_0^2}\right) \quad (12)$$

For the angle integration we will use $\theta_n - \theta_{n-1} \ll \omega$, and the approximation

$$\cos(\omega + \theta_n - \theta_{n-1}) \approx 1 - \frac{\omega^2}{2} \quad (13)$$

Note that a small difference $\theta_n - \theta_{n-1}$ does not imply in a flattened helix, since the angles are always offset by the helical twist ω , see Fig. 1. Integrating over θ , we arrive at the final approximated form of the partition function, after rearranging terms to symmetrize the integrand function

$$Z_{r\theta z}^{app} = \Gamma^N (4\zeta\Theta)^N \int_0^b \prod_{n=1}^N dr_n \sqrt{r_n r_{n-1}} \exp\left\{-\frac{\beta}{2}[V(r_n) + V(r_{n-1})]\right\} \times \exp\left\{-\frac{\beta k}{8J_0^2}[(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^2\right\} \quad (14)$$

Note that the fluctuations along and around the z -axis are given by ζ and Θ , respectively, which are now outside the remaining integration, therefore those factors will simply cancel out when calculating expectation values, Eq. (2).

The remaining variable to integrate is in r_n which can be handled by the transfer integral technique where the kernel is

$$K(x, y) = (xy)^{1/2} e^{-\frac{\beta}{2}[V(x)+V(y)]} \exp\left\{-\frac{\beta k}{8J_0^2}[(x-y)^2 + \omega^2 xy]^2\right\} \quad (15)$$

In effect, this is now equivalent to a one-dimensional radial Hamiltonian with a twist angle dependence

$$U(r_n, r_{n-1}) = V(r_n) + \frac{k}{8J_0^2} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^2 \quad (16)$$

The approximated partition function Eq. (14) can be evaluated via the transfer integral (TI) technique [1,28]. In this technique the kernel is discretized over M points in the interval $[0, b]$ and the partition function becomes

$$Z_{TI} = \Gamma^N (4\zeta\Theta)^2 \sum_{i=1}^M \lambda_i^N \quad (17)$$

The average radius $\langle r \rangle_{TI}$ is calculated as

$$\langle r \rangle_{TI} = \frac{\sum_{i=1}^M \lambda_i^N \int_0^b r |\phi_i|^2 dr}{\sum_{i=1}^M \lambda_i^N} \quad (18)$$

where ϕ_i are the eigenfunctions. For details of this procedure see Refs. 1,28,6. For the limit $N \rightarrow \infty$ this further simplifies to

$$\lim_{N \rightarrow \infty} \langle r \rangle_{TI} = \int_0^b r |\phi_1|^2 dr \quad (19)$$

where ϕ_1 is the eigenfunction with the highest eigenvalue λ_1 [1]. We will refer to the approximation calculated by the TI technique as T1 and T2, for the first and second order expansion of Eq. (8), kernels Eqs. (15) and (S4), respectively.

2.1. Numerical tests

Here, we will compare numerically the approximated Eq. (17) to the

fully integrated partition function Eq. (4). To our knowledge, the numerical evaluation of the 3D Hamiltonian, Eq. (4), has not been performed. One possible reason for this is that the numerical effort scales with N^3 . Even for the smallest possible number of base pairs, $N = 2$, it has taken us of the order of days with parallel processing. Therefore, we are limited to $N = 2$ for the evaluation of Eq. (4). On the other hand, the TI solution Eq. (17) is valid for sequences of any length N .

We will keep the periodic boundary condition, which may seem odd for a sequence of length of just $N = 2$, however there is no loss of generality for the results presented here. The reason for this is that a sequence of length $N = 2$ has two elastic constants, $k_{1,2} = k_{2,1} = k$, where the last one represents the periodic boundary condition. The open boundary condition is simply setting $k_{1,2} = k$ and $k_{2,1} = 0$ [28], which for $N = 2$ turns out to be the exact equivalent of maintaining the periodic boundary condition and setting $k_{1,2} = k_{2,1} = k/2$.

We designated the partition function calculated from Eq. (4) as Z_C , where C stands for complete,

$$Z_C = Z_{r\phi z}[N = 2, b, \zeta, \Theta] \quad (20)$$

Furthermore, we calculate Eq. (20) by adding the restrictions of Eqs. (11) and (13), which we called the restricted (R) calculation, which is a subset of the Z_C calculation,

$$Z_R = Z_C \left[\Delta z_{n,n-1} \approx J_0; \cos(\omega + \theta_n - \theta_{n-1}) \approx 1 - \frac{\omega^2}{2} \right] \quad (21)$$

and it is expected that the TI calculations should be close to R. Details of the numerical integrations are given in supplementary Section (S1).

