

Sérgio Eduardo Ferreira

Modelo Dinâmico para a molécula de DNA



1910001222





UNIVERSIDADE ESTADUAL PAULISTA

Instituto de Biociências, Letras e Ciências Exatas

DEPARTAMENTO DE FÍSICA

Modelo Dinâmico para a molécula de DNA

Sérgio Eduardo Ferreira

Dissertação de Mestrado
Pós-Graduação em Biofísica Molecular

Rua Cristovão Colombo, 2265

Caixa Postal 136

15054-000 - São José do Rio Preto - SP - Brasil

Telefone: (017) 221-2200

Fax: (017) 221-2200

+577.3
F383m

20901003

Ferreira, Sérgio Eduardo.

Modelo dinâmico para a molécula de DNA / Sérgio Eduardo Ferreira
– São José do Rio Preto : [s.n.], 2003
93 f. : il. ; 30 cm.

Orientador: Elso Drigo Filho

Co-Orientador: José Roberto Ruggiero

Dissertação (mestrado) – Universidade Estadual Paulista. Instituto de
Biociências, Letras e Ciências Exatas

1. Ácido desoxirribonucleico - Modelos. 2. Ácido
desoxirribonucleico - Promotor T7A₁. 3. Potencial de Morse. 4.
Métodos Runge-Kutta. I. Drigo Filho, Elso. II. Ruggiero, José Roberto.
III. Universidade Estadual Paulista. Instituto de Biociências, Letras e
Ciências Exatas. IV. Título.

CDU – 531.14



Modelo Dinâmico para a molécula de DNA

Sérgio Eduardo Ferreira

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, São Paulo, para a obtenção do título de Mestre em Biofísica Molecular.

Orientador: Prof. Dr. Elso Drigo Filho

Co-orientador: Prof. Dr. José Roberto Ruggiero



t. 1.222/03

São José do Rio Preto

Setembro 2003



Agradecimentos

A todos os papéis que contribuíram positivamente ao resultado desta
minha obra. Aos Professores Drs. Elso Drigo Filho e José Roberto
Ruggiero, pela orientação e incentivo ao desenvolvimento deste trabalho. Aos
Professores Marcelino Teófilo, Angélica Augustinhe e Carmine L. Rossi, que
contribuíram com discussões e sugestões que de uma forma ou de outra me
foram muito úteis. Aos amigos Gilson Nelson Marques e Fernando Lopes de
Oliveira, por me ajudarem durante alguns meses de discussão que foram
muito úteis.

*"A grandeza do trabalho é talvez, antes de tudo unir os
homens, só há um luxo verdadeiro, o das relações
humanas" Saint-Exupéry*

Dedico

aos meus pais: Timóteo e Maria José;

aos meus irmãos, Paulo e Márcia;

e à minha namorada Adriana.



Agradecimentos

À todas as pessoas que contribuíram positivamente na construção desta senda que trilhei. Aos Professores. Drs. Elso Drigo Filho e José Roberto Ruggiero, pela orientação, e incentivo no desenvolvimento deste trabalho; aos Professores Masayoshi Tsuchida, Augusto Augostinho e Ermínia L. Fanti, que contribuíram com discussões e exposições que de uma forma ou de outra me foram preciosíssimas; aos amigos Gleber Nelson Marques e Eduardo Lima de Oliveira um agradecimento sincero pelas longas horas de discussões que também adicionaram muito no trabalho e na minha formação como pessoa e ainda à minha namorada Adriana de Oliveira Dias, quem sempre soube muito bem compreender-me e apoiar-me. Agradeço também aos meus amigos deste Instituto Fernando Alessandro, Renata, Solange, Ricardo Alexandre, Lessandra, Renato Gaspar, Priscila, Flávio, Ézio, Néelson, Dênis, Fernanda R., Fernanda C., Shirley R., Luís Fernando, Irene, Mário, Joaquim, Plínio, Tatiana, Cássius e Nilton Pimenta, pelas Amizades dedicadas. Enfim a todos os meus amigos.

Aos meus pais e irmãos que sempre me apóiam.

E À Deus por ter me concedido experiências tão frutíferas e amizades tão fiéis.



Resumo

Investigamos a dinâmica de um modelo que simula o DNA (*Modelo vibracional planar: Peyrard e Bishop*). O modelo adotado simplifica a geometria helicoidal da molécula por uma estrutura plana, aproxima as interações que ocorrem na direção do eixo da hélice por um potencial harmônico homogêneo e as interações de pontes de hidrogênio que ocorrem entre os pares de bases, também denominadas interações de emparelhamento, são aproximadas por um potencial de Morse, diferenciado para cada tipo de emparelhamento ($A = T$ ou $C \equiv G$). Usando este modelo, simulamos a seqüência real da região promotora T7A₁ e observamos que, sob algumas condições ocorre a localização das maiores amplitudes de oscilação/energia em regiões ricas em emparelhamentos do tipo $A = T$.



Sumário

Abstract

We investigate the dynamic of a model that simulate the DNA (*plane vibrational model: Peyrard and Bishop*). The Adopted model simplifies helical geometry of the molecule by a plane structure. It also approximates the interactions which occur in the direction of the axis of the helix by an homogeneous harmonic potential. The the interactions of hydrogen bonds that occur between bases pair, are approximated by a Morse's potential, the parameters for this potential differ for each type of pairing ($A = T$ or $C \equiv G$). Using this model, we simulated the real sequence of the promotor T7A₁ and observed that, under some conditions, the largest amplitude of oscillations and energy are localized in regions that are rich in $A = T$ pairings.



Sumário

Introdução	1
1 Aspectos estruturais da molécula de DNA	4
1.1 A composição e estrutura dos ácidos nucleicos	5
1.2 A estrutura do DNA	8
1.2.1 A estrutura primária	8
1.2.2 A estrutura secundária	10
1.3 Principais interações que estabilizam a dupla hélice	12
1.3.1 Interações de ponte de hidrogênio	13
1.3.2 Interações de empilhamento	16
1.3.3 Interações de longo alcance	17
2 Aspectos dinâmicos do DNA: fundamentos físicos e matemáticos da dinâmica	19
2.1 Considerações iniciais	20
2.2 Modelagens para a molécula	23



2.2.1	Modelo de bastão	24
2.2.2	Mobilidade interna da molécula	24
2.2.3	Modelo de Englander	30
2.2.4	Modelos planares	30
2.2.5	Os modelo helicoidais	33
3	Modelo dinâmico para o DNA	35
3.1	O modelo vibracional planar	35
3.1.1	A equação Hamiltoniana	37
4	Localização da energia em uma cadeia de osciladores	41
4.1	Acoplamento Harmônico	42
4.2	Calculando as frequências dos modos normais	47
4.3	Introdução de uma força periódica no sistema	48
4.4	Resultados para a cadeia específica de osciladores	49
4.5	Teste de conservação da energia	57
4.6	Discussão dos resultados	59
5	Localização da energia no modelo vibracional planar	60
5.1	O potencial de emparelhamento das bases	61
5.2	Introdução de uma força periódica no sistema	63
5.3	As equações de movimento	63
5.4	Localização da energia na cadeia do promotor T7A1	64



5.5 Localização da energia em cadeias aleatórias 74

5.6 Teste de conservação da energia 86

5.7 Discussão dos resultados 87

Conclusão **89**

Introdução

Há centenas ou até milhares de tipos de moléculas que podem constituir a maquinaria bioquímica de um organismo vivo, algumas delas, devido ao papel funcional que exercem em processos essenciais ao desenvolvimento e manutenção de tais organismos são consideradas fundamentais. E o DNA é, desse ponto de vista, uma dessas moléculas.

Como uma molécula informacional, é nele que está contida a informação necessária à produção de proteínas. Processos que envolvem a interação do DNA com outras moléculas biológicas determinam qual deve ser a proteína e qual a quantidade a ser produzida. Dada a especificidade dessas estruturas uma falha nesse processo pode muitas vezes causar um desequilíbrio letal. Desse modo, um melhor entendimento do funcionamento desses mecanismos em níveis moleculares pode nos auxiliar, por exemplo, no estudo e desenvolvimento de fármacos. Devido à importância e complexidade da molécula e de seu funcionamento, ela foi e ainda tem sido largamente investigada.

Os estudos que culminaram na determinação de uma estrutura para o DNA trouxeram muita luz para uma melhor compreensão dos mecanismos utilizados pela molécula no desempenho de suas funções. Contudo, uma abordagem dos aspectos dinâmicos pode também contribuir para a investigação de tais meca-



nismos. Dado o grande número de fatores envolvidos nos processos dinâmicos é razoável supor que uma modelagem matemática da dinâmica deve ser igualmente complexa e computacionalmente custosa. Em geral, o problema é simplificado levando-se em conta somente os termos que dão as contribuições dominantes para o fenômeno considerado, esta abordagem é largamente utilizada no estudo de dinâmica do DNA.

Particularmente, o modelo vibracional de *Peyrard e Bishop* [8], foi inicialmente empregado pelos autores no estudo da desnaturação térmica do DNA, ou seja, relaciona-se com as rupturas das pontes de hidrogênio. Este aspecto é bastante interessante, pois do mesmo modo, na modelagem dinâmica que é proposta aqui, desejamos observar movimentos de oscilação de "grandes" amplitudes relacionados com a "quebra" destas pontes.

No caso específico do processo de transcrição do código genético em organismos eucariotos, feitos pela RNA-polimerase II (RNA-polimerase responsável pela transcrição da maioria das proteínas nesses organismos), a bioquímica nos indica que no princípio do processo ocorre a formação de um complexo protéico responsável por auxiliar na criação da bolha de transcrição, que é uma abertura inicial que se dá entre as fitas. Este complexo de inicialização, como é chamado, se forma em uma região localizada dentro do promotor, que é rica em nucleotídeos com bases timina e adenina, denominada TATA-box.

Neste trabalho, utilizando-se de um modelo planar para a molécula de DNA, simulamos a seqüência de 168 pares de nucleotídeos correspondentes a



região promotora T7A₁ e comparamos os resultados com os de uma sequência aleatória.

A ordem de apresentação desta dissertação é a seguinte: no capítulo 1 é feita uma rápida discussão dos elementos que interferem na dinâmica desta molécula tais como a composição, a geometria ou disposição espacial dos elementos constituintes e das principais interações existentes entre estes elementos e deles com outras moléculas; no capítulo 2 falamos de alguns fenômenos dinâmicos embasados por métodos experimentais e apresentamos a título de exemplificação algumas modelagens propostas na literatura; no capítulo 3 tratamos com detalhes o Modelo vibracional planar discutindo as simplificações envolvidas e apresentando suas equações e os parâmetros que são adotados; no capítulo 4 é feito um estudo de determinação da localização de movimentos de maiores amplitudes/energia em uma rede de osciladores unidimensional com massas homogêneas e com uma não-homogeneidade nas constantes das molas do sistema; no capítulo 5 apresentamos como é feita a diferenciação entre os tipos de emparelhamento, e comparamos os resultados obtidos para a sequência do promotor T7A₁ com os resultados obtidos para sequências "aleatórias" do mesmo tamanho.



Capítulo 1

Aspectos estruturais da molécula de DNA

Uma vez que estamos interessados em abordar uma modelagem dinâmica do DNA, se faz necessário uma discussão dos aspectos estruturais e dos fatores que interferem na dinâmica dessa molécula. Neste capítulo, apresentamos uma breve explanação dos elementos que constituem o DNA, de como estão dispostos espacialmente na molécula e das principais interações que envolvem estas estruturas.



Figure 1.1: Chemical structure of a phosphate group, showing the central phosphorus atom (P) bonded to four oxygen atoms (O). One oxygen is double-bonded to the phosphorus, and the other three are single-bonded, with one of them having a negative charge (O-).

1.1 A composição e estrutura dos ácidos nucleicos

Nas células, as macromoléculas são polímeros constituídos pela ligação entre moléculas orgânicas "pequenas", chamadas de monômeros ou subunidades, formando assim longas cadeias poliméricas. O DNA (ácido desoxirribonucleico) é um desses polímeros, que consiste de duas cadeias formadas por nucleotídeos, seus monômeros, constituídos por três estruturas:

- um grupo fosfato;
- uma pentose;
- uma base nitrogenada;

apresentadas a seguir:

Grupo fosfato

O ácido fosfórico, responsável por uma carga residual negativa no esqueleto da molécula é apresentado na figura 1.1.

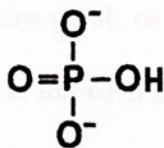


Figura 1.1: Fórmula estrutural do grupo fosfato, carregado negativamente, presente no esqueleto da molécula

Pentose

No DNA, a pentose é uma desoxirribose, ou seja, é uma ribose modificada pela substituição do grupo hidroxila (OH) ligado ao carbono* da posição 2' por um hidrogênio (H), como pode ser visto na figura 1.2

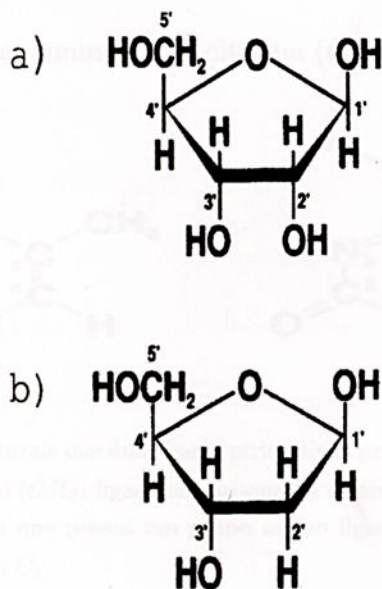


Figura 1.2: a) A fórmula estrutural da β -D-ribose encontrada nas moléculas de RNA.
b) A fórmula estrutural da β -D-2-desoxirribose que é encontrada nas moléculas de DNA.

Observamos que, de maneira geral, os átomos do anel da desoxirribose não são coplanares. Esta é uma parte muito dinâmica e flexível do DNA que pode assumir diferentes conformações, dando assim, origem a mudanças estruturais na molécula. Apresentaremos no capítulo 2 algumas considerações acerca da flexibilidade dessa estrutura.

*A numeração dos átomos de carbono do açúcar é feita usando-se a linha, para diferenciá-los dos átomos da base.

Bases

Entre as diferentes bases nitrogenadas que compõem o DNA há semelhanças que nos permitem agrupá-las em duas famílias de moléculas. Abaixo, nas figuras 1.3 e 1.4, são apresentadas as fórmulas estruturais de cada uma delas:

1) As pirimidinas: timina (T) e citosina (C)

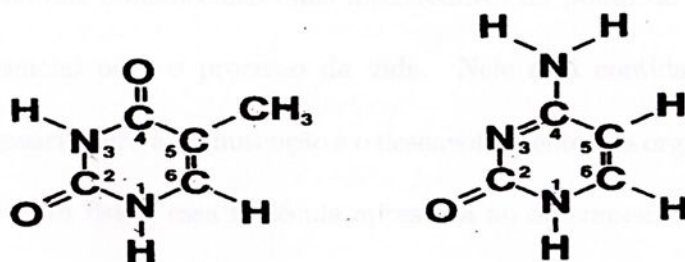


Figura 1.3: Fórmulas estruturais das duas bases pirimidinas presentes no DNA. À esquerda, a timina que apresenta um metil (CH_3) ligado ao carbono C_5 e com grupos carbonila nas posições C_4 e C_2 , à direita, a citosina que possui um grupo amino ligado ao carbono C_4 e um grupo carbonila ($C=O$) na posição C_2 .

2) As purinas: adenina (A) e guanina (G)

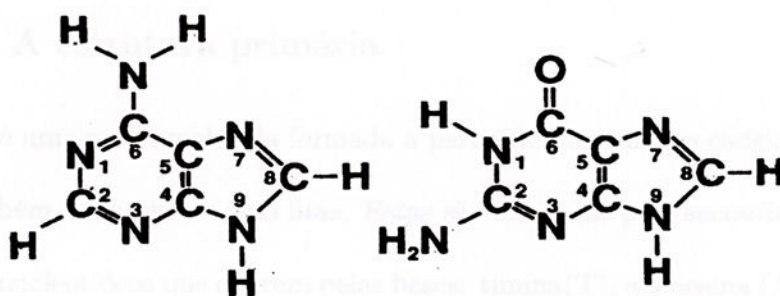


Figura 1.4: Fórmulas estruturais das duas bases purinas presentes no DNA. À esquerda, a adenina que possui um grupo amino (NH_2) ligado ao carbono C_6 e à direita, a guanina que possui um grupo amino ligado ao carbono C_2 e um grupo carbonila ($C=O$) na posição C_6 .

As três estruturas básicas apresentadas anteriormente ao se unirem, formam nucleotídeos, os quais, por sua vez, ligados quimicamente uns aos outros formam o polinucleotídeo.

1.2 A estrutura do DNA

O DNA é uma das biomoléculas mais interessantes do ponto de vista biológico porque é essencial para o processo da vida. Nele está contida a informação genética necessária para a manutenção e o desenvolvimento dos organismos vivos. Do ponto de vista físico, essa molécula apresenta no desempenho da sua função algumas características que são bastante intrigantes: ela deve ter uma grande "resistência" a fim de preservar a integridade da informação genética. Por outro lado, deve apresentar uma considerável flexibilidade e dinâmica para permitir a transferência e replicação dessa informação. Na próxima subseção apresentamos sucintamente algumas características estruturais do ácido desoxirribonucléico.

1.2.1 A estrutura primária

O DNA é uma macromolécula formada a partir de duas longas cadeias poliméricas, também conhecidas como fitas. Estas são formadas pela sucessão de quatro tipos de nucleotídeos que diferem pelas bases: timina(**T**), adenosina (**A**), guanina (**G**) e citosina (**C**). A ordem, ou a seqüência, dos diferentes monômeros nas fitas constituem a estrutura primária da molécula.



Os grupos fosfatos e as desoxirriboses dos nucleotídeos encadeados se alternam na parte externa da molécula, como pode ser visto na figura 1.5, formando uma estrutura de esqueleto, enquanto as bases nitrogenadas ficam voltadas para a parte interior dessa estrutura. A ligação glicosídica que conecta a base nitrogenada ao açúcar, a qual liga o nitrogênio da posição 1 da base ao carbono $C_{1'}$ da pentose, é dita como estando na conformação β ("up") com relação ao açúcar, em contraste com a ligação α ("down") do hidrogênio que se liga ao mesmo carbono.

Como já dissemos, os polinucleotídeos são formados a partir da junção de subunidades (os nucleotídeos). A ligação que junta dois nucleotídeos a partir do carbono $C_{3'}$ do açúcar de um nucleotídeo ao carbono $C_{5'}$ do açúcar de outro nucleotídeo é a ligação fosfodiéster formada durante a síntese bioquímica do DNA.

Outro ponto importante sobre a estrutura do polinucleotídeo é a existência de duas extremidades distintas chamadas de 5' e 3'. Química e biologicamente essas extremidades são bem diferentes, o que atribui a cada fita uma polaridade. Comumente o DNA existe com um grupo hidroxila na extremidade 3' e um grupo fosfato na extremidade 5'.



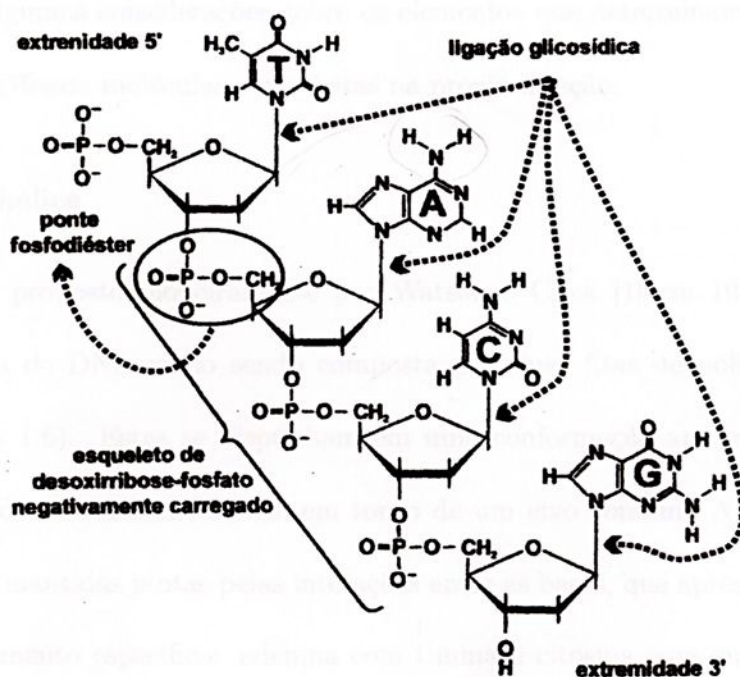


Figura 1.5: Trecho de uma cadeia polinucleotídica linear, com as ligações glicosídica e fosfodiéster indicadas pelas setas pontilhadas.

1.2.2 A estrutura secundária

A sequência dos nucleotídeos nas fitas constitui a estrutura primária da molécula, entretanto, as ligações dos três grupos que compõem cada nucleotídeo, bem como a ligação fosfodiéster (responsável pelo encadeamento dos monômeros), são ligações covalentes suscetíveis de apresentarem rotações, dando assim possibilidade de um vasto número de conformações à molécula. Contudo, a constância dos diagramas de difração de raios-x indicam que a conformação é muito estável e que, assim como acontece com as proteínas, o número de configurações possíveis é bastante restrito.

Algumas considerações sobre os elementos que determinam a estrutura secundária dessas moléculas serão feitas na próxima seção.

A dupla hélice

O modelo proposto pioneiramente por Watson e Crick [1] em 1953 descrevia a molécula de DNA como sendo composta por duas fitas de polinucleotídeos (ver figura 1.6). Estas se dispunham em uma conformação antiparalela par a par, estando torcidas em espiral em torno de um eixo comum. Além disso, as fitas eram mantidas juntas pelas interações entre as bases, que apresentavam um emparelhamento específico: adenina com timina e citosina com guanina. Para o desenvolvimento desse modelo duas informações foram essenciais: a regra de *Chargaff* e os parâmetros de difração de fibra de DNA. A regra de *Chargaff* indicava, já no início dos anos cinquenta, que a quantidade de adenina (A) era igual à quantidade de timina (T) e que a quantidade de citosina (C) era igual à de guanina (G). Observava-se que isso era verdade para DNA purificado de uma variedade de organismos e independente da quantidade total de A+T (ou C+G). Esta condição sugeria que nas duas fitas do DNA havia um emparelhamento exclusivo de A com T e C com G.