Unless noted otherwise, for the numerical tests we used the following parameters: $D = 0.2$ eV, $a = 42.5$ nm $^{-1}$, $k = 4$ eV/nm 2 , $J_0 = 0.7$ nm, corresponding to a homogeneous oligonucleotide sequence, and were largely chosen to highlight the main differences in the integration methods. The value for J_0 was adapted from Ref. 15. The equilibrium distance was taken as $R_0 = 0.1$ nm, as r represents half the distance between the base pairs, this corresponds to a hydrogen equilibrium bond distance of $2R_0 = 0.2$ nm.

3. Results and discussion

The dependence of the average radius $\langle r \rangle$ as function of temperature is shown in Fig. 2a for the numerical tests C, R, T1 and T2. In all cases, the denaturation curves exhibit the characteristic sigmoidal shape of the melting transition that has been the characteristic of the Peyrard-Bishop model [1]. The approximated calculation to first order expansion, T1, underestimates the average radius when compared to the C and R calculations, especially as temperature increases. The restrictions of Eqs. (11) and (13) do represent a substantial part of this reduction, as shown by the differences between the C and R calculations. This is to be

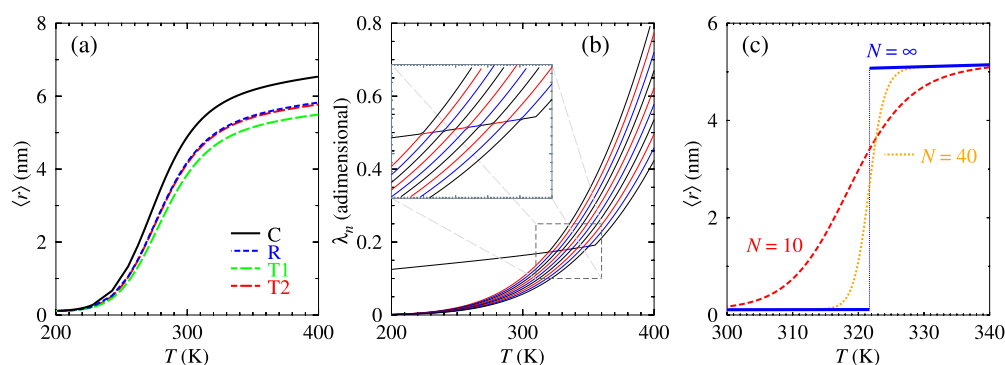


Fig. 2. Panel (a): calculated average radius $\langle r \rangle$ as function of temperature for calculations of type C (black curve), R (blue short dashed), T1 (green long dashed) and T2 (red short dashed). Panel (b): 10 highest eigenvalues λ_n for T1 (T2 in Fig. S1), the inset shows a zoom-into highlight the anti-crossings. Panel (c): T1 average radius $\langle r \rangle$ for sequence length $N = 10$ (red dashed), 40 (orange dotted) and tending to ∞ (thick blue), the vertical dotted blue line shows the discontinuity for $N \rightarrow \infty$ within resolution of $\Delta T = 0.01$ K. Integration limits are $\zeta = 10^{-2}$ nm, $\omega = 0.05$ rad, $\Theta = 0.01$ rad and $b = 20$ nm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

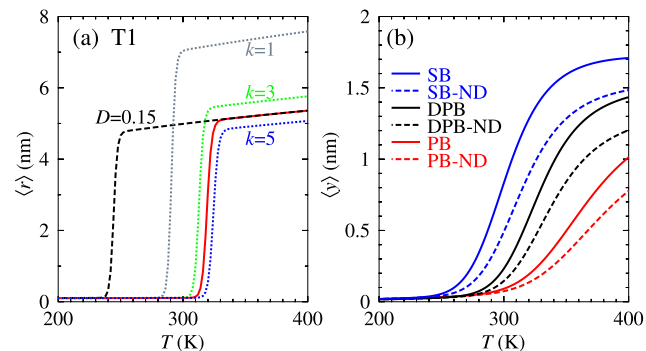


Fig. 3. Calculated average radius $\langle r \rangle$ for T1 (panel a, for T2 see Fig. S2) and calculated 1D average displacement $\langle y \rangle$ (panel b) as function of temperature for sequences of size $N = 25$. Panel (a): Morse parameters are $D = 0.2$ eV for all curves, except dashed black curve for which $D = 0.15$ eV was used; Stacking parameters are $k = 4$ eV/nm 2 for the red solid curve and black dashed curve, the remaining dotted curves are indicated in eV/nm 2 . Parameters for the T1 were $\zeta = 10^{-2}$ nm, $\omega = 0.05$ rad, $\Theta = 0.01$ rad and $b = 20$ nm. Panel (b): parameters for the 1D DPB model as in Ref. 33, and for PB and SB as in Ref. 29. Also shown as dashed curves, with suffix ND (non-divergent), are the calculations with added angle $\omega = 0.01$ rad as in Eq. (23).

expected as all three approximations, Eqs. (10, 11, 13), essentially limit the scope of the integration thus resulting in smaller $\langle r \rangle$. However, when we use the second order expansion T2, the result is very close to the R calculation, therefore the differences between T1 and R are only due to the order of the expansion of Eq. (8). The spectrum of λ_n for T1 is shown in Fig. 2b (see Fig. S1 for T2) and displays the characteristic anti-crossing between successive eigenvalues [5], which is highlighted in the zoomed-in inset. Unlike the spectra of the 1D models [29] where the eigenvalues have a substantial gap at the anti-crossings, here in Fig. 2b the gap between λ_1 and λ_2 is only 0.00042 or 0.24%. As the sequence length N increases, the transition becomes increasingly abrupt, as shown in Fig. 2c. In the limit of $N \rightarrow \infty$, see Eq. (19), a discontinuity is observed within resolution of $\Delta T = 0.01$ K. Barbi et al. [19] observed a similar transition discontinuity at $T = 350.74$ K, using their model parameters and a twist angle $\omega = 0.05$ rad, we obtain 350.60 K, which is in good agreement in view of the approximations that we are using.

For moderate sequence lengths, for instance $N = 25$, the helicoidal model already shows much steeper transitions (Fig. 3a) than other PB-type models, Fig. 3b. Some examples of different model parameters are shown in Fig. 3a. Varying the Morse potential D changes the temperature where the transitions occurs but not the $\langle r \rangle$ at high temperatures. The stacking parameter k on the other hand has an influence on both the onset of the transition and the high temperature value of $\langle r \rangle$.

To understand the differences between the helicoidal and PB-type models it is instructive to look at the symmetrized kernel Eq. (15) used for the T1 calculation, see supplementary Eq. (S4) for T2, and compare it to the PB kernel [1]

$$K_{PB}(x, y) = e^{-\frac{\beta}{2}[V(x)+V(y)]} \exp\left[-\beta k \frac{(x-y)^2}{2}\right] \quad (22)$$

One important difference is the $(x-y)$ to the fourth order in stacking term of Eq. (15), instead of second order for the PB kernel [1]. Therefore, the harmonic 3D stacking seemingly maps into an anharmonic stacking term in the helicoidal model. However, in our tests with the helicoidal model, such a fourth power term is not the main cause of a steep transition, although it has an important influence on which temperature the transition starts and how large $\langle r \rangle$ becomes at higher temperatures. What actually ensures the abrupt rise of $\langle r \rangle$ is the $(xy)^{1/2}$ factor which comes from the integration in r , which does not exist in the PB model. Supplementary Fig. S3 shows that by removing $(xy)^{1/2}$ the sigmoidal increase of $\langle r \rangle$ is no longer present. It also has an effect on magnitude of the potential parameters. For instance, the D we used to obtain a transition at higher temperatures is much closer to actual energies of the hydrogen bonds used in MD and quantum mechanical calculations, around 0.15 eV per hydrogen bond [14,30], whereas for the PB model these potentials are typically an order of magnitude smaller [11]. The elastic constants k are similar in magnitude to those already in use in the PB model and compare within order of magnitude to those used in MD, 4 to 10 eV/nm² [31,32]. The last factor in Eq. (15) contains the twist angle ω which plays a crucial role in preventing the divergence in the integration, we will discuss this in more detail next.

All PB-type models suffer from a numerical divergence, this is becomes especially apparent for the anharmonic Dauxois-Peyrard-Bishop (DPB) model and was discussed in detail by Zhang et al. [28]. One tentative approach to circumvent this divergence was to add a twist angle ω to Eq. (22) [7],

$$(x-y)^2 \rightarrow (x^2 - 2xy\cos\omega + y^2) \quad (23)$$

which mimics a small out of plane angle. This procedure avoids the divergence for any PB model [29] but also reduces the steepness of the transition, see dashed curves Fig. 3b. In general, the solvent-barrier (SB) model [29], another PB-type Hamiltonian, has a much steeper increase of the displacement than the anharmonic DBP [33] or the original harmonic PB model [1]. Some authors have used Eq. (23) as a twist angle model with large angles and basically observe the suppression of