Entretanto, mesmo com o auxílio dessas informações, a elucidação dessa estrutura não foi tão direta, pois Watson e Crick não tinham meios de saber que o DNA era necessariamente composto de duas fitas.



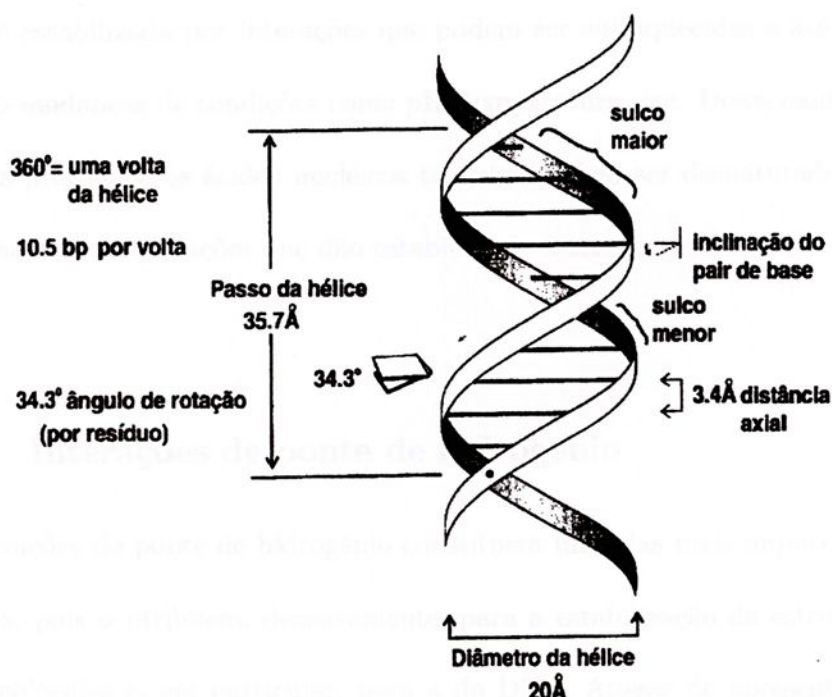


Figura 1.6: A dupla hélice de Watson e Crick, em detalhes os parâmetros estruturais da hélice

1.3 Principais interações que estabilizam a dupla hélice

Na investigação da estrutura do DNA foram levantados vários elementos que levaram a considerar a existência de uma estrutura mais complexa do que a de uma fita simples e completamente distendida. A simples análise da viscosidade e densidade indica que o DNA nativo possui uma estrutura altamente compacta. Entretanto, algumas mudanças nas propriedades físico-químicas podem causar alterações significativas em propriedades, tais como, absorbância de UV, apontando para uma troca de conformação. Tais indicações mostraram que a estrutura

nativa é estabilizada por interações que podem ser enfraquecidas e até destruídas, sob mudanças de condições como pH, temperatura, etc. Desse modo, assim como as proteínas, os ácidos nucleicos também podem ser desnaturados. Uma discussão dessas interações que dão estabilidade à estrutura nativa é feita nesta seção.

1.3.1 Interações de ponte de hidrogênio

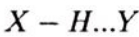
As interações de ponte de hidrogênio constituem uma das mais importantes interações, pois contribuem, decisivamente, para a estabilização da estrutura das macromoléculas e, em particular, para a do DNA. Apesar de apresentar baixa energia, e ser uma interação relativamente fraca [5] tabela 1.1, o grande número dessas pontes que se formam entre as bases do DNA resultam numa contribuição energética fundamental na determinação e manutenção da estrutura de dupla hélice descrita anteriormente.



Tabela 1.1: Comparação de alguns valores de energia envolvidos nas ligações covalentes e pontes de hidrogênio

Ligação	Comprimento (Å)	Energia da ligação (Kcal/mol)	Energia requerida p/ estirar em 0.1 Å (Kcal/mol)
<i>Covalentes</i>			
C-C	1.54 ± 0.02	83.1	3.25
C-H (no etano)	1.09 ± 0.02	98.8	3.6
<i>Ponte de hidrogênio</i>			
O-H...O	2.7 ± 0.2	3 a 6	0.1

A ponte de hidrogênio ocorre comumente entre um átomo "doador", X, ligado covalentemente ao hidrogênio e um átomo "receptor", Y que tem um grau de elétronegatividade. Este estado pode ser esquematizado como se segue:



Como a interação tem um caráter predominantemente eletrostático, a intensidade de uma ponte de hidrogênio depende das cargas dos átomos envolvidos e também das distâncias entres esses átomos. Contudo, a interação apresenta um caráter direcional, ou seja, há também uma dependência dos ângulos entre as ligações dos átomos envolvidos, sendo que a intensidade da ponte de hidrogênio é máxima quando estes átomos estão alinhados. No DNA, as interações de ponte



de hidrogênio formadas entre os pares de base são de dois tipos:



A eletronegatividade do "receptor" provém dos elétrons do oxigênio da carbonila (C=O) ou de um nitrogênio (N).

No emparelhamento do tipo Watson-Crick, o par guanina-citosina apresenta três pontes de hidrogênio e o par adenina-timina duas, como mostrado na figura 1.7.

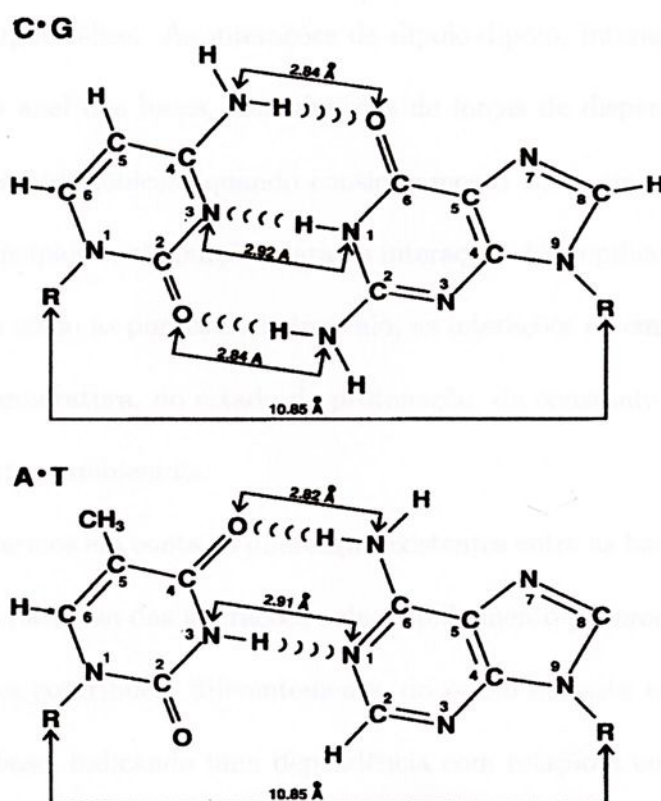


Figura 1.7: Esquema de emparelhamento das bases do tipo Watson e Crick.

Comparando as energias envolvidas nas interações (tabela 1.1) observamos que a ponte de hidrogênio apresenta uma energia de vinte a trinta vezes menor que a energia de uma ligação covalente (C-C) ou (C-H). Conseqüentemente, há uma grande diferença na rigidez, ou seja, a ponte de hidrogênio apresenta uma maior facilidade de ser dobrada ou esticada.

1.3.2 Interações de empilhamento

Além das pontes de hidrogênio há outras interações a serem consideradas, por exemplo, as interações de empilhamento, que também contribuem para a estabilidade da dupla hélice. As interações de dipolo-dipolo, interações do elétron π oscilante do anel das bases, contribuições de forças de dispersão de London e de interações hidrofóbicas (quando consideramos o DNA em solução aquosa) formam as principais contribuições para as interações de empilhamento.

Assim como as pontes de hidrogênio, as interações de empilhamento dependem da temperatura, do estado de protonação, da constante dielétrica e de outros parâmetros ambientais.

Se levarmos em conta as diferenças existentes entre as bases A, C, T e G (seção 1.1) e a natureza das interações e de empilhamento podemos perceber que essas interações contribuem diferentemente, do ponto de vista energético, para cada tipo de base, indicando uma dependência com relação à composição. Por exemplo, um trecho da hélice mais rico em emparelhamentos A-T deve apresentar energia de estabilização diferente de um trecho mais rico em emparelhamentos



C-G. Contudo, as interações de empilhamento apresentam não somente uma dependência da composição, mas também, uma dependência da ordem ou sequência, conforme indicado pelos resultados da química quântica [5] (tabela 1.2).

Tabela 1.2: Energias de empilhamento em kcal/mol de dímeros para os possíveis empilhamentos em uma B-DNA.

Dímeros empilhados	Energias de empilhamento [kcal/mol de dímeros]
$\uparrow \frac{C-G}{G-C} \downarrow$	-14.59
$\uparrow \frac{C-G}{A-T} \downarrow$ ou $\uparrow \frac{T-A}{G-C} \downarrow$	-10.51
$\uparrow \frac{C-G}{T-A} \downarrow$ ou $\uparrow \frac{A-T}{G-C} \downarrow$	-9.81
$\uparrow \frac{G-C}{C-G} \downarrow$	-9.69
$\uparrow \frac{G-C}{G-C} \downarrow$ ou $\uparrow \frac{C-G}{C-G} \downarrow$	-8.26
$\uparrow \frac{T-A}{A-T} \downarrow$	-6.57
$\uparrow \frac{G-C}{T-A} \downarrow$ ou $\uparrow \frac{A-T}{C-G} \downarrow$	-6.57
$\uparrow \frac{G-C}{T-A} \downarrow$ ou $\uparrow \frac{T-A}{C-G} \downarrow$	-6.78
$\uparrow \frac{A-T}{A-T} \downarrow$ ou $\uparrow \frac{T-A}{T-A} \downarrow$	-5.37
$\uparrow \frac{A-T}{T-A} \downarrow$	-3.82

1.3.3 Interações de longo alcance

As interações de longo alcance, de natureza columbiana, são devidas, principalmente, aos grupos fosfatos carregados negativamente, presentes na molécula de



DNA (figura 1.5). A menor distância entre fosfatos de fita diferentes é cerca de 20 Å. Consequentemente, suas interações podem ser consideradas fracas. Em contraste, a menor distância entre fosfatos na mesma fita pode ser de apenas 5 Å. Isto significa que a repulsão mútua pode ser bem forte. Esta repulsão deve ser minimizada para que se mantenha a estrutura do DNA assim, a forma nativa da molécula deve estar em um meio apropriado.

Os fosfatos são, então, blindados por "contra-íons" presentes no meio; esta blindagem é muito estável com a concentração de "contra-íons" mantendo-se constante mesmo com uma variação de 0.5mM até 0.5M. Neste intervalo de concentração, o número médio de $0.88Na^+$ (Sódio) para cada grupo fosfato do esqueleto é mantido. Entretanto, variações de parâmetros ambientais tais como, temperatura, pH, valência do "contra-íon", entre outros, presentes em condições fisiológicas podem alterar a blindagem.



Capítulo 2

Aspectos dinâmicos do DNA: fundamentos físicos e matemáticos da dinâmica

Após uma breve introdução acerca da estrutura e de interações do DNA, feita anteriormente, passaremos a abordar aspectos da modelagem dessa biomolécula. Neste capítulo apresentamos alguns fenômenos que ocorrem no DNA, e que indicam uma grande mobilidade das suas estruturas, justificando assim uma modelagem dinâmica. Discutimos também alguns graus de liberdade e movimentos internos que estão presentes no DNA e apresentamos sucintamente, alguns modelos para esta molécula propostos na literatura.



esqueleto da molécula seja transposto, isto é, ocorra a exposição das bases para o reconhecimento como é ilustrado na figura 2.1.

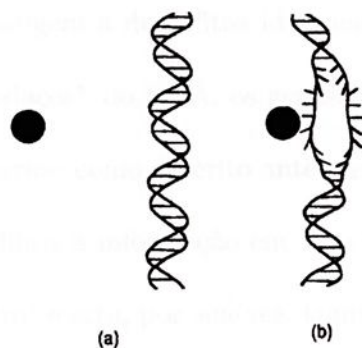


Figura 2.1: (a) uma proteína e a dupla hélice, (b) a dupla hélice desnatada localmente para o reconhecimento das bases pela proteína.

Um exemplo deste mecanismo é a interação da proteína RNA-polimerase II com a região promotora. Essa interação ocorre no princípio do processo de transcrição, quando começa a montagem do complexo de iniciação da transcrição numa região próxima ao sítio de inicialização. O complexo constitui-se de várias proteínas, conhecidas como fatores gerais de transcrição, que se ligam num processo de montagem. Este complexo é responsável por auxiliar a RNA-polimerase II a iniciar a transcrição. A montagem inicia-se com a ligação do fator geral de transcrição **TFIID** [22, 23] a uma sequência rica em nucleotídeos com timina (T) e adenina (A), conhecida como TATAbox. Com a ligação do **TFIID** ocorre uma distorção local no DNA e outros fatores são montados juntamente com a RNA-polimerase II. Ocorre, então, a formação de uma bolha conhecida como bolha de transcrição, ou seja, o rompimento das pontes de hidrogênio numa pequena região, de 15-20 pares de base [24].

Grosso modo, podemos associar a função do DNA, basicamente, a dois processos dinâmicos principais: a *replicação*, no qual a molécula se divide longitudinalmente para dar origem a duas fitas idênticas, e a *transcrição*, no qual são gerados cópias de "pedaços" do DNA, os genes. Esses genes são transcritos do DNA pela RNA polimerase como descrito anteriormente. Essa enzima que lê a fita molde do DNA codifica a informação em uma fita complementar que ela sintetiza, (RNA mensageiro) e esta, por sua vez, transfere a informação ao RNA ribossômico para a construção das cadeias polipeptídicas (as proteínas).

Outros movimentos são observados, tais como as aberturas flutuacionais ou "suspiros" nos quais alguns pares de bases têm seus emparelhamentos temporariamente rompidos e ficam expostos por um tempo de aproximadamente $10^{-7}s$. Essas aberturas flutuacionais são consideradas precursoras da desnaturação [15].

Todos esses fenômenos dinâmicos evidenciam a grande flexibilidade e mobilidade dessa molécula e fundamentam uma modelagem dinâmica. Como vimos diversos são os fenômenos dinâmicos associados ao DNA e assim também há uma variedade de modelos propostos para a molécula. A seguir a fim de exemplificar apresentaremos brevemente algumas destas modelagens que tem sido propostas.



2.2 Modelagens para a molécula

As análises de fenômenos tais como aberturas flutuacionais ou da transição podem trazer idéias ou informações relevantes para um melhor entendimento do DNA. Entretanto, esses processos *in vivo* são extremamente complexos. Além disso, são processos que ocorrem num estado de não equilíbrio, o que torna uma abordagem termodinâmica bem complicada. Dada a grande complexidade estrutural do sistema e o grande número de fatores envolvidos, é razoável supor que uma modelagem matemática "completa", que leva em conta esse grande número de fatores e a interdependência entre eles, seja igualmente complexa e computacionalmente custosa. O problema, contudo, pode ser simplificado se construirmos um modelo que leve em conta somente os termos responsáveis pelas principais contribuições para um processo a ser considerado; esta abordagem é largamente utilizada no estudo de dinâmica do DNA.

A seleção dos termos dominantes pode ser feita de muitas maneiras e isto explica a diversidade de modelos propostos. De forma geral, o DNA tem sido considerado como um sistema no qual efeitos não lineares são essenciais, e nessas abordagens a estrutura do DNA é proposta como sendo perfeitamente regular [3, 7, 8, 11], isto é, o espaçamento entre as bases são iguais e as suas massas são homogêneas, embora alguns trabalhos tenham incluído não homogeneidades para representar o código genético [20, 21]. Os modelos são concebidos para terem soluções que podem imitar algum aspecto da dinâmica da molécula conhecido



experimentalmente. Em particular, são investigados efeitos referentes a trocas de conformação local que podem se propagar ao longo da molécula, como na transcrição; movimentos de grandes amplitudes das bases que possam indicar aberturas flutuacionais; a desnaturação térmica da molécula, entre outros.

2.2.1 Modelo de bastão

O modelo mais simples para a molécula do DNA conhecido como o modelo de bastão, que aproxima a molécula por um bastão elástico uniforme. Este modelo leva em conta três tipos de movimentos da molécula: estiramento, torção, dobramento ou encurvamento da estrutura [2]. Notamos que em um modelo como este o detalhamento das estruturas da molécula fica bastante empobrecido, deste modo, percebemos que na medida em que assumimos simplificações na construção de um modelo limitamos nossas possibilidades de detalhamentos. Assim se desejamos observar detalhes de estruturas internas da molécula necessitamos considerar na modelagem aspectos referentes à mobilidade destas estruturas.

2.2.2 Mobilidade interna da molécula

Descrevemos no capítulo anterior a estrutura da molécula de um ponto de vista praticamente estático. Entretanto, como já mencionamos, a estrutura do DNA é estabilizada pelas interações entre os átomos que a compõem, estas interações possuem um caráter muito dinâmico, isto é, elas podem apresentar variações em função da alteração de parâmetros físico-químicos. Por exemplo, o aumento do pH



pode diminuir a blindagem de grupos carregados e, conseqüentemente, aumentar a intensidade de atração ou repulsão entre átomos (ou grupos) podendo, desta maneira, aumentar ou diminuir a rigidez de determinadas regiões da molécula e propiciar uma maior flexibilidade ou mobilidade a elas.

O fato do DNA estar imerso em um banho térmico, em contato com outras moléculas (água, sais etc.) e assim, sujeito a choques e interações com drogas, também pode ser destacado como uma possível fonte de mobilidade da molécula. Numa modelagem, mediante argumentos o problema complexo pode ser simplificado, a seguir algumas considerações sobre a mobilidade de estruturas tais como as bases e o esqueleto da molécula.

Graus de liberdade das bases

As bases nitrogenadas são constituídas de anéis cíclicos, os quais podem ser considerados estruturas planas, uma vez que seus átomos estão deslocados de um plano por quantidades muito pequenas (menores que $0.1\text{\AA}$). Estes anéis são relativamente rígidos devido a interação do elétron π oscilante que confere uma estabilidade adicional e, dessa forma, a mobilidade interna dessas estruturas pode eventualmente ser desconsiderada. Isto permite adotar a simplificação de representar as bases por placas planas e discutir, então, somente as possibilidades de movimentos de uma base com relação à outra, o que é denominado graus de liberdade intra-bases, e os movimentos de um par de base com relação a outro, ou graus de liberdade inter-bases.



Para descrever, de um modo claro, os graus de liberdade das bases, fixamos um sistema de coordenadas *Oxyz* ortogonais. Admitindo ainda que as bases estão ortogonalmente posicionadas com relação ao eixo da hélice e que este é coincidente com o eixo *z*, o eixo *x* intercepta as pontes de hidrogênio e o eixo *y* passa pelos anéis das bases. As quatro figuras subsequentes (2.2, 2.3, 2.4 e 2.5) ilustram as possibilidades de movimento das bases com estas simplificações.

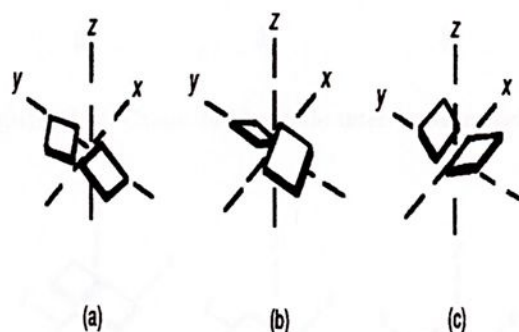


Figura 2.2: Graus de liberdade intra-bases rotacional.

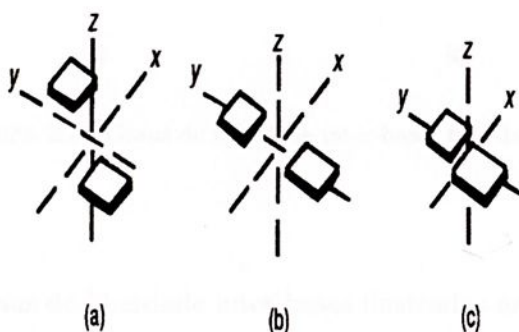


Figura 2.3: Graus de liberdade intra-bases translacional.

Notemos, entretanto, que as bases não estão totalmente livres para girar ou transladar, há impedimentos que impõem resistência a esses movimentos. No caso dos movimentos ilustrados nas figuras 2.2 e 2.3, a resistência oferecida deve-

se, principalmente, às pontes de hidrogênio que se formam entre os pares de bases.

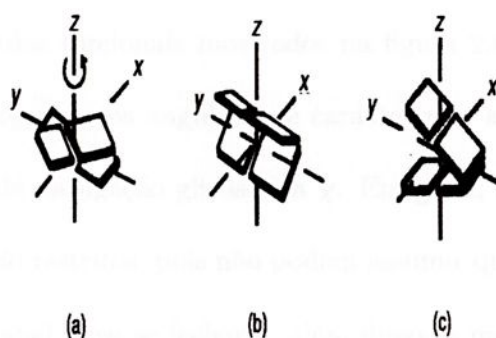


Figura 2.4: Graus de liberdade inter-bases rotacional.

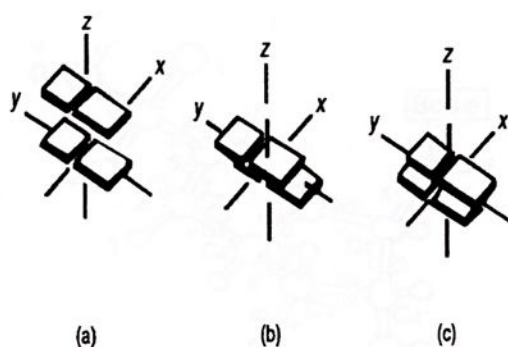
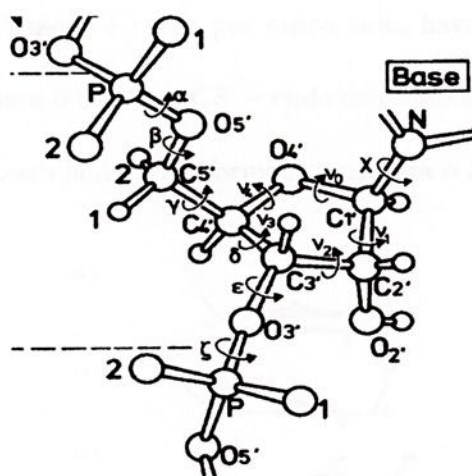


Figura 2.5: Graus de liberdade inter-bases translacional.