the transition steepness [34]. The helicoidal model also shows the divergence problem if the twist angle is zero, $\omega = 0$, as shown in Fig. 4. The radius $\langle r \rangle$ diverges much more strongly than the partition function $Z_{r\theta z}$ due to the additional variable r in the integration of Eq. (2). Therefore the onset of the divergence for $\langle r \rangle$, Fig. 4b, occurs at a much shorter b than for $Z_{r\theta z}$, Fig. 4a. The divergence appears equally for the C and TI calculations, and consequently is not a particularity introduced by the approximations or by the transfer integral technique. Setting the twist angle ω to a non-zero value, no matter how small, removes the divergence entirely and therefore brings some justification to the similar approach used in the PB model, Eq. (23) [29]. When we increase the upper limit b we observe that both the partition function and radius, Fig. 4, converge to constant values for $b > 12$ nm, including at higher temperatures, see supplementary Fig. S4. Therefore, by using this type of analysis we ensure that the results do not depend on integration cut-off values.

The average radius $\langle r \rangle$ follows in general that of the restricted numerical integration R. Deviations of T1, T2 and R, from the unrestricted calculation C, become larger if we move away from the conditions where the approximations are valid, which is for small angles ($\Theta < 0.01$ rad) and small longitudinal displacements ($\zeta < 2 \times 10^{-2}$ nm), see Fig. 5a,b. The limit Θ of the angle θ and the upper limit ζ for the z variable both appear as constant factors in Eq. (14) and therefore are cancelled in the calculation of the average radius. As a consequence, the average radius is constant for θ and ζ , for the approximated calculation as shown in Fig. 5a,b. The same happens for R, which validates the T1 and T2 within these restriction. For the twist angle ω we observe a progressive reduction of the average radius $\langle r \rangle$ after $\omega = 0.02$ rad. For larger angles, $\langle r \rangle$ tends towards the equilibrium radius r_0 in all cases, which is consistent with the idea that the strands can not separate without unwinding the helix.

The analytical form of the Hamiltonian Eq. (16) and its associated TI kernel Eq. (15) allows for a straightforward use in established PB calculations and implementations [35]. This means that the new model can now be adapted to heterogeneous sequences and parameterised from melting temperatures in a similar way as for the PB model in DNA and RNA [11]. The improved compatibility of the model with MD means that we are able to use similar model parameters and compare the displacements $\langle r \rangle$ to hydrogen bond lengths resulting from atomistic simulations [31,14].

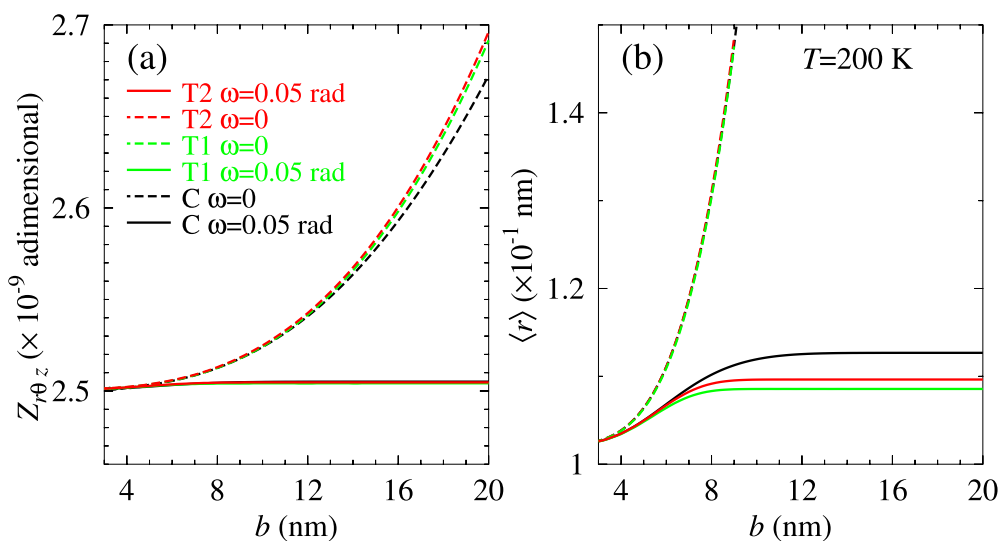


Fig. 4. Calculated (a) configurational part of the partition function $Z_{r\theta z}$ and (b) average radius $\langle r \rangle$ as function of the upper limit b of the integration variable r , at 200 K, for C (black), T1 (green) and T2 (red). Full curves are for a twist angle of $\omega = 0.05$ rad and dashed curves for $\omega = 0$. Integration limits are $\zeta = 10^{-2}$ nm, $\Theta = 0.01$ rad. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

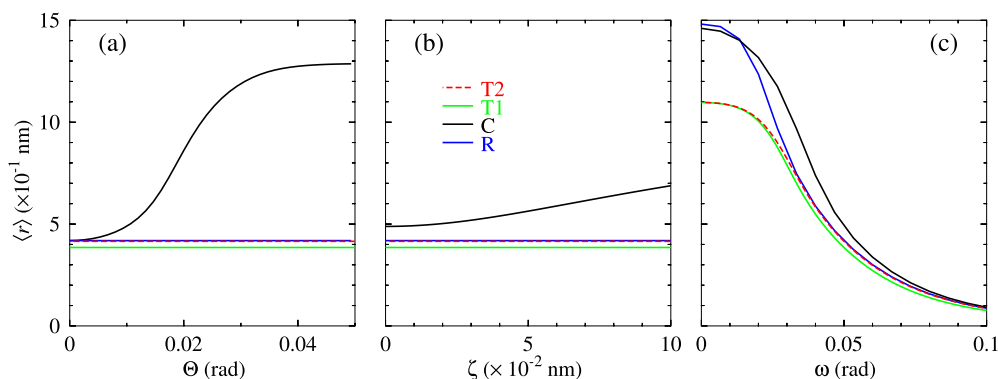


Fig. 5. Average radius $\langle r \rangle$ as function of the upper limit (a) Θ , (b) ζ and (c) ω , at temperature 300 K and $\omega = 0.05$ rad; for C (black), R (blue), T1 (green) and T2 (red). Integration limits are (a,b) $b = 20$ nm, (c) $b = 15$ nm, and (a) $\zeta = 10^{-2}$ nm or (b) $\Theta = 0.01$ rad. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Conclusions

We have developed and tested several approximations that allow the 3D Hamiltonian to be analytically integrated and resulted in a new 1D Hamiltonian with twist angle dependence. A first-order-like transition is observed, much steeper than for any PB model. This transition arises naturally in the helicoidal model, without the need of additional anharmonic potentials.

The results of the new helicoidal model, when compared to the restricted and unrestricted 3D calculations, points to a quantitative agreement in regime of small integration angles. Therefore, this approximated model, in particular represented by the Hamiltonian of Eq. (16), is expected to be useful for situations where the DNA helix is completely unwound. This is typically the case close to the temperature of DNA denaturation. We believe that its primary use will be for replacing PB-like Hamiltonians in melting temperature calculations [7], as it can be used within the framework of the TI method that already exists for the PB models [35]. The helicoidal model considers a similar structural definition as used in molecular dynamics [14], which enables the use of compatible parameters, such as hydrogen bond equilibrium distances. In addition, we showed that the Morse potential parameters are now much closer to those that are obtained from quantum mechanical calculations [30].

CRedit authorship contribution statement

Mateus Rodrigues Leal: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Visualization. **Gerald Weber:** Conceptualization, Methodology, Supervision, Writing - original draft, Visualization, Project administration.

Declaration of Competing Interest

None.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.cplett.2020.137781>.