Para os graus de liberdade inter-bases ilustrados nas figuras 2.4 e 2.5 as restrições ou impedimentos são impostos pela resistência devido às interações ao longo do eixo da hélice.

Graus de liberdade do complexo desoxirribose-fosfato

O complexo desoxirribose-fosfato constitui o chamado esqueleto da molécula e apresenta onze ângulos torcionais mostrados na figura 2.6. Dentre eles, destacamos: v_0 , v_1 , v_2 , v_3 e v_4 , os ângulos que caracterizam a conformação do anel do açúcar, e o ângulo da ligação glicosídica χ . Em geral, os valores dos ângulos torcionais do anel são restritos, pois não podem assumir qualquer valor, uma vez que a estrutura do anel deve se fechar e, além disso, a maioria dos ângulos são interdependentes.



anel não são coplanares. Duas classes de conformações são apresentadas: primeiro quando um dos átomos ($C2'$ ou $C3'$) é deslocado do plano dos outros quatro átomos, formando uma estrutura que lembra um envelope aberto, chamada de forma **E**; ou alternativamente, quando dois átomos adjacentes ($C2'$ e $C3'$) são deslocados em direções opostas com relação ao plano obtém-se a denominada forma **T**. Se o átomo é deslocado na mesma direção do carbono $C5'$ é chamado de *endo*, se o átomo é deslocado na direção oposta então, é chamado de *exo*. Na forma **T** os dois átomos que são deslocados não estão necessariamente em posições simétricas. Se isso acontecer e o átomo $C3'$ for *endo*, por exemplo, a notação dada para descrever esta situação é : ${}^3T$; por outro lado, havendo um deslocamento assimétrico e maior para o carbono $C3'$ – *endo* denotamos por ${}^3T_2$. A figura 2.7 apresenta algumas possibilidades conformacionais para o anel da desoxirribose.

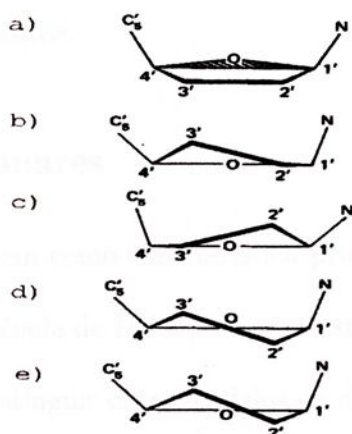


Figura 2.7: Ilustração de algumas possibilidades conformacionais para o anel da desoxirribose (a) os cinco átomos estão no mesmo plano, situação não observada; (b) estrutura **E** com $C3'$ – *endo*; (c) estrutura **E** com $C2'$ – *endo*; (d) estrutura **T** simétrica com $C3'$ – *endo* e (e) estrutura **T** com um maior desvio para o $C3'$ – *endo*.

2.2.3 Modelo de Englander

Um modelo bastante simples que procurava evidenciar aspectos dinâmicos do DNA foi proposto por Englander et al. [3], e propunha a possibilidade de excitações sólitons se propagando ao longo da molécula, o que leva à proposta de um modelo com interações não lineares. O modelo de Englander é composto por uma cadeia de pêndulos como a mostrada na figura 2.8

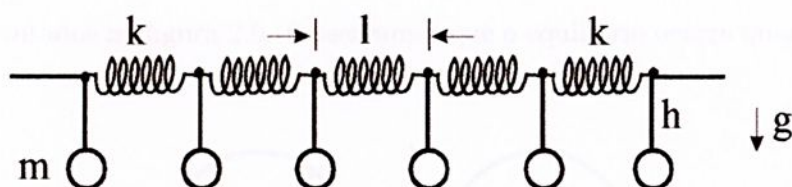


Figura 2.8: Modelo dinâmico para o DNA proposto por Englander e colaboradores.

Posteriormente outros modelos um pouco mais complexos que o modelo de Englander foram propostos.

2.2.4 Modelos planares

Esta classe de modelos tem como característica principal a simplificação da estrutura helicoidal da molécula de DNA para uma estrutura planar. Dentro desta classe, podemos ainda distinguir entre modelos de dinâmica torcional e modelos de dinâmica vibracional.

O **modelo de dinâmica torcional** proposto por Yakushevich [7] é formado por duas cadeias de discos giratórios que são acoplados longitudinalmente

por molas, as quais, simulam as interações ao longo do eixo da hélice e há também um acoplamento transversal entre as cadeias, o qual, simula as interações de emparelhamento.

O acoplamento longitudinal depende dos ângulos de torção α_i , α_{i-1} , β_i e β_{i-1} enquanto que o acoplamento transversal depende do estiramento x_i provocado pela torção dos discos, entretanto, este estiramento depende de uma forma não linear dos ângulos α_i e β_i . Detalhes do emparelhamento do modelo torcional são apresentados na figura 2.9. Observamos que o equilíbrio ocorre quando $\alpha = 0$ e $\beta = 0$.

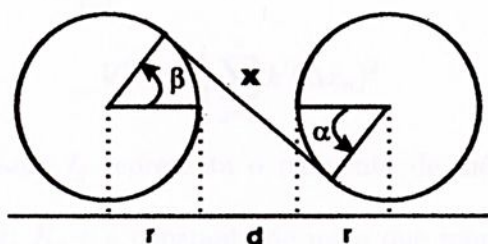


Figura 2.9: Modelo de dinâmica torcional, os discos representam os pares de bases e a distância x calculada em função dos ângulos β e α representa o tamanho das pontes de hidrogênio.

O quadrado do tamanho das pontes de hidrogênio, x^2 , é determinado, pela própria geometria do sistema, e é dado por:

$$x^2 = [d + 2r - r(\cos\beta + \cos\alpha)]^2 + [r(\sin\beta - \sin\alpha)]^2.$$

Portanto o estiramento da ponte de hidrogênio Δx , é obtido pela diferença entre o comprimento desta ligação estressada, x , e o comprimento dela relaxada, $x_0 = d$ (na figura 2.9), assim obtemos:

$$\Delta x = x - x_0 = [(2r + x_0 - r\cos\alpha - r\cos\beta)^2 + (r\sin\alpha - r\sin\beta)^2]^{\frac{1}{2}} - x_0,$$

dessa forma a função Hamiltoniana do sistema é:

$$H = T + V^{(1)} + V^{(2)} , \quad (2.1)$$

onde T representa a energia cinética, $V^{(1)}$ é a energia potencial torcional relativa ao empilhamento e $V^{(2)}$ é a energia potencial relativa às pontes de hidrogênio.

Em termos matemáticos obtém-se:

$$T = \frac{1}{2} \sum_{j=1}^n I_j \dot{\alpha}_j^2 + I_j \dot{\beta}_j^2 , \quad (2.2)$$

$$V^{(1)} = \frac{1}{2} \sum_{j=1}^n K_j [(\alpha_j - \alpha_{j-1})^2 + (\beta_j - \beta_{j-1})^2] , \quad (2.3)$$

$$V^{(2)} = \frac{1}{2} \sum_{j=1}^n k' (\Delta x_n)^2 . \quad (2.4)$$

Nesta expressão I_j representa o momento de inércia de cada um dos discos do j -ésimo par; K_j é a constante de mola que representa a interação de empilhamento e k' é a constante de mola associada às pontes de hidrogênio.

O **modelo de dinâmica vibracional** proposto por *Peyrard e Bishop* [8] também é composto por duas cadeias de osciladores, onde a dinâmica é analisada em termos do movimento radial das bases com relação ao eixo da dupla hélice, as coordenadas u_n e v_n descrevem o deslocamento das n -ésimas bases com relação às suas posições de equilíbrio. Em tal modelo, o emparelhamento entre as bases é feito pelo potencial de Morse, e o acoplamento longitudinal (aquele que simula as interações ao longo do eixo da hélice) é feito por um potencial harmônico. A



Hamiltoniana do modelo é:

$$H = \frac{1}{2} \sum_{j=1}^n \left(\frac{p_{u,j}^2}{m_j} + \frac{p_{v,j}^2}{m_j} \right) + \frac{1}{2} \sum_{j=1}^{n+1} K_j [(u_j - u_{j-1})^2 + (v_j - v_{j-1})^2] + \frac{1}{2} \sum_{j=1}^n V_j (u_j - v_j)^2, \quad (2.5)$$

onde a primeira somatória indica a energia cinética do sistema, a segunda representa a energia potencial das interações ao longo do eixo da hélice e a última corresponde à energia potencial das pontes de hidrogênio simuladas pelo potencial de Morse.

$$V(u_j - v_j) = \xi [\exp(-a(u_j - v_j)) - 1]^2, \quad (2.6)$$

onde ξ corresponde à profundidade do poço do potencial, ou energia de dissociação, e a está associado à largura do poço.

Este modelo é tratado em detalhes no capítulo 3 e sua dinâmica estudada mais adiante.

2.2.5 Os modelo helicoidais

Em contraposição aos modelos planares que introduzem simplificações geométricas no modelo da estrutura da macromolécula, o modelo helicoidal propõe a possibilidade de se levar em conta a geometria helicoidal da dupla hélice. Na estrutura helicoidal (que é mais realística) grupos carregados estão espacialmente próximos de modo que eles podem interagir, causando a formação de pontes de filamentos de água entre os grupos considerados. Esta modificação foi primeiro proposta para o modelo de Yakushevich [11] e depois também aplicada ao modelo



de Peyrard-Bishop [12, 13]. A Hamiltoniana do modelo helicoidal com dinâmica torcional é dada por:

$$H = \frac{1}{2} \left\{ \sum_{j=1}^n I \dot{\alpha}_j^2 + I \dot{\beta}_j^2 + \sum_{j=1}^n K_L [(\alpha_j - \alpha_{j-1})^2 + (\beta_j - \beta_{j-1})^2] + \sum_{j=1}^n K_T (\Delta x_n)^2 + \sum_{j=1}^n K_H d^2 (\alpha_{j+h} - \beta_j)^2 + \sum_{j=1}^n K_H d^2 (\alpha_{j-h} - \beta_j)^2 \right\},$$

e a Hamiltoniana do modelo helicoidal de dinâmica vibracional é:

$$H = \frac{1}{2} \left\{ \sum_{j=1}^n \left(\frac{p_{u,j}^2}{m} + \frac{p_{v,j}^2}{m} \right) + \sum_{j=1}^{n+1} k [(u_j - u_{j-1})^2 + (v_j - v_{j-1})^2] + \sum_{j=1}^n V (u_j - v_j)^2 + \sum_{j=1}^n K [(u_j - v_{j+h})^2 + (u_j - v_{j-h})^2] \right\}.$$

A diferença existente entre as Hamiltonianas dos modelos helicoidais com relação às Hamiltonianas dos equivalentes modelos planares está na última somatória, relacionado com a energia da interação entre aqueles grupos carregados que ficam próximos quando é considerada a geometria helicoidal da molécula.

Capítulo 3

Modelo dinâmico para o DNA

No capítulo 2, foram apresentados diferentes tipos de modelos para a molécula de DNA. Na abordagem proposta neste trabalho adotaremos o modelo vibracional planar proposto por Peyrard e Bishop. A seguir detalharemos as principais características e simplificações deste modelo.

3.1 O modelo vibracional planar

Este modelo considera somente os movimentos vibracionais das bases em torno de suas posições de equilíbrio na direção radial ao eixo da dupla hélice. Ou seja, assume-se que os movimentos ao longo do eixo da hélice são muito reduzidos devido à rigidez da estrutura nessa direção. No caso do modelo vibracional planar não se leva em conta a geometria helicoidal do DNA, mas adota-se uma estrutura plana como a mostrada na figura 3.1. Os movimentos serão reduzidos ao caso

unidimensional, como verificaremos.

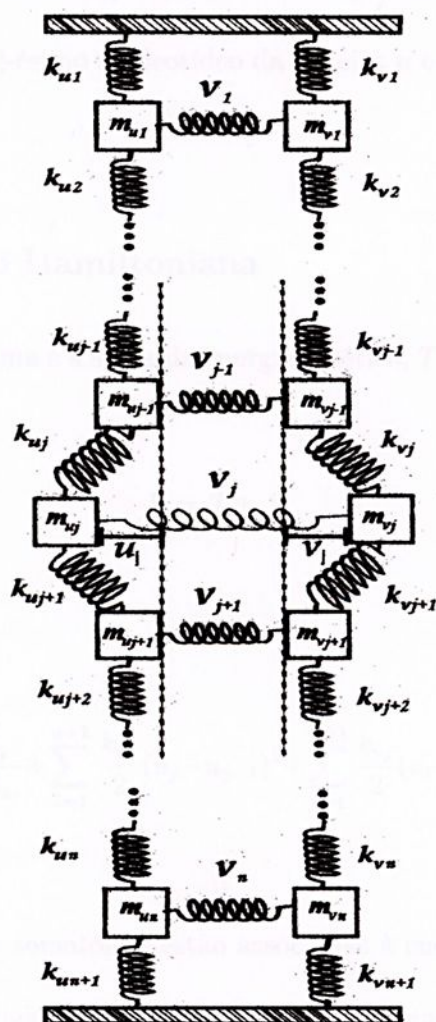


Figura 3.1: Modelo mecânico plano que consiste de duas cadeias de osciladores massa-mola acopladas e com extremidades.

No modelo mecânico, ilustrado na figura 3.1, V_j identifica o potencial que acopla as duas cadeias e simula as interações de emparelhamento (ou pontes de hidrogênio). V_j depende do deslocamento relativo $(u_j - v_j)$ e enquanto k_{u_j} e k_{v_j}

são, respectivamente, as constantes de força das molas, que simulam a resistência imposta pelas interações ao longo do eixo da hélice, ao movimento nas direções de $\mathbf{u}_j$ e $\mathbf{v}_j$. As massas do j -ésimo nucleotídeo das cadeias u e v são, respectivamente, m_{u_j} e m_{v_j} .

3.1.1 A equação Hamiltoniana

A energia total do sistema é a soma da energia cinética, T , e da energia potencial, V ,

$$H = T + V,$$

ou seja,

$$H = \sum_{j=1}^n \frac{p_{u_j}^2}{2m_{u_j}} + \sum_{j=1}^n \frac{p_{v_j}^2}{2m_{v_j}} + \sum_{j=1}^{n+1} \frac{k_{u_j}}{2} (u_j - u_{j-1})^2 + \sum_{j=1}^{n+1} \frac{k_{v_j}}{2} (v_j - v_{j-1})^2 + \sum_{j=1}^n V_j(u_j - v_j), \quad (3.1)$$

onde as duas primeiras somatórias estão associadas à energia cinética, as outras duas seguintes associadas à energia potencial das interações ao longo das fitas, que são simuladas por um potencial harmônico, e a última somatória representa a energia potencial das interações de pontes de hidrogênio, simuladas por um potencial que depende somente da diferença $u_j - v_j$. Observamos ainda que nessa expressão $p_{u_j} = m_{u_j} \dot{u}_j$ e $p_{v_j} = m_{v_j} \dot{v}_j$ são os momentos da j -ésima massa da cadeia $\mathbf{u}$ e $\mathbf{v}$, respectivamente.

Notemos que as equações de movimento obtidas a partir da hamiltoniana

(3.1) estarão acopladas, isto é, as equações que descrevem os movimentos nas coordenadas u_j dependerão da coordenada v_j e vice-versa. Entretanto, introduzindo algumas simplificações e com uma substituição de variáveis adequada podemos obtê-las de forma que elas estejam desacopladas.

Lembramos que na molécula de DNA os nucleotídeos se emparelham de forma específica. Então, com base nos valores das massas de cada nucleotídeo, exibidos na tabela 3.1 (em unidades de massa atômica), podemos observar que a diferença entre os tipos de pares é de aproximadamente 0.16%, embora as massas individuais de cada base apresentem uma diferença maior.

Tabela 3.1: Comparação das massas dos nucleotídeos em unidades de massa atômica.

Bases	Composição química da base	Composição química do complexo desoxirribose-fosfato	Massa total do nucleotídeo em (u.m.a.)
<i>Adenina</i>	$C_5N_5H_2$	$C_5O_5PH_8$	311
<i>Timina</i>	$C_5N_2O_2H_4$	$C_5O_5PH_8$	303
<i>Citosina</i>	$C_4N_3OH_2$	$C_5O_5PH_8$	287
<i>Guanina</i>	$C_5N_5OH_3$	$C_5O_5PH_8$	328

Uma simplificação introduzida em [8, 13] considera todas as massas homogêneas* de forma que:

$$m_{u_j} = m_{v_j} = m = 300 \text{ u.m.a.} \approx 5,0 \times 10^{-25} \text{ Kg} .$$

*Fator de conversão : 1 u.m.a.— 1.6606×10^{-27} Kg



Também não será levado em conta as diferenças entre as interações ao longo das fitas [13, 12] e, conseqüentemente, as constantes das molas que representam essas interações são consideradas todas iguais:

$$k_{u_j} = k_{v_j} = k = 1.5 \text{ eV}/\text{\AA}^2 \approx 24 \text{ N/m} ,$$

assim, a equação 3.1 pode ser reescrita da seguinte forma:

$$H = \sum_{j=1}^n \frac{p_{u_j}^2}{2m} + \sum_{j=1}^n \frac{p_{v_j}^2}{2m} + \sum_{j=1}^{n+1} \frac{k}{2} (u_j - u_{j-1})^2 + \sum_{j=1}^{n+1} \frac{k}{2} (v_j - v_{j-1})^2 + \sum_{j=1}^n V_j (u_j - v_j) . \quad (3.2)$$

Para desacoplar as equações introduz-se as novas variáveis:

$$x_j = \frac{u_j + v_j}{\sqrt{2}} \quad e \quad y_j = \frac{u_j - v_j}{\sqrt{2}} ,$$

e substituindo na equação (3.2), obtém-se:

$$H = H_x + H_y ,$$

onde,

$$H_x = \sum_{j=1}^n \frac{p_{x_j}^2}{2m} + \sum_{j=1}^{n+1} \frac{k}{2} (x_j - x_{j-1})^2 , \quad (3.3)$$

$$H_y = + \sum_{j=1}^n \frac{p_{y_j}^2}{2m} + \sum_{j=1}^{n+1} \frac{k}{2} (y_j - y_{j-1})^2 + \sum_{j=1}^n V_j (\sqrt{2} y_j) . \quad (3.4)$$

Como estamos interessados na dinâmica de abertura das pontes de hidrogênio, e as equações agora não dependem uma da outra, focalizaremos nossa

atenção na equação (3.4). Um sistema equivalente para esta Hamiltoniana é representado na figura 3.2

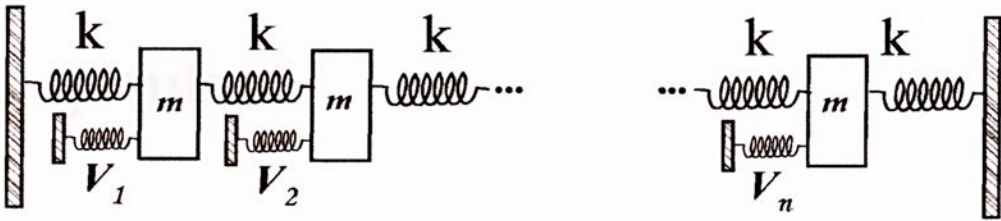


Figura 3.2: Sistema mecânico que representa o modelo simplificado unidimensional para a dinâmica do DNA, dado pela Hamiltoniana (3.4). Consiste de uma cadeia de osciladores massa-mola com extremidades fixas. Notemos que as massas estão acopladas entre si por um potencial harmônico e acopladas a um ponto fixo por um potencial V .

Observamos, então, que nas aproximações usadas a diferença entre os tipos de nucleotídeos que compõe o DNA deve-se apenas às pontes de hidrogênio, ou seja, ela está apenas no potencial $V_j(\sqrt{2}y_j)$.

Capítulo 4

Localização da energia em uma cadeia de osciladores

Vimos que o modelo vibracional planar se reduz a um modelo unidimensional, isto deve-se à simplificação que considera os deslocamentos ortogonais ao eixo da hélice muito maiores que aqueles que ocorrem na direção do eixo. Notemos que no modelo unidimensional simplificado do DNA há dois tipos de acoplamentos: o que conecta os osciladores entre si, harmônico, e aquele que conecta cada oscilador a um suporte fixo, V . No presente capítulo investigamos a localização das maiores amplitudes de vibração/energia, em uma pequena cadeia de osciladores acoplados por um potencial harmônico e com massas homogêneas.



4.1 Acoplamento Harmônico

Em um sistema como uma cadeia de osciladores acoplados, livre de amortecimento, a energia cinética dos osciladores é constantemente transformada em energia potencial e vice-versa. Dessa forma, se a cadeia é homogênea, isto é, tem massas todas iguais e as constantes das molas são também todas iguais, então é fácil perceber que se um oscilador oscila com uma amplitude maior em relação aos outros em algum tempo, então a energia associada a ele será igualmente maior nesse mesmo tempo. Assim para uma cadeia homogênea é direta a relação entre oscilações de grandes amplitudes e localização de grandes energias. Entretanto, se a cadeia não é homogênea esta relação pode não ser direta. Vejamos o caso mais simples com osciladores isolados.

Consideremos um oscilador A de massa m_A e constante da mola K_A , que oscila com a mesma amplitude de um outro oscilador B de massa m_B e constante da mola K_B . Comparando a energia dos dois, temos os seguintes casos:

- $m_A = m_B$ e $K_A > K_B$: a energia associada ao oscilador A será necessariamente maior que a energia associada ao oscilador B .
- $m_A > m_B$ e $K_A = K_B$: a energia associada ao oscilador A será novamente maior que a energia associada ao oscilador B .
- $m_A \neq m_B$ e $K_A \neq K_B$: nesse caso a comparação pode não ser imediata, e então, uma solução é calcular a energia associada a cada um para compará-las.