References

- [1] M. Peyrard, A.R. Bishop, Statistical Mechanics of a Nonlinear Model for DNA denaturation, *Phys. Rev. Lett.* 62 (23) (1989) 2755–2757.
- [2] T. Dauxois, Dynamics of breather modes in a nonlinear “helicoidal” model of DNA, *Phys. Lett. A* 159 (8–9) (1991) 390–395.
- [3] G. Kalosakas, K. Rasmussen, A. Bishop, Non-exponential decay of base-pair opening fluctuations in DNA, *Chem. Phys. Lett.* 432 (1) (2006) 291–295.
- [4] S. Buyukdagli, M. Joyeux, Mapping between the order of thermal denaturation and the shape of the critical line of mechanical unzipping in one-dimensional DNA models, *Chem. Phys. Lett.* 484 (4) (2010) 315–320.
- [5] M. Peyrard, T. Dauxois, DNA melting: A phase transition in one dimension, *Math. Comput. Simul.* 40 (3–4) (1996) 305–318.
- [6] G. Weber, N. Haslam, J.W. Essex, C. Neylon, Thermal Equivalence of DNA Duplexes for Probe Design, *J. Phys.: Condens. Matter* 21 (2009a) 034106, doi:10.1088/0953-8984/21/3/034106.
- [7] G. Weber, N. Haslam, N. Whiteford, A. Prügel-Bennett, J.W. Essex, C. Neylon, Thermal Equivalence of DNA Duplexes Without Melting Temperature Calculation, *Nat. Phys.* 2 (2006) 55–59, <https://doi.org/10.1038/nphys189>.
- [8] R.V. Maximiano, G. Weber, Deoxyinosine Mismatch Parameters Calculated with a Mesoscopic Model Result in Uniform Hydrogen Bonding and Strongly Variable Stacking Interactions, *Chem. Phys. Lett.* 631–632 (2015) 87–91, <https://doi.org/10.1016/j.cplett.2015.04.045>.
- [9] T.D. Amarante, G. Weber, Evaluating hydrogen bonds and base stackings of single, tandem and terminal GU in RNA mismatches with a mesoscopic model, *J. Chem. Inf. Model.* 56 (1) (2016) 101–109, <https://doi.org/10.1021/acs.jcim.5b00571> URL <https://doi.org/10.1021/acs.jcim.5b00571>.
- [10] E.d.O. Martins, V.B. Barbosa, G. Weber, DNA/RNA hybrid mesoscopic model shows strong stability dependence with deoxypyrimidine content and stacking interactions similar to RNA/RNA, *Chem. Phys. Lett.* 715C (2019) 14–19, <https://doi.org/10.1016/j.cplett.2018.11.015>.
- [11] G. Weber, J.W. Essex, C. Neylon, Probing the Microscopic Flexibility of DNA from Melting Temperatures, *Nat. Phys.* 5 (2009) 769–773, <https://doi.org/10.1038/nphys1371>.
- [12] T.E. Ouldridge, A.A. Louis, J.P. Doye, Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model, *J. Chem. Phys.* 134 (8) (2011) 02B627.
- [13] I. Likhachev, V. Lakhno, The direct investigation of DNA denaturation in Peyrard-Bishop-Dauxois model by molecular dynamics method, *Chem. Phys. Lett.* 727 (2019) 55–58.
- [14] K. Drukker, G. Wu, G.C. Schatz, Model simulations of DNA denaturation dynamics, *J. Chem. Phys.* 114 (1) (2001) 579–590.
- [15] S. Cocco, R. Monasson, Statistical mechanics of torque induced denaturation of DNA, *Phys. Rev. Lett.* 83 (1999) 5178–5181.
- [16] M. Barbi, S. Cocco, M. Peyrard, Helicoidal model for DNA opening, *Phys. Lett. A* 253 (1999) 358–369.
- [17] S. Cocco, R. Monasson, Theoretical study of collective modes in DNA at ambient temperature, *J. Chem. Phys.* 112 (22) (2000) 10017–10033.
- [18] W.K. Olson, et al., A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry, *J. Mol. Biol.* 313 (2001) 229–237.
- [19] M. Barbi, S. Lepri, M. Peyrard, N. Theodorakopoulos, Thermal denaturation of a helicoidal DNA model, *Phys. Rev. E* 68 (2003) 061909.
- [20] T. Michael, Y. Van de Peer, Helicoidal transfer matrix model for inhomogeneous DNA melting, *Phys. Rev. E* 73 (1) (2006) 011908.
- [21] C.B. Tabi, A. Mohamadou, T.C. Kofané, Modulational instability and exact soliton solutions for a twist-opening model of DNA dynamics, *Phys. Lett. A* 373 (29) (2009) 2476–2483.
- [22] S. Behnia, M. Panahi, A. Akhshani, A. Mobaraki, Mean Lyapunov exponent approach for the helicoidal Peyrard-Bishop model, *Phys. Lett. A* 375 (41) (2011) 3574–3578.
- [23] G. Torrellas, E. Maciá, Twist–radial normal mode analysis in double-stranded DNA chains, *Phys. Lett. A* 376 (45) (2012) 3407–3410.
- [24] M. Zoli, End-to-end distance and contour length distribution functions of DNA