Deve ser observado que em uma cadeia de osciladores unidimensional, como a cadeia mostrada na figura 4.1, cada oscilador está acoplado a duas molas e não mais a uma única, como o oscilador isolado. Além disso, cada mola também está acoplada a duas massas (exceto as molas das extremidades).

Neste capítulo investigamos a localização da energia em uma cadeia unidimensional de osciladores massa-mola, livre de forças dissipativas. A cadeia contendo massas iguais acopladas por molas de constantes diferentes (não homogênea) e com condições de contorno de extremidades fixas pode ser vista na figura 4.1

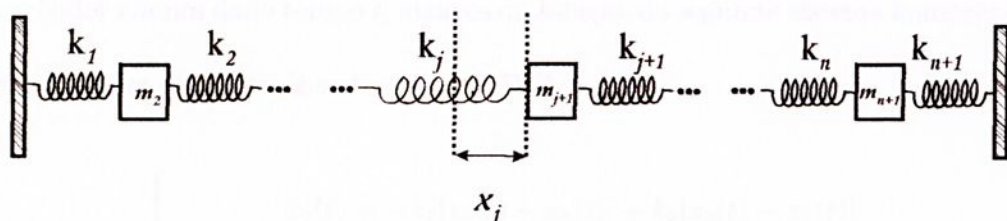


Figura 4.1: Cadeia de osciladores massa-mola com extremidades fixas e massas homogêneas, onde k_j é a constante da mola que conecta a massa do j -ésimo sítio à massa do sítio posterior. Exceto a primeira e a última mola que conectam as correspondentes massas às extremidades fixas. m é a massa de cada oscilador, x_j representa a quantidade pela qual a j -ésima mola está distendida, $x_j > 0$, ou, comprimida, $x_j < 0$.

Sendo que x_1 e x_{n+2} referem-se às extremidades fixas, então $x_1(t) = x_{n+2}(t) = 0$ para qualquer t . A equação Hamiltoniana de um sistema desse tipo, com n massas e $n + 1$ molas é dada por:

$$H = \sum_{j=2}^{n+1} \frac{p_j^2}{2m_j} + \sum_{j=2}^{n+2} \frac{k_{j-1}}{2} (x_j - x_{j-1})^2, \quad (4.1)$$

onde $p_j = m_j \dot{x}_j$ é o momento da j -ésima massa.

Assim, a partir da equação (4.1) obtemos as equações de movimento:

$$m_j \ddot{x}_j(t) = -k_{j-1}[x_j(t) - x_{j-1}(t)] + k_j[x_{j+1}(t) - x_j(t)] , \quad (4.2)$$

$$j = \{2, 3, \dots, n+1\}.$$

Como no sistema considerado a distribuição de massa é homogênea, isto é, todas as massas são iguais, assim, por simplicidade consideraremos:

$$m_2 = m_3 = \dots = m_{n+1} = m = 1.$$

Usando as equações de movimento (4.2), determina-se a posição de cada oscilador em um dado tempo t , através da solução do seguinte sistema homogêneo de equações diferenciais ordinárias (E.D.O.):

$$\left\{ \begin{array}{l} \ddot{x}_2(t) = -k_1[x_2(t) - x_1(t)] + k_2[x_3(t) - x_2(t)] \\ \ddot{x}_3(t) = -k_2[x_3(t) - x_2(t)] + k_3[x_4(t) - x_3(t)] \\ \ddot{x}_4(t) = -k_3[x_4(t) - x_3(t)] + k_4[x_5(t) - x_4(t)] \\ \vdots \\ \ddot{x}_{n+1}(t) = -k_n[x_{n+1}(t) - x_n(t)] + k_{n+1}[x_{n+2}(t) - x_{n+1}(t)] \end{array} \right. \quad (4.3)$$

Reescrevendo o sistema (4.3) na forma matricial temos:

$$\begin{bmatrix} \ddot{x}_2 \\ \ddot{x}_3 \\ \vdots \\ \ddot{x}_j \\ \vdots \\ \ddot{x}_{n+1} \end{bmatrix} = (-1) \begin{bmatrix} (k_1+k_2) & -k_2 & 0 & 0 & \dots & \dots & 0 \\ -k_2 & (k_2+k_3) & -k_3 & 0 & 0 & \dots & 0 \\ 0 & -k_3 & (k_3+k_4) & -k_4 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & \ddots & & & \\ \vdots & & & & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & 0 & 0 & -k_{n+1} & (k_{n+1}+k_n) \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_j \\ \vdots \\ x_{n+1} \end{bmatrix},$$

ou, em notação matricial:

$$\ddot{\mathbf{x}} = -\mathbf{K}\mathbf{x} \quad (4.4)$$

onde $\mathbf{K}$ é a matriz $n \times n$, tridiagonal das constantes de mola.

Sabemos que tais sistemas podem executar oscilações harmônicas, assim, todas as partículas oscilam numa mesma frequência. Entretanto, isso só ocorre para as frequências dos modos normais de vibração [18]. Por conseguinte procuraremos as soluções da forma:

$$x_j(t) = C_j e^{-i\omega_N t}, \quad (4.5)$$

onde C_j é uma constante complexa e ω_N é a frequência de um modo normal de vibração e i o número imaginário $\sqrt{-1}$.

A solução (4.5) pode ainda ser expressa na forma polar:

$$x_j(t) = A_j \cos(\omega_N t) - iB_j \sin(\omega_N t) , \quad (4.6)$$

sendo A_j e B_j constantes reais. Entretanto, a solução física do sistema é identificada somente com a parte real dessa função, logo:

$$x_j(t) = A_j \cos(\omega_N t) . \quad (4.7)$$

Substituindo a equação (4.7) no sistema de equações diferenciais (4.3), obtemos a seguinte forma matricial:

$$-\lambda \mathbf{x} = -\mathbf{K} \mathbf{x} , \quad (4.8)$$

onde $\lambda = \omega_N^2$.

Assim, para determinar as frequências dos modos normais de vibração basta calcularmos os auto-valores associados à matriz $\mathbf{K}$.

A condição necessária e suficiente para que o sistema (4.8) não tenha apenas a solução trivial é que o determinante da matriz, $\mathbf{K} - \lambda \mathbf{I}$ ($\mathbf{I}$ = a matriz identidade), se anule. Ou seja:

$$\det(\mathbf{K} - \lambda \mathbf{I}) = 0 \quad (4.9)$$

Então, diagonalizando a matriz $\mathbf{K}$ encontramos os auto-valores, que são os quadrados das frequências dos modos normais.



4.2 Calculando as frequências dos modos normais

Consideremos o exemplo específico de uma cadeia contendo 10 massas e onze molas, cujas constantes de mola são dadas na tabela, como sugerido na referência 4.1 [17] .

Tabela 4.1: Constantes de molas da cadeia de osciladores.

$k_1 = 0.704$	$k_3 = 0.707$	$k_5 = 0.754$	$k_7 = 0.006$	$k_9 = 0.470$	$k_{11} = 0.904$
$k_2 = 0.388$	$k_4 = 0.525$	$k_6 = 0.721$	$k_8 = 0.479$	$k_{10} = 0.574$	

A matriz tridiagonal das constantes de mola do sistema descrito pelos valores dados na tabela 4.1 é:

$K = \begin{bmatrix} 1.092 & -0.338 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.388 & 1.095 & -0.707 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.707 & 1.232 & -0.525 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.525 & 1.279 & -0.754 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.754 & 1.475 & -0.721 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.721 & 0.727 & -0.006 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.006 & 0.485 & -0.479 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.479 & 0.949 & -0.470 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.470 & 1.044 & -0.574 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.574 & 1.478 \end{bmatrix}$

Como podemos notar não foi mencionado em que unidades estão dadas as constates das molas ou as massas, isto porque não há uma dependência com

relação ao sistema de unidades adotado. Assim se adotamos por exemplo o S.I. a constante de mola é dada em newtons por metro (N/m), a massa em quilogramas (Kg) e o tempo em segundo (s).

Por meio do software *Mathematica* diagonalizamos a matriz $\mathbf{K}$ e obtemos as seguintes frequências:

$$\begin{array}{ll} \omega_{N1} = 1.5664 & \omega_{N6} = 0.9466 \\ \omega_{N2} = 1.4034 & \omega_{N7} = 0.7669 \\ \omega_{N3} = 1.3955 & \omega_{N8} = 0.5659 \\ \omega_{N4} = 1.1483 & \omega_{N9} = 0.2825 \\ \omega_{N5} = 1.1174 & \omega_{N10} = 0.1844 \end{array}$$

4.3 Introdução de uma força periódica no sistema

Para verificarmos a localização da energia nesse sistema, introduziremos para cada um dos osciladores da cadeia uma força periódica, F , que varia senoidalmente com o tempo. Assim, a mesma quantidade de energia é adicionada simultaneamente a cada oscilador. Podemos escrever a força periódica como:

$$F(t) = \mathbb{F}_0 \text{sen}(\omega_f t + \theta_0) \quad (4.10)$$

onde: $\mathbb{F}_0$ é a amplitude da força, ω_f é a frequência da força periódica e θ_0 é um ângulo de fase.

Com a introdução de $F(t)$ as equações de movimento (4.2) tornam-se:

$$\ddot{x}_j(t) = -k_{j-1}[x_j(t) - x_{j-1}(t)] + k_j[x_{j+1}(t) - x_j(t)] + \mathbb{F}_0 \text{sen}(\omega_f t + \theta_0) \quad , \quad (4.11)$$

$$j = \{2, 3, \dots, 11\}.$$

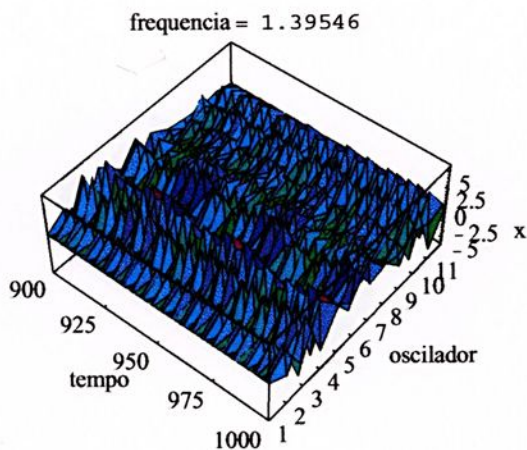
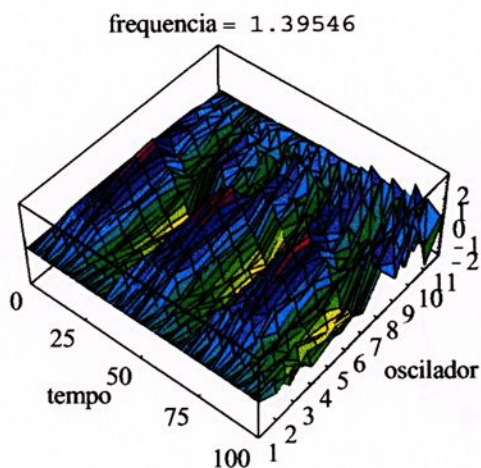
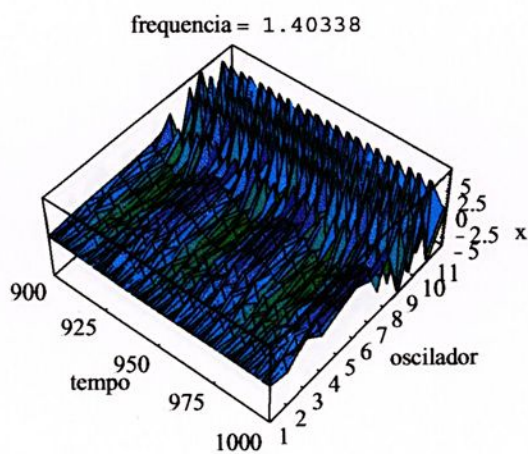
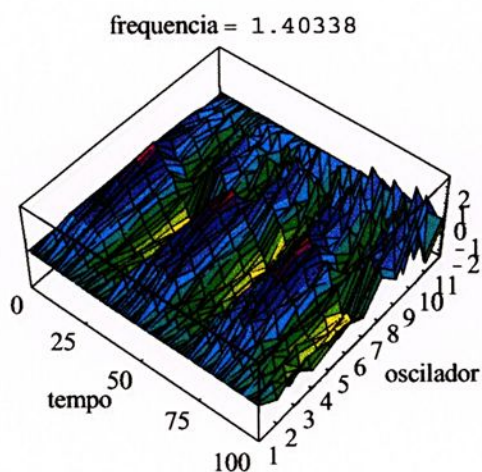
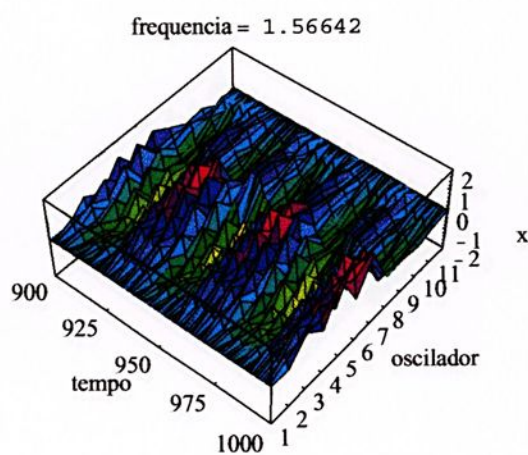
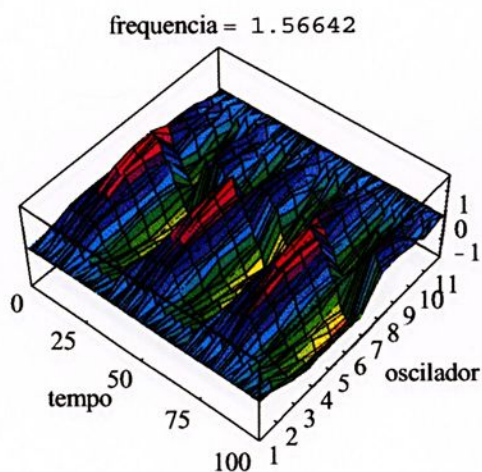
4.4 Resultados para a cadeia específica de osciladores

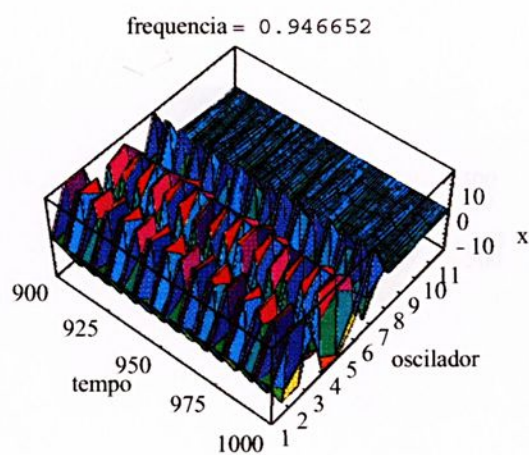
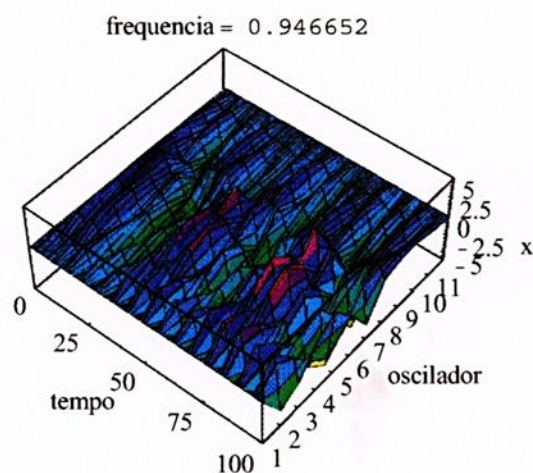
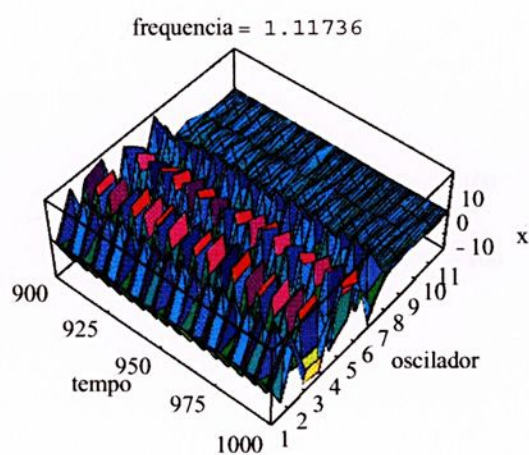
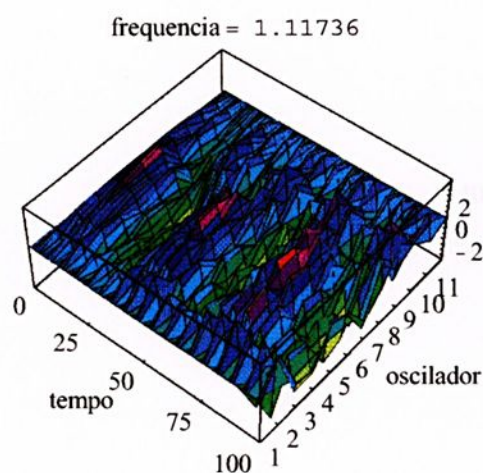
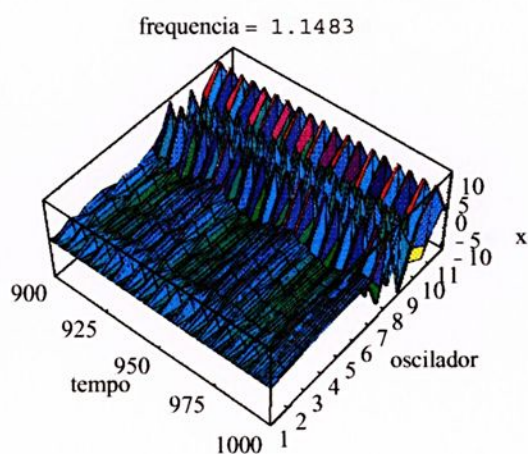
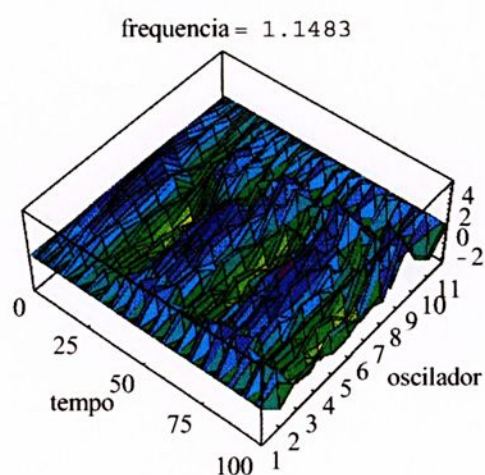
Resolvemos o sistema de equações (4.11), numericamente, utilizando o software *Mathematica* [25] e como integrador foi utilizado o algoritmo de Runge-Kutta com um passo de 10^{-3} .

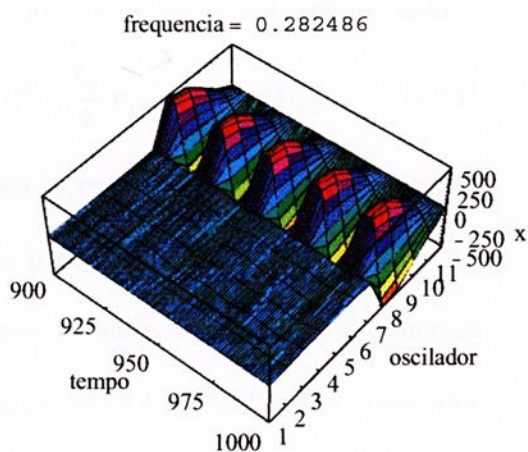
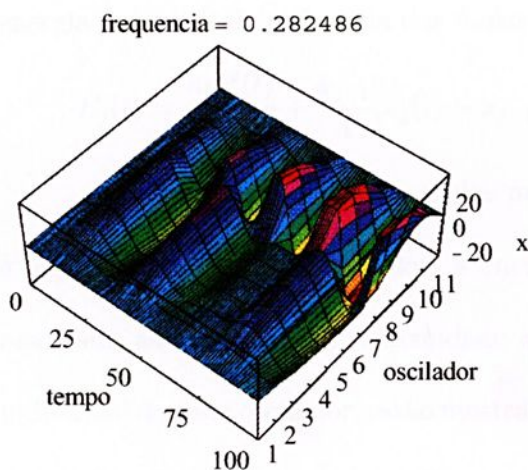
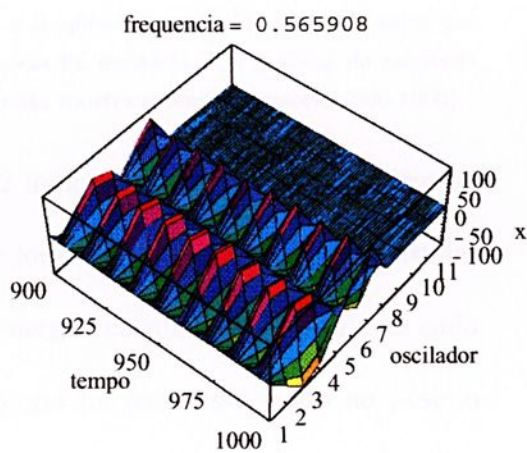
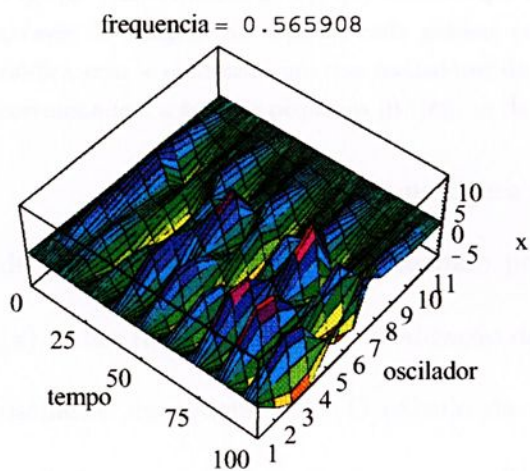
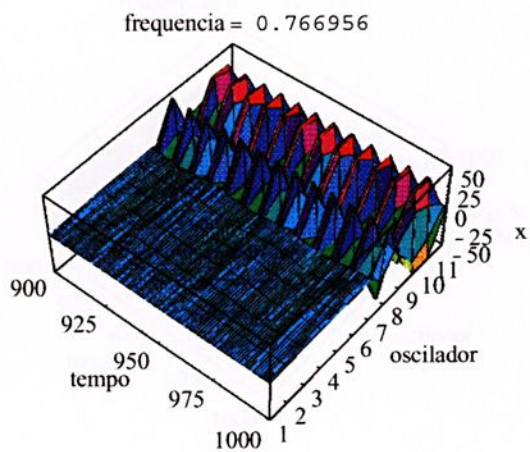
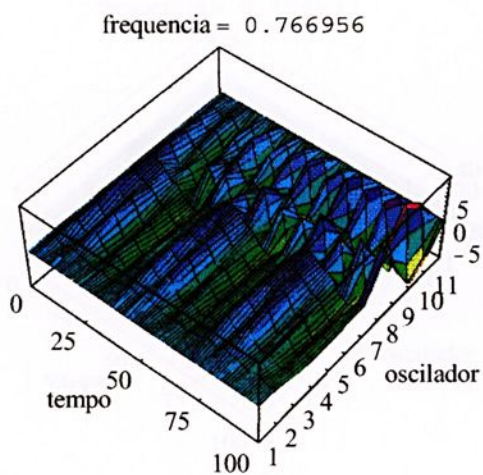
Fixando $\theta_0 = 0$, $F_0 = 0.3$, testamos os efeitos das diferentes frequências da força perturbativa, experimentando as frequências dos modos normais, ou seja, $\omega_f = \{\omega_{N_1}, \omega_{N_2}, \omega_{N_3}, \dots \dots \omega_{N_{10}}\}$.

A seguir, na figura 4.2, os gráficos tridimensionais da coordenada x_j em função do tempo, t , e da localização de cada oscilador na cadeia, j , indicam diferentes padrões de localização das maiores amplitudes para as diferentes frequências da força perturbativa experimentada.

Note que tanto na figura 4.2 quanto na figura 4.3 há dois gráficos para cada frequência testada (um ao lado do outro). No gráfico à esquerda está mostrada a evolução do sistema nos instantes iniciais $t=0$ até $t=100$ (ou $t=50$ no caso dos gráficos da energia). No gráfico à direita está mostrada a dinâmica do sistema no intervalo de tempo que vai de $t=900$ até $t=1000$. Esta divisão em dois momentos da dinâmica tem por objetivo somente possibilitar uma melhor visualização dos gráficos.







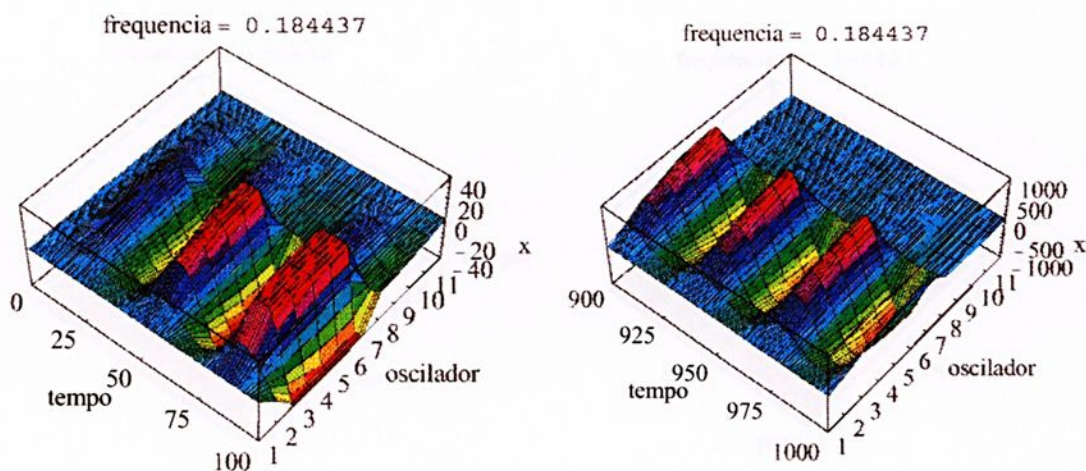
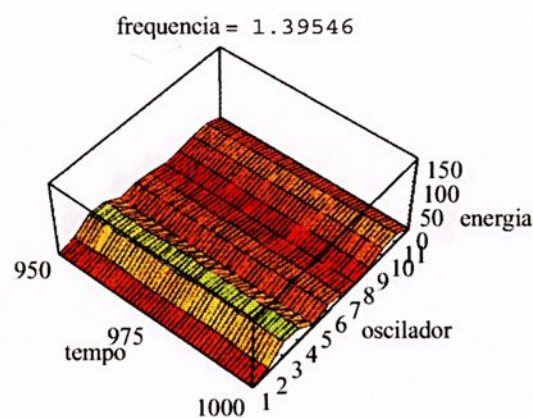
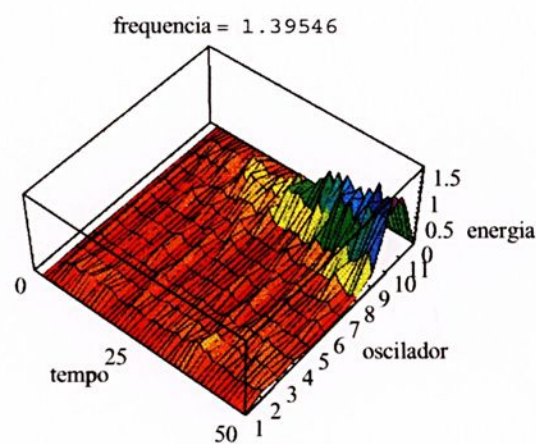
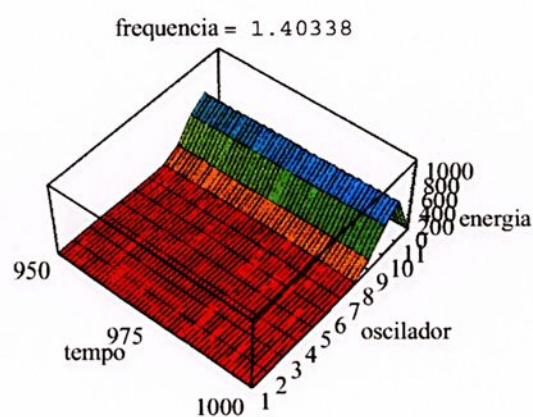
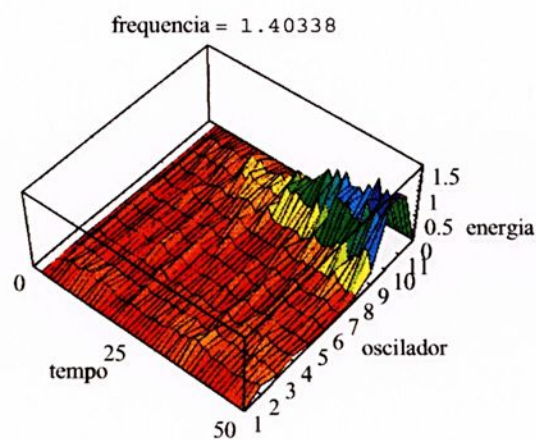
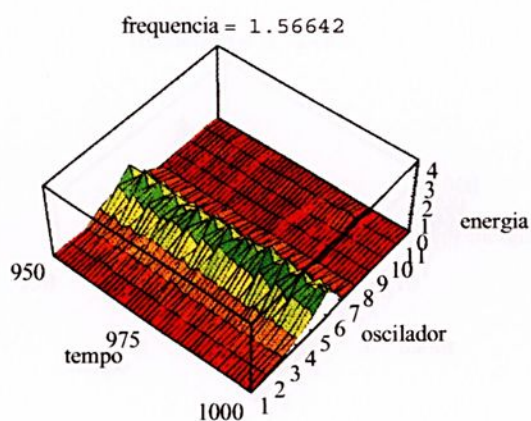
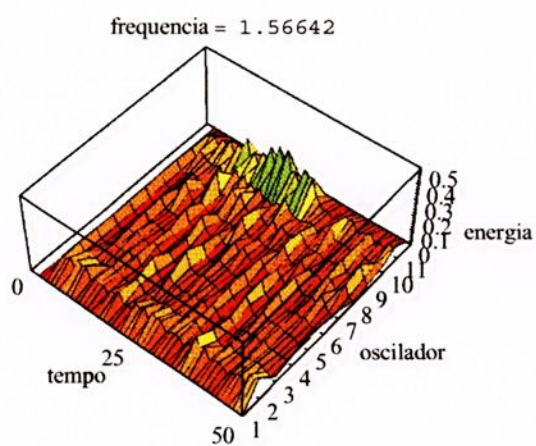


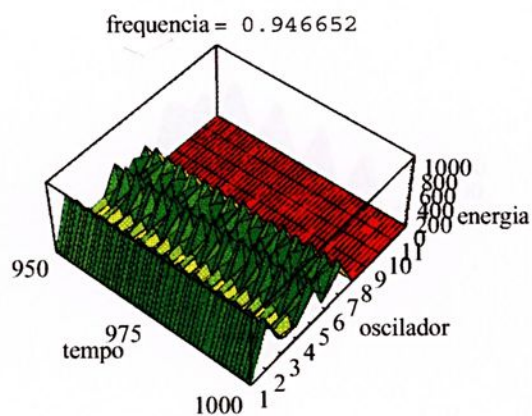
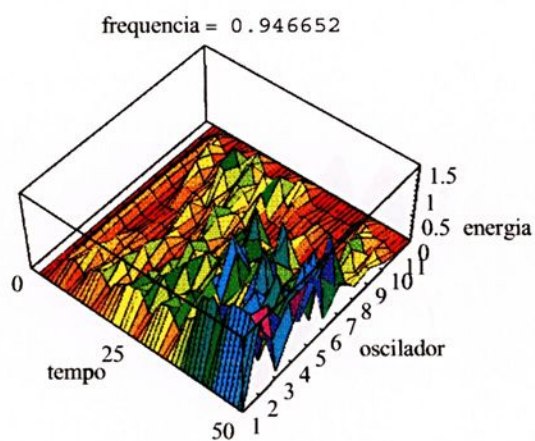
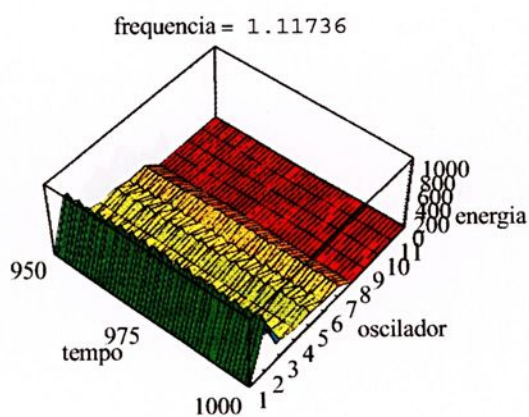
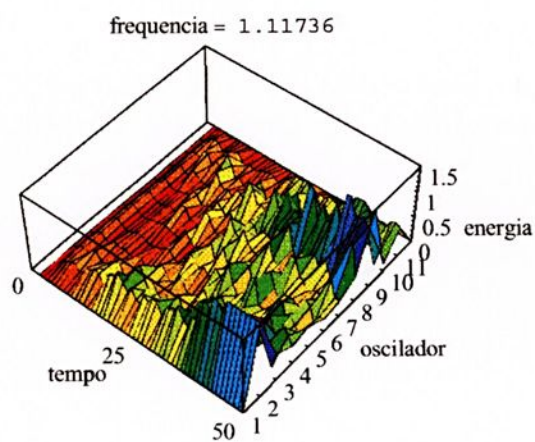
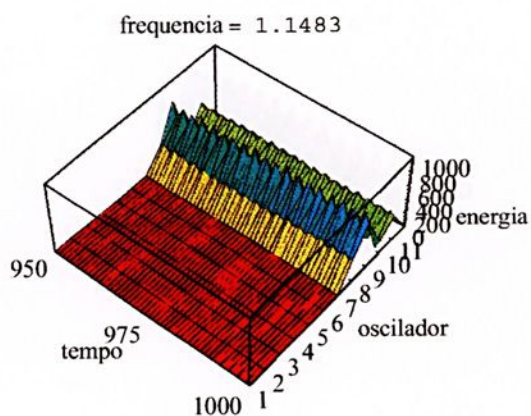
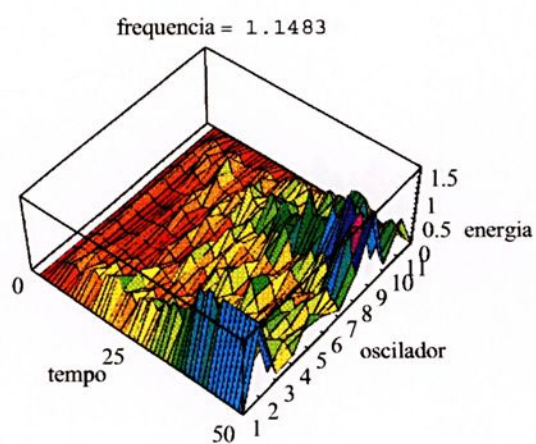
Figura 4.2: Gráficos de x_j , que indicam quanto cada mola está distendida ou comprimida, através do tempo. No topo de cada gráfico, está a frequência ω_f da força perturbativa periódica com a qual cada um dos osciladores da cadeia foi excitado. Os gráficos da esquerda correspondem a tempos pequenos (0-100), os da direita mostram tempos maiores (900-1000).

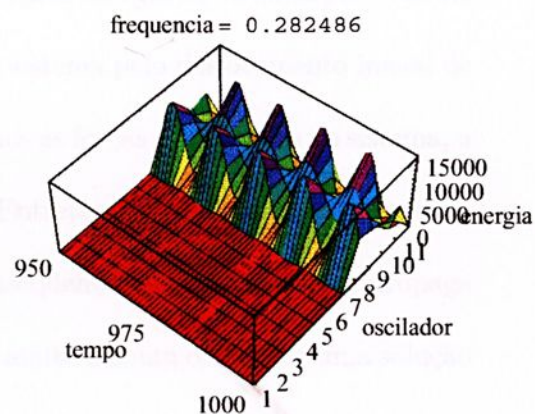
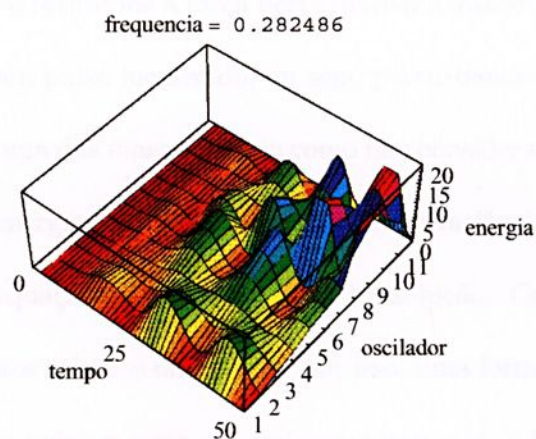
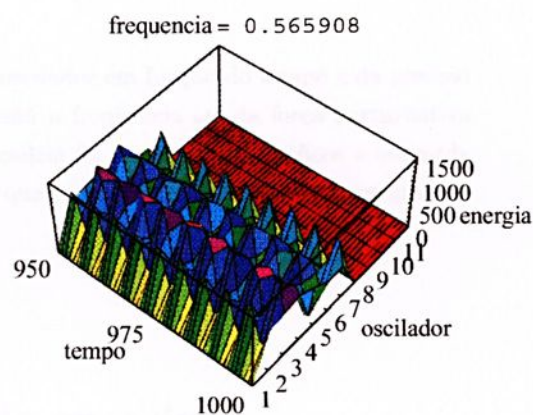
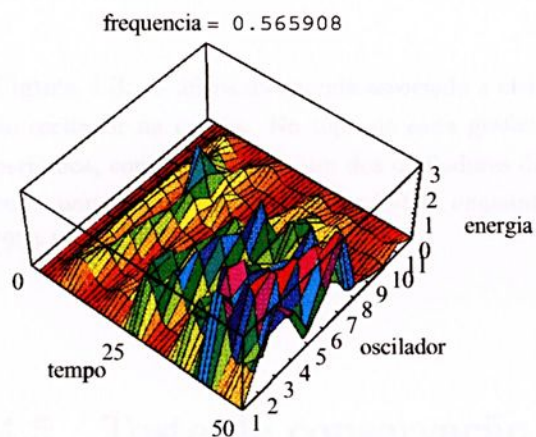
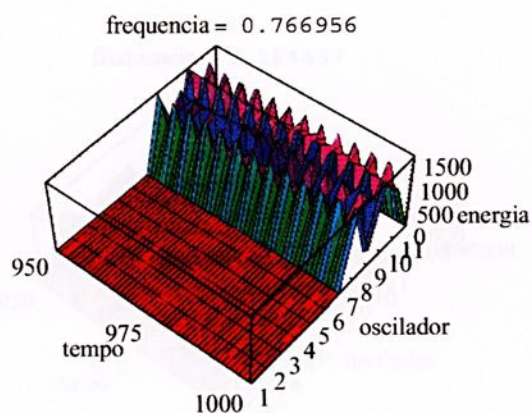
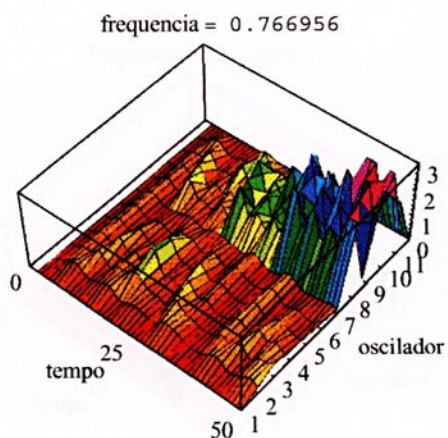
Os gráficos mostrados na figura 4.2 indicam a existência de localização de energia. Isto pode ser observado pela localização das maiores amplitudes (x). Afim de verificarmos a localização da energia, calculamos a energia de cada oscilador da cadeia [19]. O cálculo da energia foi feito associando ao j -ésimo oscilador, a energia cinética correspondente à j -ésima massa mais a metade da energia potencial de cada uma das molas às quais a massa está acoplada, assim:

$$E_j(t) = \frac{m\dot{x}_j^2(t)}{2} + \frac{k_{j-1}}{4}[x_j(t) - x_{j-1}(t)]^2 + \frac{k_j}{4}[x_{j+1}(t) - x_j(t)]^2. \quad (4.12)$$

Deve-se lembrar que as molas presas às extremidades fixas são exceção à expressão acima. Nestes casos, a energia potencial da mola é integralmente associada ao oscilador da extremidade correspondente. Os gráficos da energia individual de cada oscilador, estão mostrados na figura 4.3 e confirmam os padrões de localização indicados pelos gráficos anteriores relacionados às amplitudes.







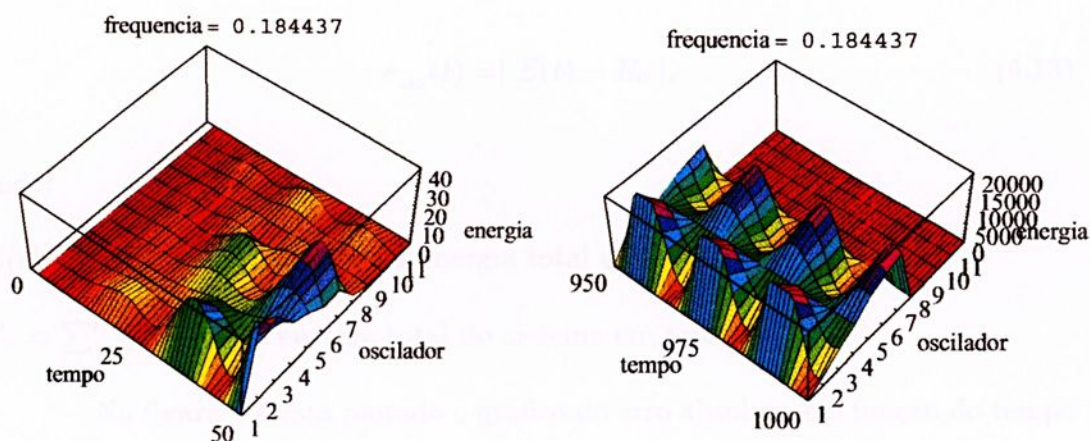


Figura 4.3: Gráficos da energia associada a cada oscilador em função do tempo e da posição do oscilador na cadeia. No topo de cada gráfico está a frequência ω_f da força perturbativa periódica, com a qual cada um dos osciladores da cadeia foi excitado. Os gráficos à esquerda correspondem a tempos pequenos (0-50), enquanto que os da direita mostram tempos maiores (950-1000).

4.5 Teste de conservação da energia

Se retiramos a força perturbativa e introduzirmos energia no sistema por meio de um pulso localizado, ou seja, perturbando o sistema pelo deslocamento inicial de uma das massas, então como não consideramos as forças dissipativas no sistema, a energia total deverá permanecer invariável. Entretanto, os cálculos numéricos das equações introduzem erros na solução. Conseqüentemente, este erro se propaga nos cálculos da energia, por isso, uma forma muito comum de avaliar uma solução numérica, para um sistema como este, é testando a estabilidade da solução por

meio da conservação da energia. Para isto, definimos o erro absoluto:

$$e_{abs}(t) = |E(t) - E_0|, \quad (4.13)$$

onde:

$E(t) = \sum_{j=2}^{n+1} E_j(t)$ é o valor da energia total do sistema no tempo t e

$E_o = \sum_{j=2}^{n+1} E_j(0)$ é a energia total do sistema em $t=0$.

Na figura 4.4 está plotado o gráfico do erro absoluto em função do tempo de $t = 0$ até $t = 1000$.

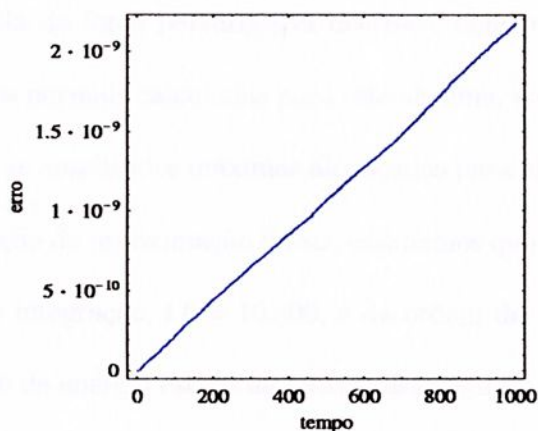


Figura 4.4: Gráfico do erro absoluto em função do tempo, de $t=0$ a $t=1000$.