- helices, *J. Chem. Phys.* 148 (21) (2018) 214902.
- [25] S. Zdravković, D. Chevizovich, A.N. Bugay, A. Maluckov, Stationary solitary and kink solutions in the helicoidal Peyrard-Bishop model of DNA molecule, *Chaos* 29 (5) (2019) 053118.
- [26] S.K. Nomidis, E. Skoruppa, E. Carlon, J.F. Marko, Twist-bend coupling and the statistical mechanics of the twistable wormlike-chain model of DNA: Perturbation theory and beyond, *Phys. Rev. E* 99 (3) (2019) 032414.
- [27] E.d.O. Martins, G. Weber, An asymmetric mesoscopic model for single bulges in RNA, *J. Chem. Phys.* 147 (155102) (2017), <https://doi.org/10.1063/1.5006948>.
- [28] Y.-L. Zhang, W.-M. Zheng, J.-X. Liu, Y.Z. Chen, Theory of DNA melting based on the Peyrard-Bishop model, *Phys. Rev. E* 56 (6) (1997) 7100–7115.
- [29] G. Weber, Sharp DNA denaturation due to solvent interaction, *Europhys. Lett.* 73 (5) (2006) 806–811, <https://doi.org/10.1209/epl/i2005-10466-6>.
- [30] K. Wendler, J. Thar, S. Zahn, B. Kirchner, Estimating the hydrogen bond energy, *J. Phys. Chem. A* 114 (35) (2010) 9529–9536.
- [31] F. Zhang, M.A. Collins, Model simulation of DNA dynamics, *Phys. Rev. E* 52 (4) (1995) 4217–4224.
- [32] S. Mukherjee, S. Kailasam, M. Bansal, D. Bhattacharyya, Stacking interactions in RNA and DNA: Roll-slide energy hyperspace for ten unique dinucleotide steps, *Biopolymers* 103 (3) (2015) 134–147, <https://doi.org/10.1002/bip.22566> URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.22566>.
- [33] T. Dauxois, M. Peyrard, A.R. Bishop, Entropy-driven DNA denaturation, *Phys. Rev. E* 47 (1) (1993) R44–R47.
- [34] M. Zoli, Twist versus nonlinear stacking in short DNA molecules, *J. Theo. Biol.* 354 (2014) 95–104.
- [35] G. Weber, TfReg: Calculating DNA and RNA Melting Temperatures and Opening Profiles with Mesoscopic Models, *Bioinformatics* 29 (2013) 1345–1347, <https://doi.org/10.1093/bioinformatics/btt133> URL <http://bioinformatics.oxfordjournals.org/content/29/10/1345>.

Supplementary information: Sharp DNA denaturation in a helicoidal mesoscopic model

Mateus Rodrigues Leal Gerald Weber
Departamento de Física
Universidade Federal de Minas Gerais
31270-901 Belo Horizonte-MG, Brazil

S1 Numerical integrations

For all numerical calculations we used the Gauss-Legendre quadrature. The integration steps were 1500 for a fixed b , which is the most critical integration due to the occurrence of numerical instabilities and divergences. For the varying integration limit b we used the rule of $200 + 60(b/\text{nm})$ integration points. For ζ and Θ , 10 integration points were found to be sufficient for achieving numerical convergence.

Eq. (8) expanded to second order of $\Delta z_{n,n-1}^2$

$$J_{n,n-1} \approx \Delta z_{n,n-1} \left[1 + \frac{1}{2} \frac{f_{n,n-1}^2}{\Delta z_{n,n-1}^2} - \frac{1}{8} \left(\frac{f_{n,n-1}^2}{\Delta z_{n,n-1}^2} \right)^2 \right] \quad (\text{S1})$$

The equivalent of Eq. (12) is

$$\begin{aligned} Z_{r\theta z} &= \Gamma^N (2\zeta)^N \int_0^b \int_{-\Theta}^{\Theta} \prod_{n=1}^N d\theta_n r_n dr_n \exp \left[-\frac{\beta}{2} (V(r_n) + V(r_{n-1})) \right] \\ &\times \exp \left[-\frac{\beta k}{8J_0^2} \left(f_{n,n-1}^4 - \frac{f_{n,n-1}^6}{2J_0^2} + \frac{5f_{n,n-1}^8}{16J_0^4} \right) \right] \end{aligned} \quad (\text{S2})$$

The equivalent of Eq. (14) is

$$\begin{aligned} Z_{r\theta z}^{app.} &= \Gamma^N (4\zeta\Theta)^N \int_0^b \prod_{n=1}^N dr_n \sqrt{r_n r_{n-1}} \exp \left\{ -\frac{\beta}{2} [V(r_n) + V(r_{n-1})] \right\} \\ &\times \exp \left\{ -\frac{\beta k}{8J_0^2} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^2 \right\} \\ &\times \exp \left\{ \frac{\beta k}{16J_0^4} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^3 \right\} \\ &\times \exp \left\{ -\frac{5\beta k}{128J_0^6} [(r_n - r_{n-1})^2 + \omega^2 r_n r_{n-1}]^4 \right\} \end{aligned} \quad (\text{S3})$$

And the kernel to second order, equivalent o Eq. (15), is

$$\begin{aligned} K(x, y) &= \sqrt{xy} \exp \left\{ -\frac{\beta}{2} [V(x) + V(y)] \right\} \\ &\times \exp \left\{ -\frac{\beta k}{8J_0^2} [(x - y)^2 + \omega^2 xy]^2 \right\} \\ &\times \exp \left\{ \frac{\beta k}{16J_0^4} [(x - y)^2 + \omega^2 xy]^3 \right\} \\ &\times \exp \left\{ -\frac{5\beta k}{128J_0^6} [(x - y)^2 + \omega^2 xy]^4 \right\} \end{aligned} \quad (\text{S4})$$