Observamos que na equação (4.14) obtida com a aproximação linear para os pontos do gráfico da figura 4.4, o termo relacionado com o coeficiente angular da reta é da ordem de 10^{-12} o que mostra que a velocidade com que o erro cresce é consideravelmente pequena.

$$e_{aprox}(t) = -1.36 \times 10^{-12} + 2.15 \times 10^{-12}t, \quad (4.14)$$

4.6 Discussão dos resultados

Podemos verificar que para esta cadeia de dez osciladores, nas condições em que foram descritas, o padrão de localização das maiores amplitudes de oscilações concorda com o padrão de localização da energia, ou seja, para as frequências testadas, a região da cadeia em que ocorre uma localização das maiores amplitudes coincide com a região em que ocorre a localização da energia. Concluimos, portanto, que houve uma relação direta entre a localização das maiores amplitudes e a localização da energia. Os gráficos da energia também mostram que, quando a frequência da força perturbativa decresce, dentro do conjunto de frequências dos modos normais calculadas para este sistema, a energia localizada (e conseqüentemente as amplitudes máximas alcançadas para algum t fixo aumenta.

Pela projeção da aproximação linear, estimamos que o erro absoluto para um tempo final de integração, $tf = 10.000$, é da ordem de 10^{-8} , ou seja, o erro esperado no cálculo da energia estará na oitava casa decimal, o que é considerado um resultado satisfatório, pois a localização da energia se dá em um tempo bem menor.



Capítulo 5

Localização da energia no modelo vibracional planar

Analogamente ao que foi feito no capítulo 4, onde verificamos a localização da energia em uma cadeia unidimensional de osciladores, investigaremos aqui a localização da energia no modelo vibracional planar, adotado neste trabalho para modelar a dinâmica do DNA. Utilizando este modelo, simulamos uma sequência de 168 pares de nucleotídeos correspondente à região promotora $T7A_1$ [20, 21]. Simulamos também sequências "aleatórias" de mesmo comprimento e comparamos os resultados com aqueles da sequência do promotor. A energia é introduzida no sistema por uma força periódica que atua sobre cada oscilador. Verificamos também a estabilidade da solução, testando a conservação da energia no sistema. Para isto, retiramos a força periódica e introduzimos energia apenas por meio de um pulso, isto é, somente pelo deslocamento inicial de uma das massas.



5.1 O potencial de emparelhamento das bases

Como já discutido antes, a não homogeneidade do código genético é considerada neste modelo pela diferenciação do potencial de interação do emparelhamento entre as bases.

Esse emparelhamento, responsável em grande parte por manter juntas as fitas da dupla hélice, deve-se às interações de pontes de hidrogênio que se formam entre uma base pirimidina de uma fita e uma base purina da outra fita ou, mais especificamente, entre citosina e guanina ou entre adenina e timina. É usado, para representar estas interações, um potencial de Morse [8, 12] dado por:

$$V_j(\sqrt{2}y_j) = \xi \{ \exp[-a\sqrt{2}y_j] - 1 \}^2. \quad (5.1)$$

Como podemos ver, o potencial de Morse possui dois parâmetros: ξ e a que estão associados, respectivamente, à profundidade e à largura do poço de potencial. ξ representa, portanto, a energia de dissociação, ou seja, energia necessária para romper as pontes de hidrogênio formadas entre os emparelhamentos descritos acima. Cálculos dessa energia [16] apontam os seguintes resultados*: para o emparelhamento do tipo A-T,

$$\xi_{A-T} = 7.00 \text{ Kcal/mol} \simeq 0.304 \text{ eV}, \quad (5.2)$$

e para o emparelhamento do tipo C-G,

$$\xi_{C-G} = 16.79 \text{ Kcal/mol} \simeq 0.728 \text{ eV}. \quad (5.3)$$

*fatores de conversão: $1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$ e $1 \text{ cal} = 4.184 \text{ J}$, e o número de avogrado $N_A = 6,0220 \times 10^{23} \text{ partículas/mol}$

Portanto, obtemos com esses valores uma diferenciação entre os dois tipos de emparelhamento. Com relação ao parâmetro a do potencial de Morse adotamos o mesmo para ambos tipos de associações [13, 12],

$$a = 2.0 \text{ \AA}^{-1}.$$

Assim, o potencial específico para as associações do tipo A-T ou T-A é:

$$V_{A-T}(\sqrt{2}y_j) = V_{T-A}(\sqrt{2}y_j) = \xi_{A-T} \{ \exp[-a\sqrt{2}y_j] - 1 \}^2, \quad (5.4)$$

enquanto o potencial para as associações do tipo C-G ou G-C é:

$$V_{C-G}(\sqrt{2}y_j) = V_{G-C}(\sqrt{2}y_j) = \xi_{C-G} \{ \exp[-a\sqrt{2}y_j] - 1 \}^2. \quad (5.5)$$

Na figura 5.1 estão traçados os gráficos de cada um dos potenciais (5.4) e (5.5) em função de y . Podemos observar que para uma mesma energia ($E_0 = 0.15$ eV, por exemplo) o potencial do emparelhamento A-T, mais "flexível", alcança um deslocamento y maior que aquele alcançado pelo potencial C-G, mais "rígido".

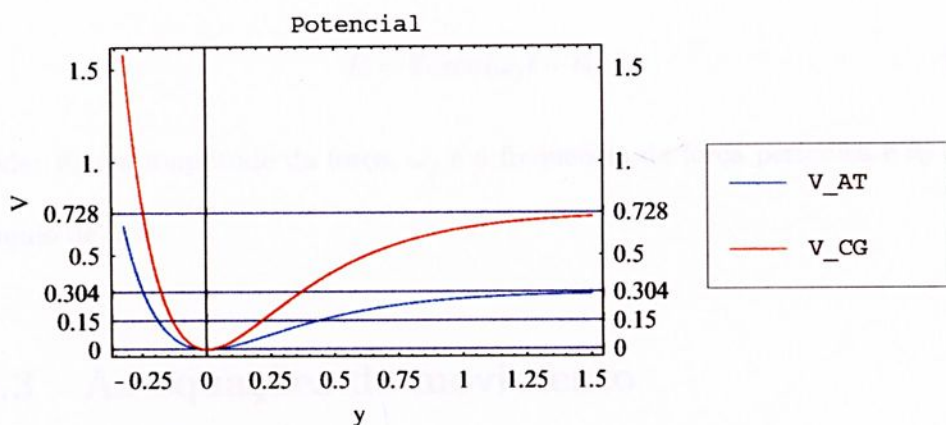


Figura 5.1: Gráfico dos potenciais V_{A-T} e V_{C-G} dados nas equações (5.4) e (5.5). Aqui y representa a distância que a massa se desvia da posição de equilíbrio.

Então a Hamiltoniana do modelo, obtida a partir da equação (3.4) é

$$H = \sum_{j=1}^n \frac{p_j^2}{2m} + \sum_{j=1}^{n+1} \frac{k}{2} (y_j - y_{j-1})^2 + \sum_{j=1}^{n+1} \xi \{ \exp[-a\sqrt{2}y_j] - 1 \} , \quad (5.6)$$

se o j -ésimo sítio na cadeia representar um nucleotídeo que contém uma adenina ou uma timina o termo ξ assume o valor dado em (5.2), ou no caso em que representar um nucleotídeo com uma guanina ou uma citosina ξ equivale então ao valor dado em (5.3).

5.2 Introdução de uma força periódica no sistema

A exemplo do que foi feito no capítulo 4, para investigarmos as propriedades de localização da energia neste modelo introduzimos, para cada uma das massas, uma força periódica que varia senoidalmente com o tempo. Desta forma, a mesma quantidade de energia é inserida em cada oscilador simultaneamente. A denominada força perturbativa periódica é dada por:

$$F = F_0 \sin(\omega_f t + \theta_0) , \quad (5.7)$$

onde: F_0 é a amplitude da força, ω_f é a frequência da força periódica e θ_0 é um ângulo de fase.

5.3 As equações de movimento

A partir da Hamiltoniana (5.6) podemos obter as equações de movimento, sem considerar a ação das forças externas, para cada tipo de nucleotídeo na fita. Uma

vez que a força periódica é igual para qualquer oscilador, para considerá-la basta inserir o termo relativo à força (5.7) em cada uma das equações de movimento. Assim, se o j -ésimo oscilador da cadeia representa um nucleotídeo com uma base adenina ou com uma timina, a equação de movimento associada é:

$$m\ddot{y}_j = +k(-2y_j + y_{j-1} + y_{j+1}) + 2a\xi_{A-T} \exp[-ay_j] \{ \exp[-ay_j] - 1 \} + F_0 \sin(\omega_j t + \theta_0), \quad (5.8)$$

e para o caso em que ele representa um nucleotídeo com uma citosina ou uma guanina, a equação de movimento é:

$$m\ddot{y}_j = +k(-2y_j + y_{j-1} + y_{j+1}) + 2a\xi_{C-G} \exp[-ay_j] \{ \exp[-ay_j] - 1 \} + F_0 \sin(\omega_j t + \theta_0). \quad (5.9)$$

Observamos que nestas equações não é considerado nenhum tipo de amortecimento. A única força externa considerada é a força perturbativa periódica.

5.4 Localização da energia na cadeia do promotor T7A₁

Tomamos a sequência com 168 pares de nucleotídeos correspondentes à região promotora T7A₁ [20, 21], que possui uma das suas fitas com a sequência dada na figura 5.2.

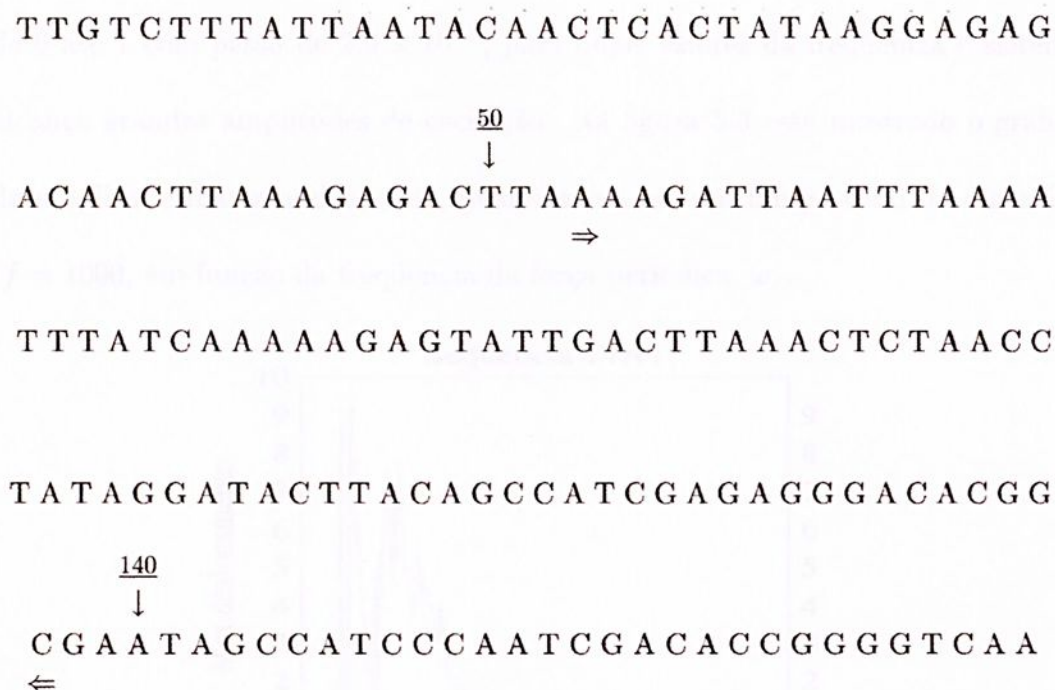


Figura 5.2: Seqüência de uma das fitas do promotor $T7A_1$. Lembramos que a outra fita pode ser deduzida pela regra de emparelhamento de *Chargaff* mencionada no capítulo 1. A região entre as duas setas horizontais delimitam a região com a qual a RNA polimerase pode interagir com a dupla hélice [20].

Associando cada nucleotídeo da seqüência dada na figura 5.2 a uma equação de movimento correspondente (5.8) ou (5.9), obtemos um sistema de equações diferenciais não homogêneas. O sistema foi resolvido numericamente utilizando o software *Mathematica* e como integrador novamente usamos o algoritmo de Runge-Kutta com um passo de 10^{-3} .

Como é razoável supor uma dependência da dinâmica do sistema com relação às frequências da força periódica, investigamos numericamente, no intervalo de 0 até 1 com passo de 2.5×10^{-3} , para quais valores da frequência o sistema alcança grandes amplitudes de oscilação. Na figura 5.3 está mostrado o gráfico da amplitude máxima alcançada, pelo sistema em um tempo total de dinâmica $tf = 1000$, em função da frequência da força periódica, ω_f .

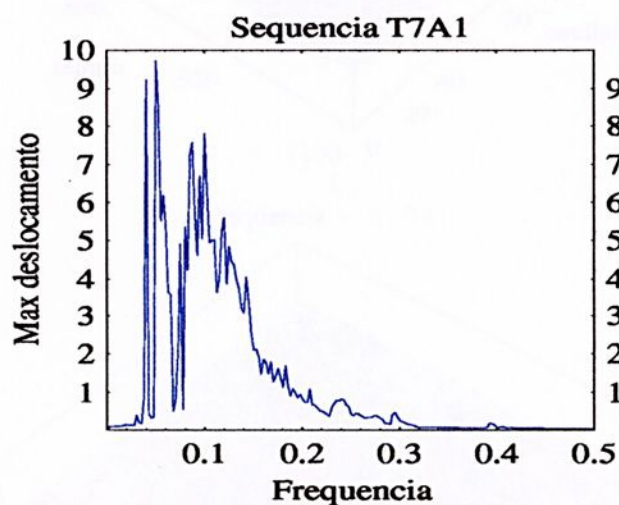


Figura 5.3: Gráfico da máxima amplitude alcançada pelo sistema em função da frequência da força periódica. No gráfico em detalhe os resultados para o intervalo de frequências experimentadas de 0 a 0.5 com um passo de 2.5×10^{-3} na frequência.

Os maiores valores para as amplitudes máximas alcançadas ocorrem em ordem decrescente para as seguintes frequências 0.05; 0.04; 0.1; e 0.0875.

Na figura 5.4 estão mostrados os gráficos tridimensionais do *deslocamento a partir da posição de equilíbrio de cada oscilador*, y_j (Å), em função do *tempo*, t (em picosegundos), e do *sítio do oscilador*, j , para estas frequências encontradas. Observamos que nestes gráfico as cores azul claro, azul escuro e vermelho indicam nesta ordem amplitudes crescentes de oscilação.

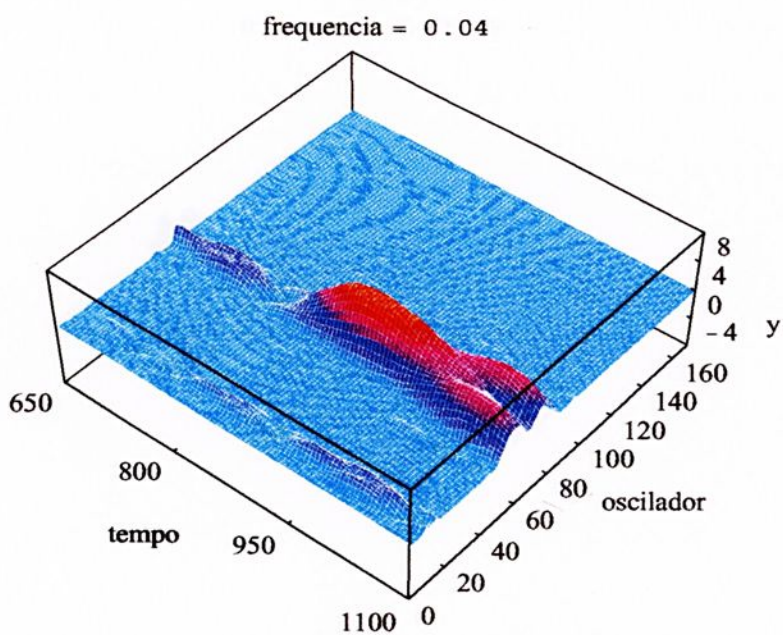
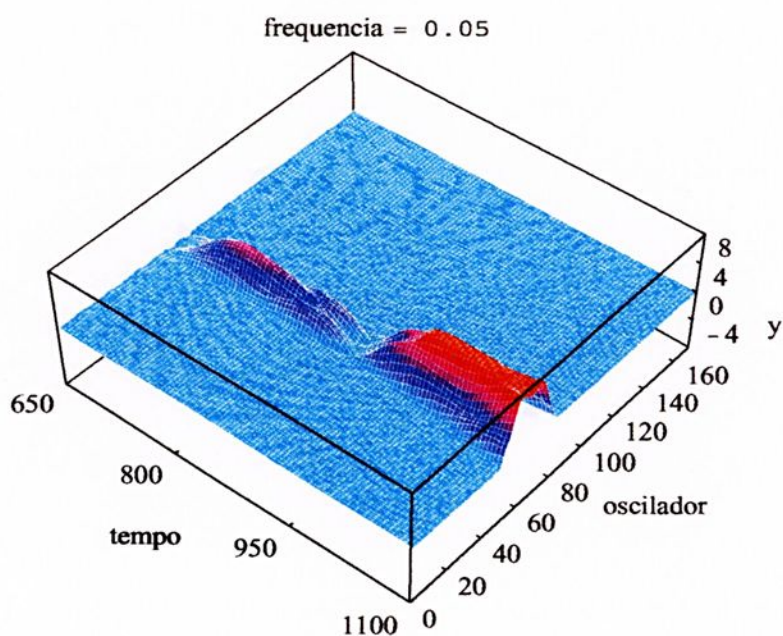


Figura 5.4: Localização tridimensional da densidade de energia no espaço de fase da cadeia de osciladores. A) em tempo, B) em frequência. As superfícies representam a densidade de energia no espaço de fase. As cores representam a densidade de energia, variando de azul (baixa densidade) a vermelho (alta densidade).

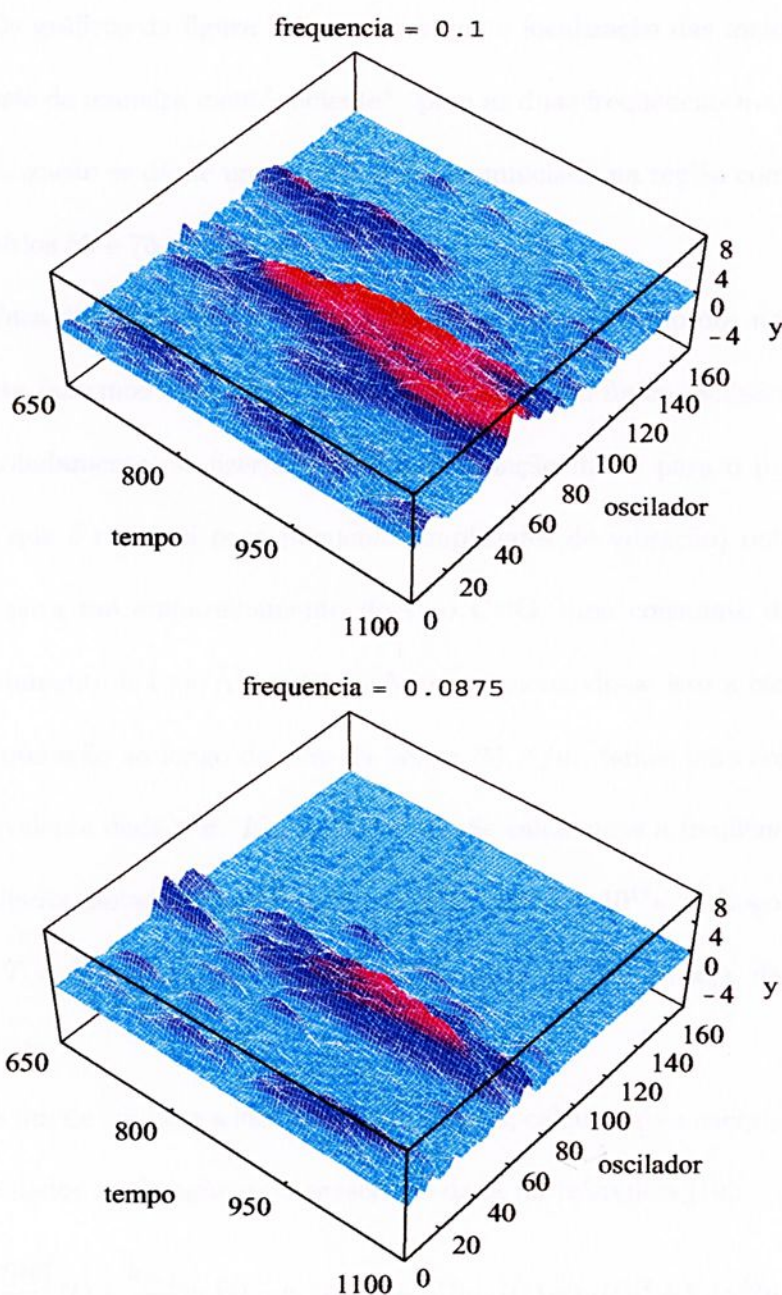


Figura 5.4: Gráficos tridimensionais do deslocamento a partir da posição de equilíbrio de cada oscilador ($\text{\AA}$) em função do tempo (ps) e do sítio do oscilador (localização na cadeia). A frequência da força periódica, ω_f , experimentada está indicada no topo de cada gráfico.

Os gráficos da figura 5.4 mostram que a localização das maiores amplitudes ocorre de maneira mais "eficiente" para as duas frequências mais baixas, e que a localização se dá de uma forma mais pronunciada na região compreendida entre os sítios 55 e 75 da cadeia.

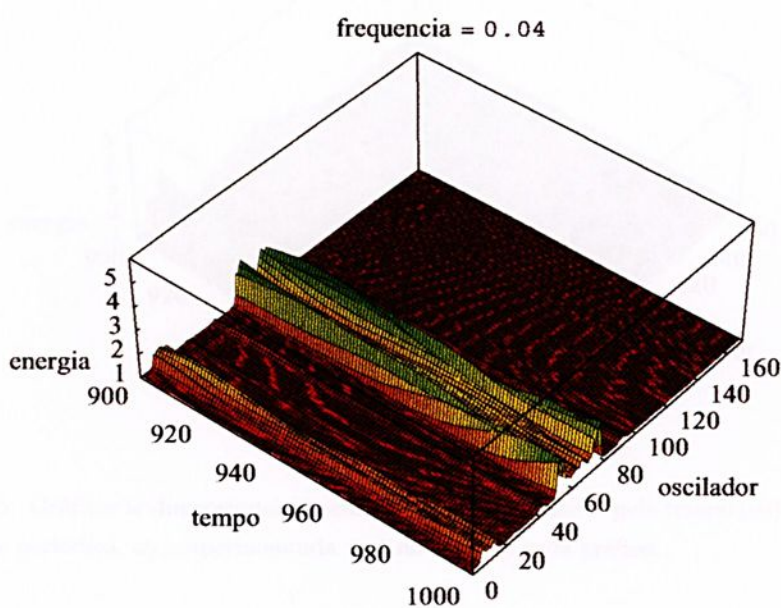
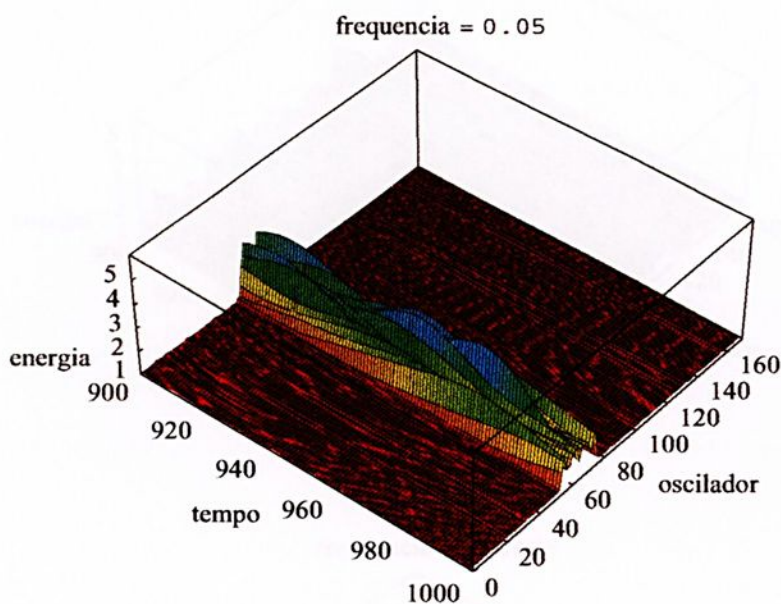
Para que possamos ter uma idéia da escala do tempo dos nossos resultados basta fazermos o seguinte exercício: tomando um único oscilador da nossa cadeia, isoladamente, se fizermos uma aproximação linear para o potencial de Morse (o que é razoável para pequenas amplitudes de vibração) obtemos, por exemplo, para um emparelhamento do tipo $C \equiv G$, uma constante de mola de aproximadamente $5.4 \text{ eV}/\text{\AA}^2$ ou 86.50 N/m , adicionando-se isto à constante de mola da interação ao longo da eixo da hélice, 24 N/m , temos uma constante de mola equivalente dada por: $K_{eq} \simeq 110 \text{ N/m}$. Se calculamos a frequência natural desde oscilador isolado obtemos $\omega = \sqrt{\frac{110}{5 \times 10^{-25}}} \simeq 4.7 \times 10^{12} \text{ s}^{-1}$. Logo o período dado por $T = \frac{2\pi}{\omega}$ vale aproximadamente $T \simeq 1.3 \times 10^{-12} \text{ s}$, ou seja, da ordem de picosegundos.

A fim de verificar a localização da energia, calculamos a energia associada a cada oscilador analogamente à prescrição dada na referência [19]:

$$E_j(t) = \frac{m\dot{y}_j^2}{2}(t) + \frac{k_{j-1}}{4}[y_j(t) - y_{j-1}(t)]^2 + \frac{k_j}{4}[y_{j+1}(t) - y_j(t)]^2 + V_j(\sqrt{2}y_j) \quad (5.10)$$

Os gráficos tridimensionais da *energia*, E_j (em elétron-volt), em função do *tempo*, $t(\text{ps})$ e do *sítio do oscilador*, j , está mostrado na figura 5.5 para cada um dos valores das frequências determinadas antes (0.05; 0.04; 0.1; 0.0875). Nestes

gráficos as cores vermelho, alaranjado, amarelo, verde claro e verde escuro indicam nesta ordem valores crescentes para a energia.



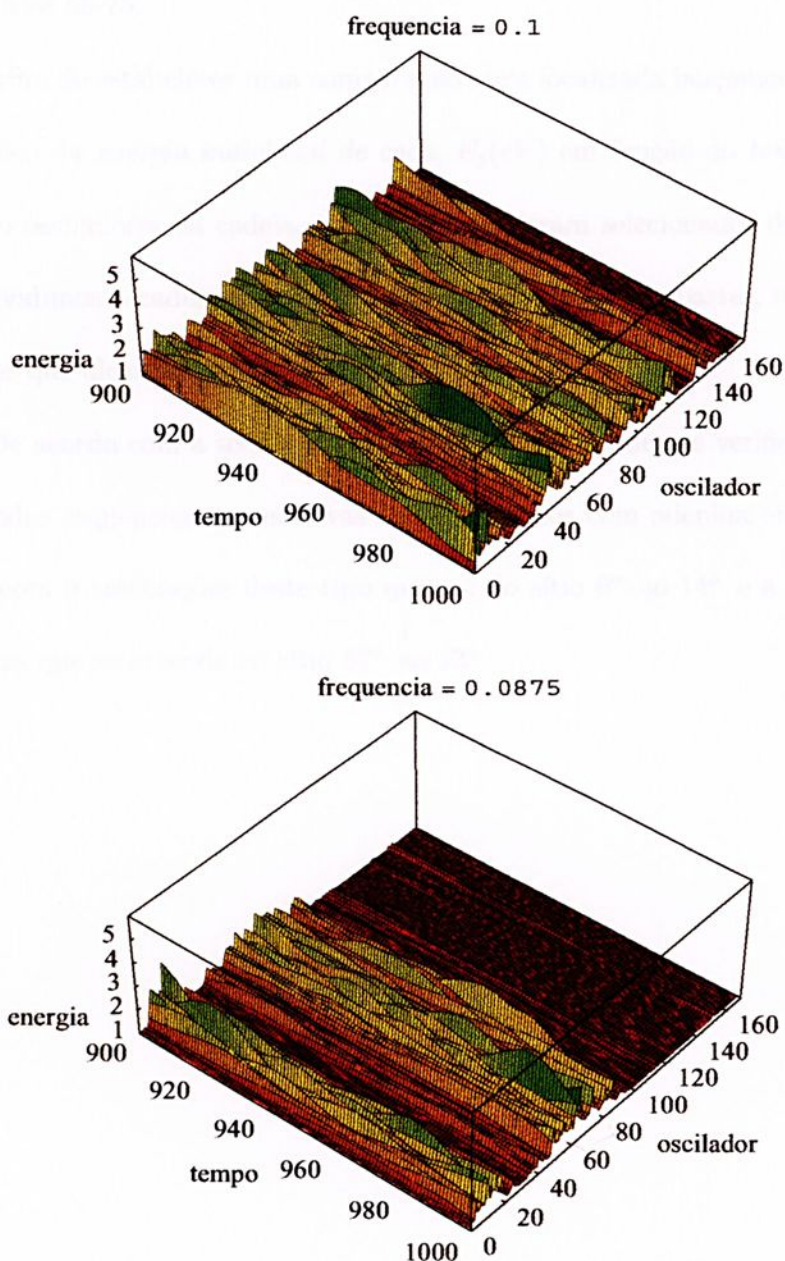


Figura 5.5: Gráficos tridimensionais da energia (eV) por oscilador pelo tempo (ps). A frequência da força periódica, ω_f , experimentada está no topo de cada gráfico.

Dos gráficos da figura 5.5, podemos verificar que a localização da energia ocorre nas mesmas regiões onde se concentraram as maiores amplitudes, ou seja,

entre os sítios 55-75.

Afim de estabelecer uma comparação mais focalizada lançamos na figura 5.6 o gráfico da *energia individual de cada*, $E_j(\text{eV})$ em função do *tempo*, $t(\text{ps})$, para cinco osciladores da cadeia. Os osciladores foram selecionados da seguinte forma: dividimos a cadeia com os 168 nucleotídeos em cinco partes, e tomamos o oscilador que alcança a maior energia de cada parte.

De acordo com a seqüência dada na figura 5.2 podemos verificar que há duas grandes seqüências consecutivas de nucleotídeos com adenina ou timina a primeira com 9 associações deste tipo que vai do sítio 6º ao 14º e a segunda e maior delas que se estende do sítio 57º ao 73º.

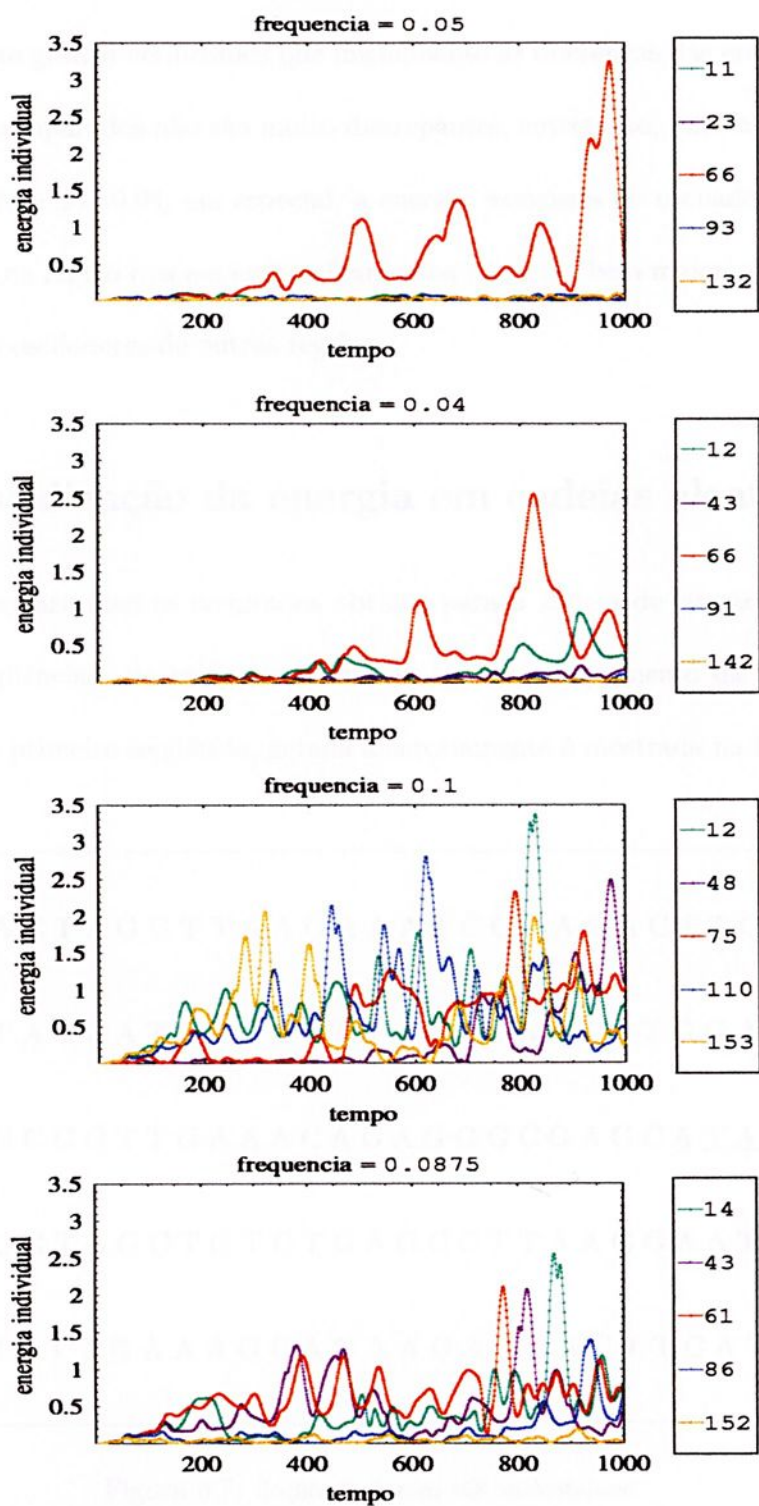


Figura 5.6: Gráfico de energia individual do j -ésimo oscilador pelo tempo. Na legenda estão indicados os sítios de cada oscilador segundo a sequência do promotor T7A₁.

Neste gráfico verificamos que inicialmente as diferenças das energias dos osciladores comparados não são muito discrepantes, entretanto, para as frequências $\omega_f=0.05$, e $\omega_f=0.04$, em especial, a energia associada ao oscilador do sítio 66 (inserido na região rica em emparelhamentos A-T) são bem maiores do que as energias dos osciladores de outras regiões.

5.5 Localização da energia em cadeias aleatórias

Afim de compararmos os resultados obtidos para a cadeia do promotor T7A₁ geramos seqüências de nucleotídeos com o mesmo comprimento da cadeia do promotor, a primeira seqüência, gerada aleatoriamente é mostrada na figura 5.7

AGCAAAGCTAGGTTAAGAAATCCGAGACTTCCCTA
CGATCTACCATATCCTTTACGTAATGTGGACCTA
TTAGTGCGGTTGAAACAGAGGGCGAGCATATAGG
GACAACGTAGCTGTCTGAGCCTTAAGGAATGGAT
AACCGTTTGAAAGCAGAAGACGACTTCATGAG

Figura 5.7: Seqüência-I, com 168 nucleotídeos.

A segunda seqüência difere da seqüência-I, como podemos ver na figura 5.8, por somente dois nucleotídeos que foram substituídos afim de "quebrar" uma regularidade que ocorreu na seqüência aleatória (cinco nucleotídeos do tipo A ou T consecutivos), no sítio 69 substituímos um T por um C e no sítio 98 substituímos um A também por um C.

AGCAAAGCTAGGTTAAGAAATCCGAGACTTCCCTA
CGATCTACCATATCCTTTACGTAATGTGGACCTA
CTAGTGCGGTTGAAACAGAGGGCGAGCATCTAGG
GACAACGTAGCTGTCTGAGCCTTTAAGGAATGGAT
AACCGTTTGAAAGCAGAAGACGACTTCATGAG

Figura 5.8: Seqüência-II, com 168 nucleotídeos.

Na terceira seqüência, mostrada na figura 5.9, inserimos modificações adicionais com o objetivo de excluir a ocorrência de trechos com mais de dois nucleotídeos consecutivos do tipo A ou T.

AGCACACTAGGTCACGACATCCGAGACTTCTCTA
CGATCTACCACATCCTGTACGTCATGTGGACCTA
CTAGTGCGGTTGACACAGAGTGCGAGCATCTGAT
GACAACGTAGCTGTCTGAGTCTCAAGGACTGAGT
CACTGTGTGACAGCAGAAGACGACTTCATGAG

Figura 5.9: Seqüência-III, com 168 nucleotídeos.

Da mesma forma investigamos, numericamente, para quais freqüências da força periódica ocorrem as maiores amplitudes de oscilação nestas seqüências (I-II e III). Na figura 5.10 está plotado o gráfico das máximas amplitudes alcançadas pelos sistemas simulados em função da freqüência da força periódica.

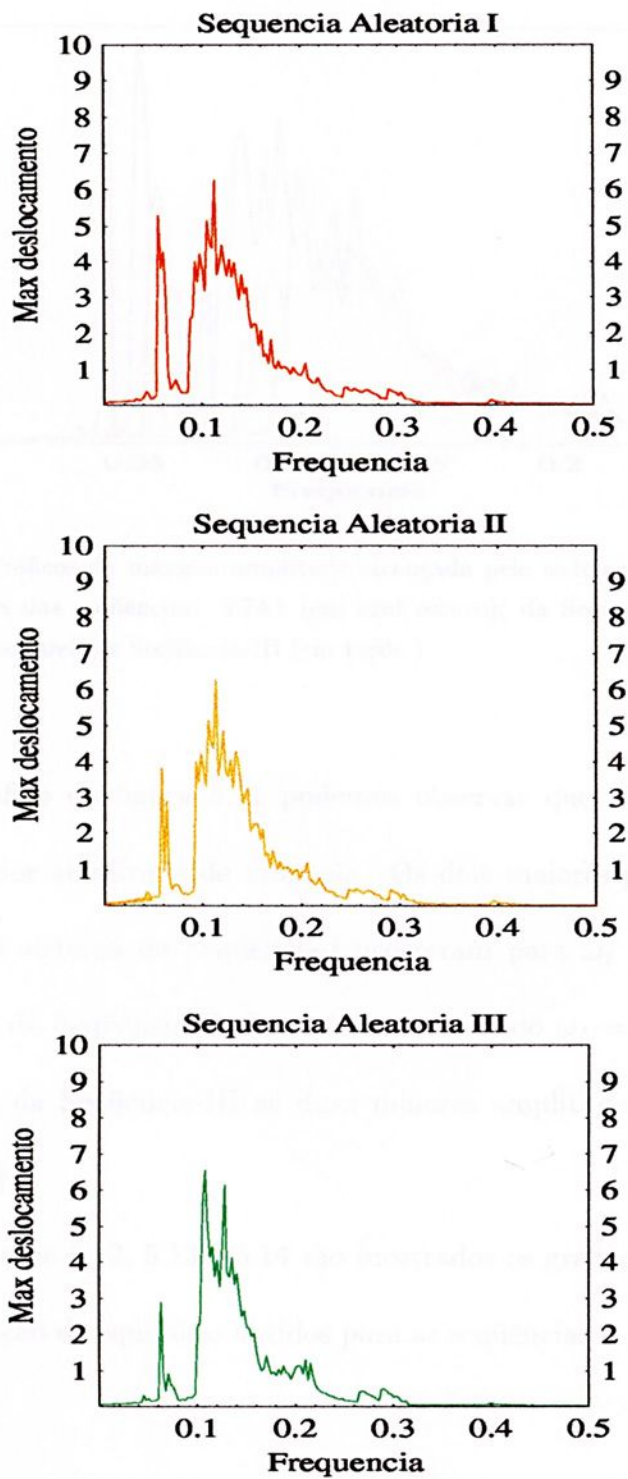


Figura 5.10: Gráficos da máxima amplitude alcançada pelo sistema em função da frequência da força periódica.

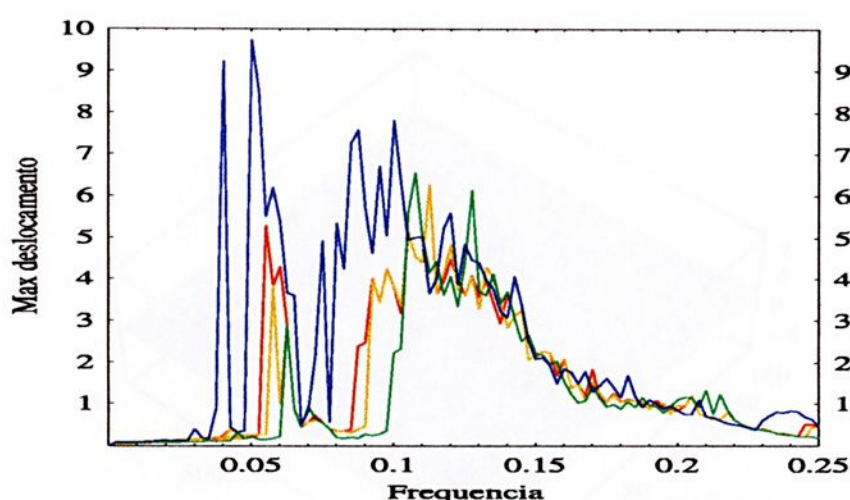


Figura 5.11: Gráficos da máxima amplitude alcançada pelo sistema função da frequência da força periódica das seqüências: T7A1 (em azul escuro); da Seqüência-I (em vermelho); Seqüência-II (em amarelo) e Seqüência-III (em verde).

No gráfico da figura 5.11 podemos observar que a cadeia do promotor alcança maior amplitude de vibração. Os dois maiores picos da amplitude máxima para o sistema da Seqüência-I ocorreram para $\omega_f = \{0.055, 0.1055\}$; para o sistema da Seqüência-II eles ocorreram quando $\omega_f = \{0.0575, 0.1125\}$ e para o sistema da Seqüência-III as duas maiores amplitudes se dão em $\omega_f = \{0.1025, 0.1275\}$.

Nas figuras 5.12, 5.13 e 5.14 são mostrados os gráficos do deslocamento a partir da posição de equilíbrio obtidos para as seqüências I, II e III, respectivamente.