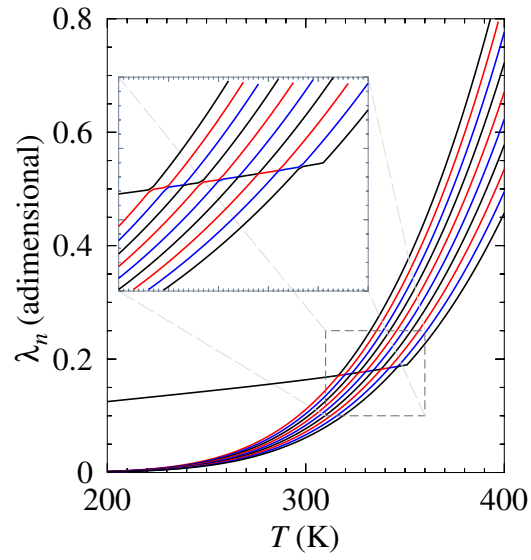


Figure S1: Equivalent to Fig. 2b for T2.

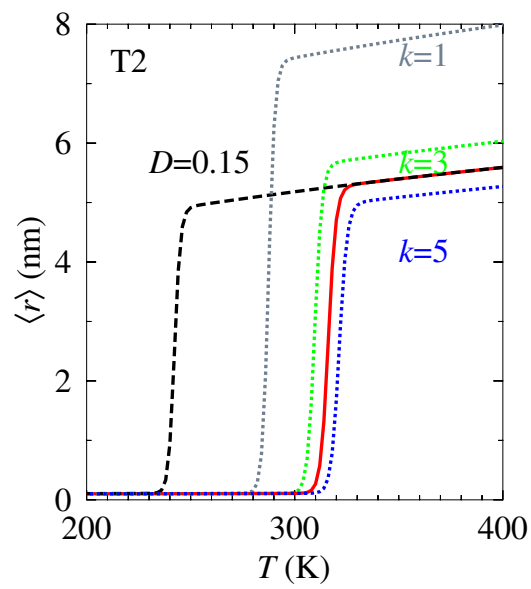


Figure S2: Equivalent to Fig. 3 for T2.

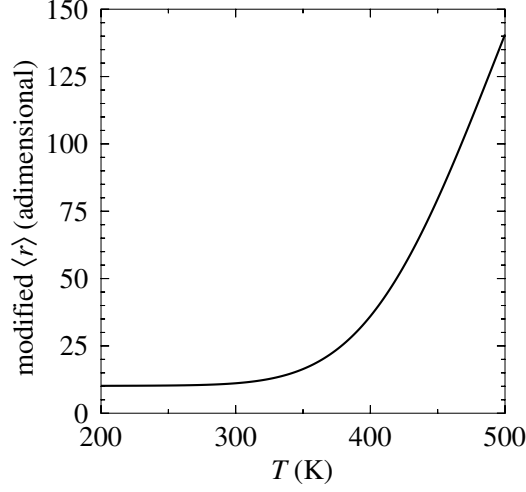


Figure S3: Numerical test artificially removing the factor $(xy)^{1/2}$ from the kernel eq. (15). Note that this is not dimensionally consistent, as removing $(xy)^{1/2}$ renders $\langle r \rangle$ adimensional, and is only shown for the purpose of testing the influence of this factor. Model parameters are the same as for Fig. 2a.

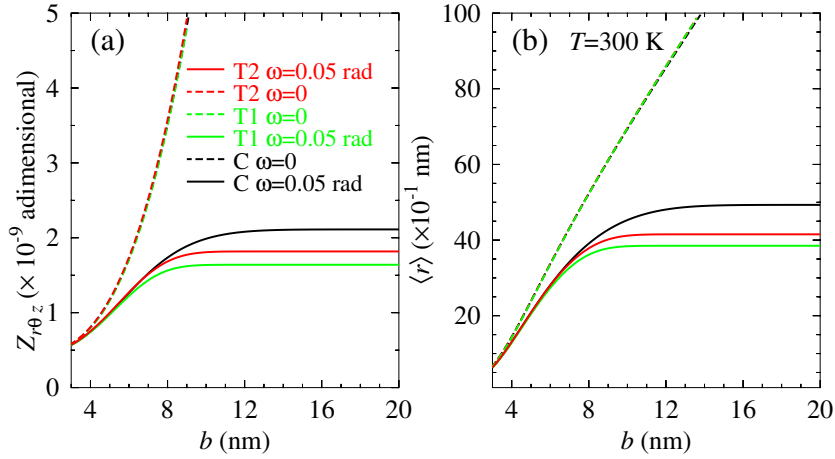


Figure S4: Equivalent to Fig. 4 at 300 K.