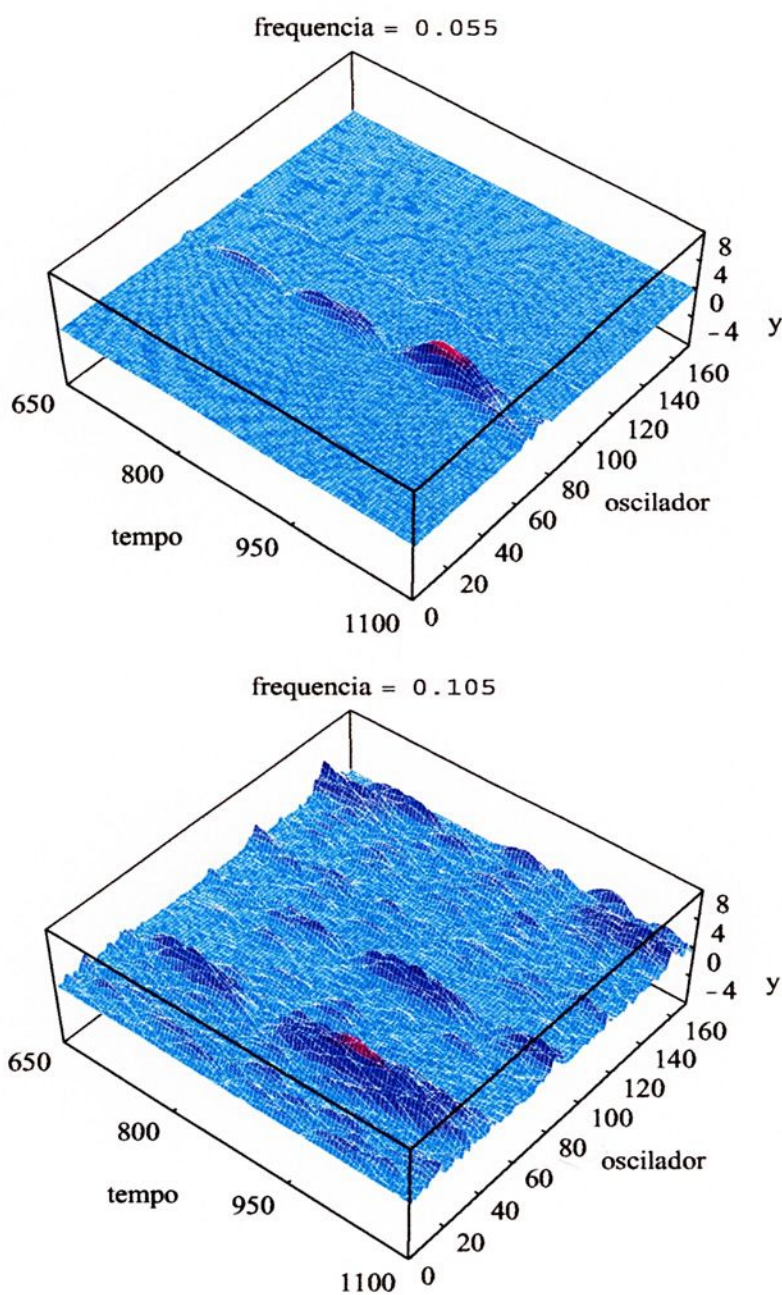


Figura 5.12: Gráficos tridimensionais do deslocamento a partir da posição de equilíbrio ($\text{\AA}$) por oscilador pelo do tempo (ps) para a Seqüência-I. A freqüência da força periódica, ω_f , experimentada está no topo de cada gráfico.

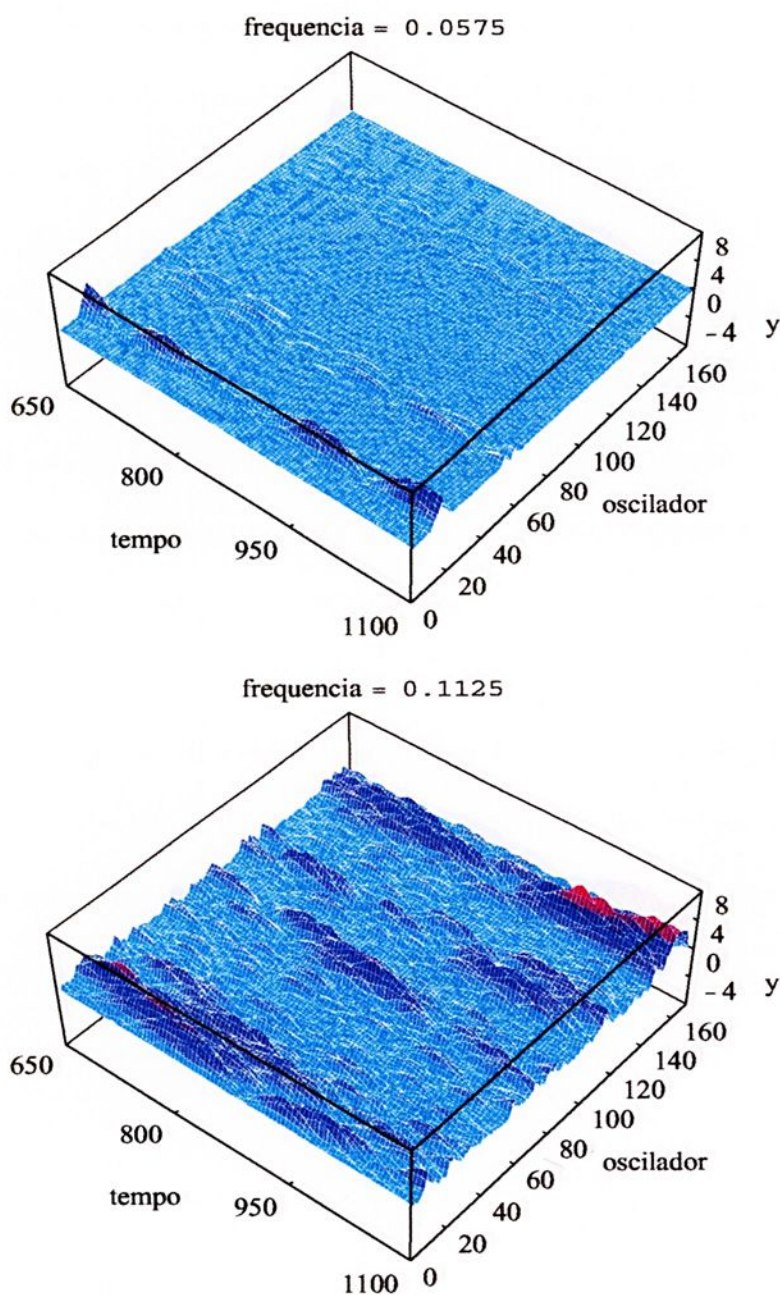


Figura 5.13: Gráficos tridimensionais do deslocamento a partir da posição de equilíbrio ($\text{\AA}$) por oscilador pelo do tempo (ps) para a Sequência-II. A frequência da força periódica, ω_f , experimentalada está no topo de cada gráfico.

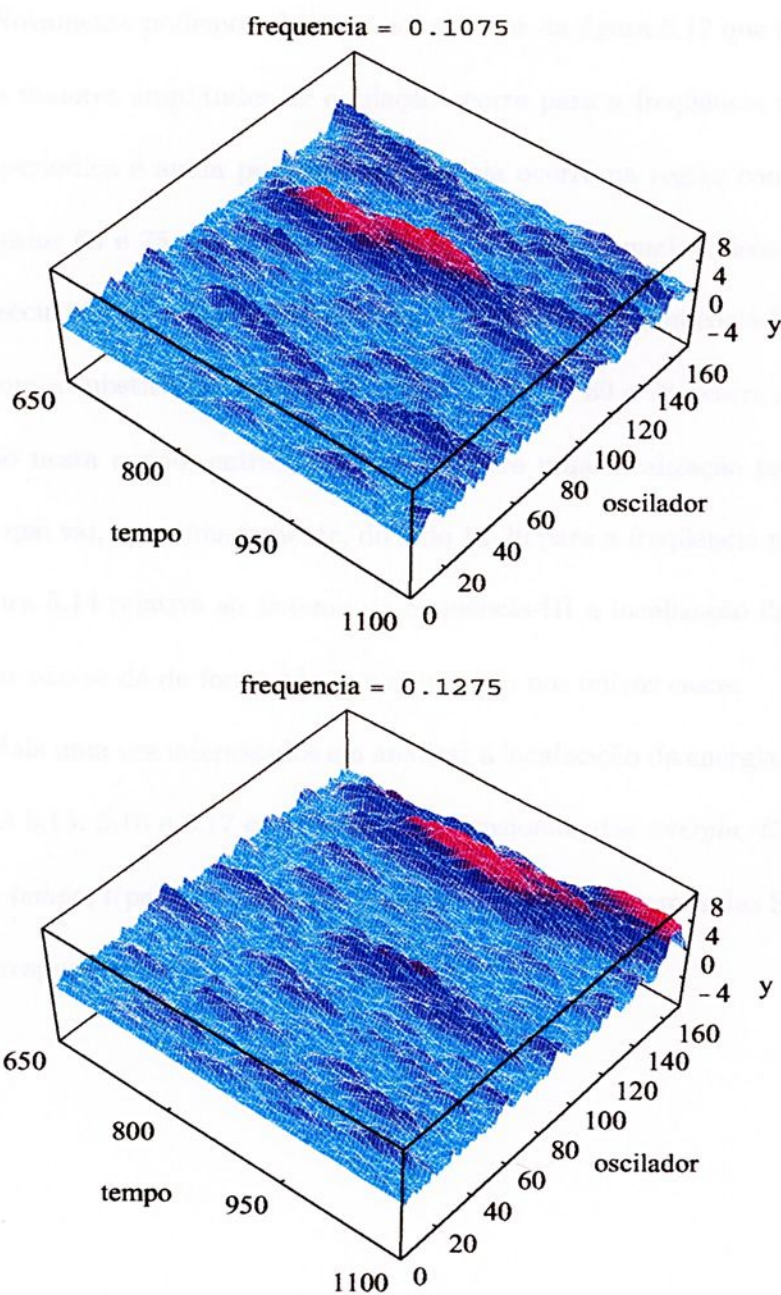


Figura 5.14: Gráficos tridimensionais do deslocamento a partir da posição de equilíbrio ($\text{\AA}$) por oscilador pelo do tempo (ps) para a Sequência-III. A frequência da força periódica, ω_f , experimentalada está no topo de cada gráfico.

Novamente podemos observar nos gráficos da figura 5.12 que uma localização das maiores amplitudes de oscilação ocorre para a frequência mais baixa da força periódica e ainda podemos ver que ela ocorre na região compreendida entre os sítios 65 e 75, onde há uma sequência de cinco nucleotídeos de tipo A ou T consecutivos. Na figura 5.13 referente ao sistema da Sequência-II podemos ver que com a substituição dos nucleotídeos dos sítios 69 e 98 ocorre a perda da localização nesta região, entretanto, ainda ocorre uma localização pronunciada na região que vai, aproximadamente, do sítio 10-20 para a frequência mais baixa. Já na figura 5.14 relativa ao sistema da Sequência-III a localização das maiores amplitudes não se dá de forma tão evidente como nos outros casos.

Mais uma vez interessados em analisar a localização da energia plotamos, nas figuras 5.15, 5.16 e 5.17 os gráficos tridimensionais das *energia*, $E_j(\text{eV.})$, em função do *tempo*, $t(\text{ps})$ e do *sítio do oscilador*, j , para os sistemas das Sequências I, II e III, respectivamente.



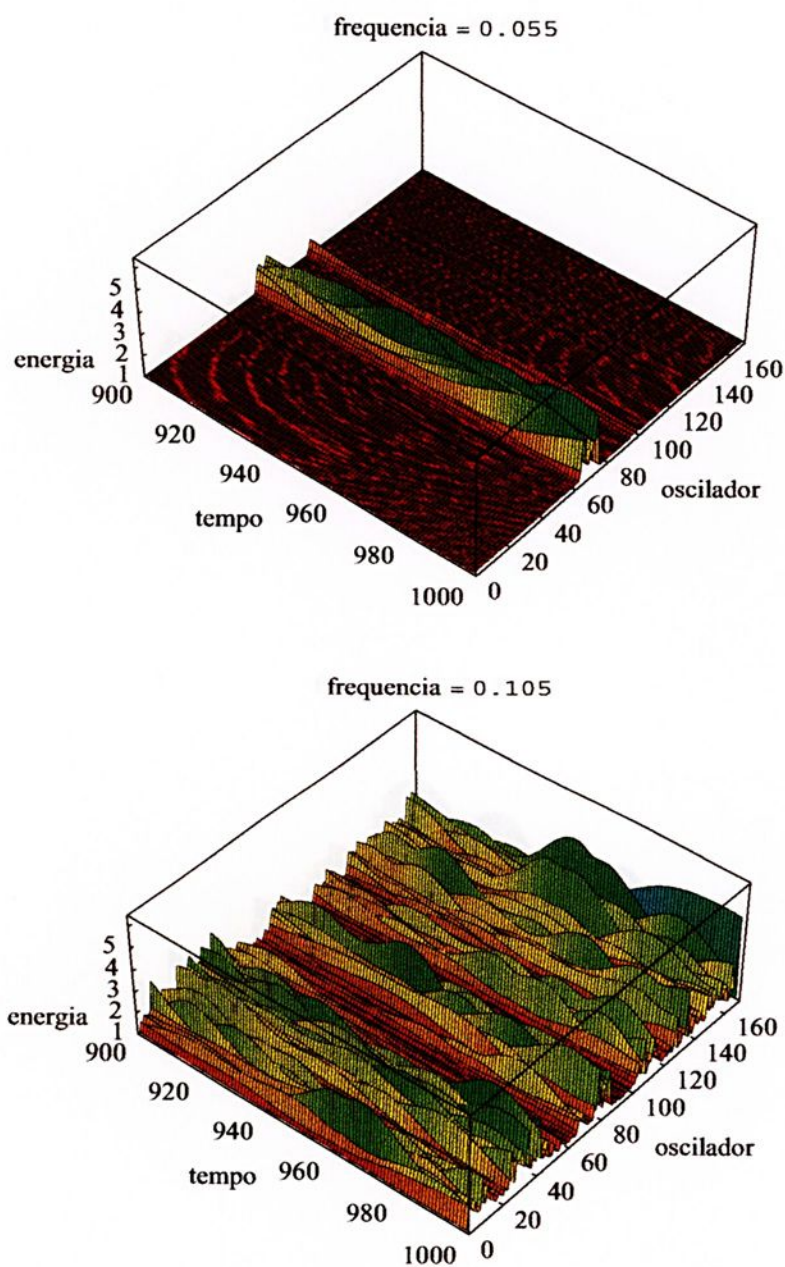


Figura 5.15: Gráficos tridimensionais da energia (eV.) por oscilador pelo tempo (ps), da Sequência-I. A frequência da força periódica, ω_f , experimentada está no topo de cada gráfico.

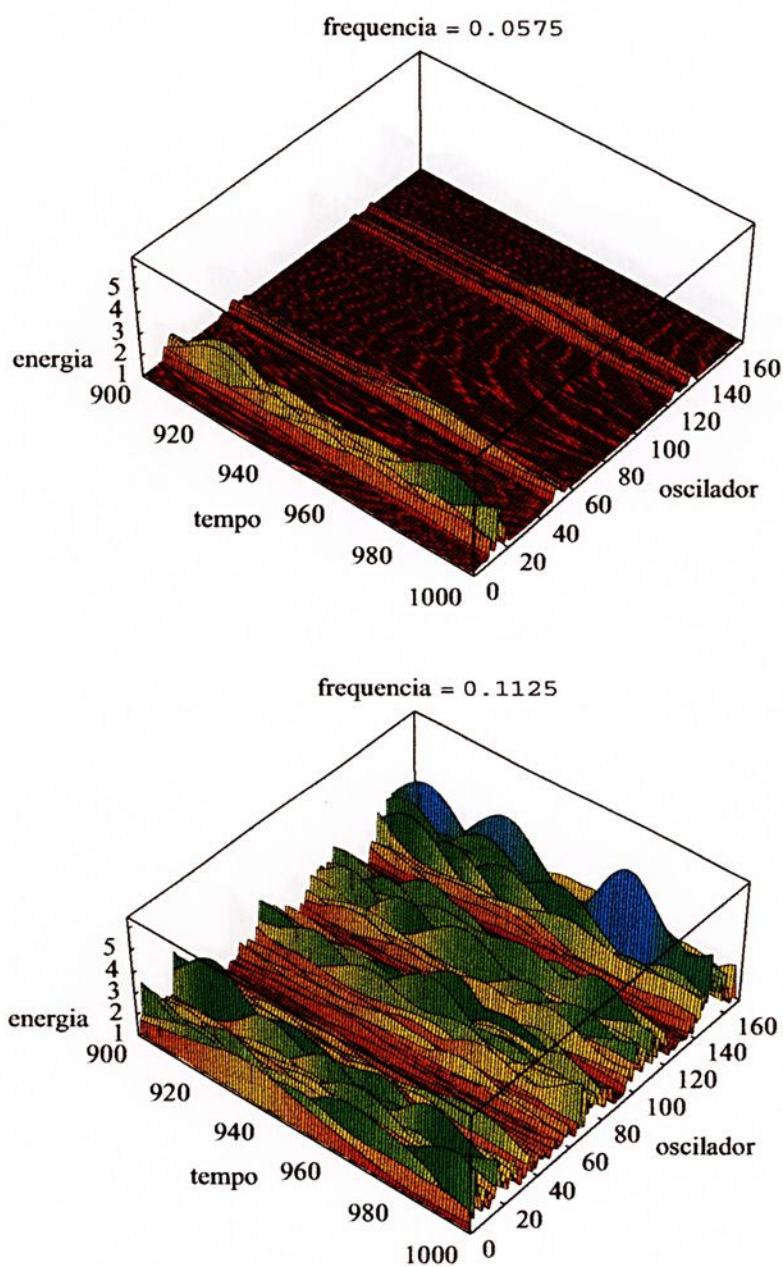


Figura 5.16: Gráficos tridimensionais da energia (eV.) por oscilador pelo tempo (ps), da Sequência-II. A frequência da força periódica, ω_f , experimentada está no topo de cada gráfico.

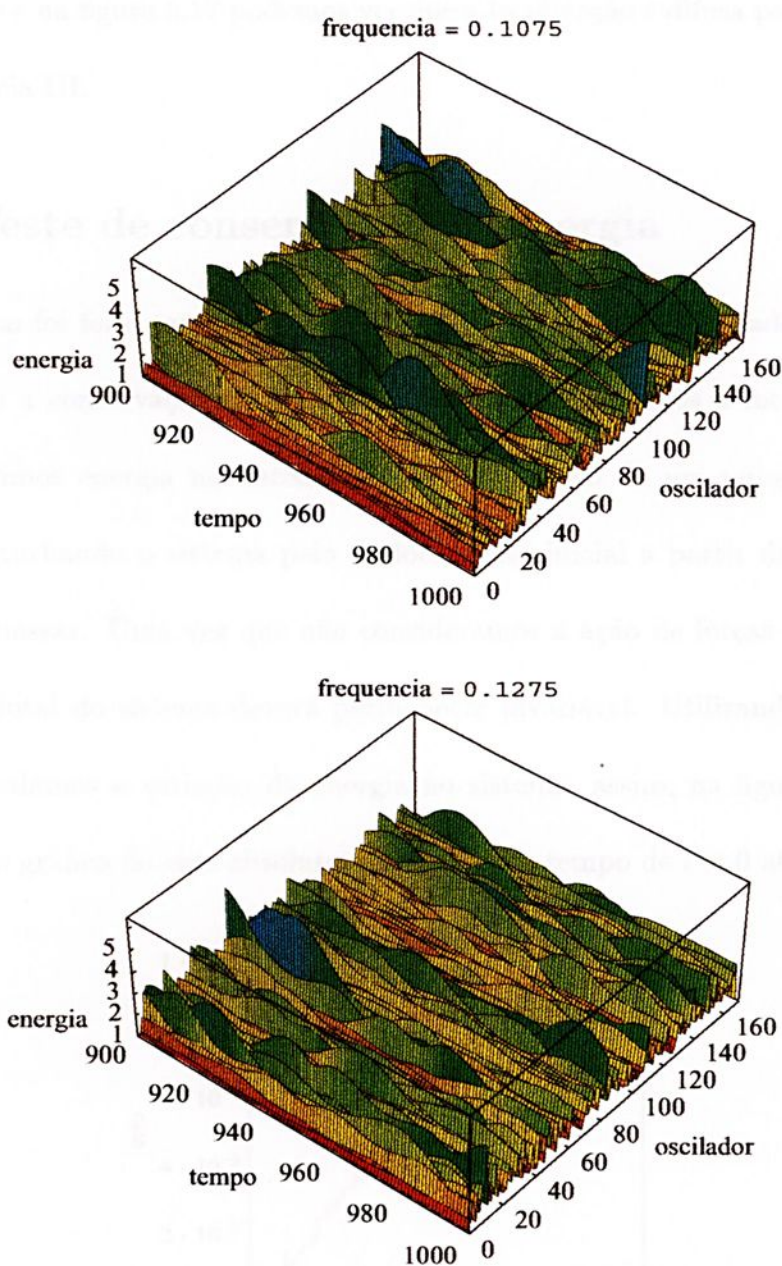


Figura 5.17: Gráficos tridimensionais da energia (eV.) por oscilador pelo tempo (ps), da Sequência-III. A frequência da força periódica, ω_f , experimentada está no topo de cada gráfico.

Nos gráficos das figuras 5.15 e 5.16 podemos observar que, novamente, a localização da energia ocorre nas mesmas regiões onde se concentraram as maiores

amplitudes e na figura 5.17 podemos ver que a localização é difusa para o sistema da Sequência-III.

5.6 Teste de conservação da energia

Assim como foi feito também no capítulo 4, testamos a estabilidade da solução verificando a conservação da energia. Para tanto, retiramos a força periódica e introduzimos energia no sistema somente por meio de um pulso localizado, isto é, perturbando o sistema pelo deslocamento inicial a partir da posição de uma das massas. Uma vez que não consideramos a ação de forças dissipativas, a energia total do sistema deverá permanecer invariável. Utilizando a equação (4.13) calculamos a variação da energia no sistema, assim, na figura 5.18 está mostrado o gráfico do erro absoluto em função do tempo de $t = 0$ até $t = 1000$.

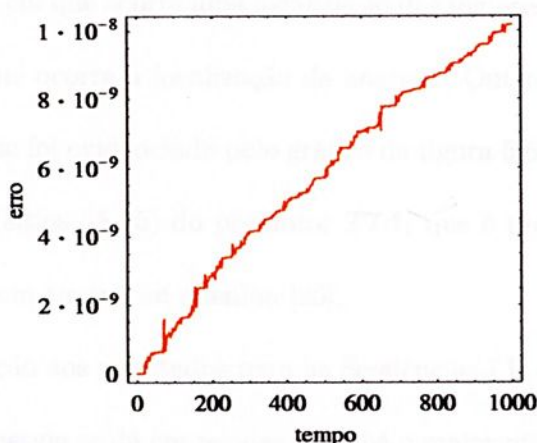


Figura 5.18: Gráfico do erro absoluto em função do tempo, de $t = 0$ a $t = 1000$.

A aproximação linear para o conjunto dos pontos do gráfico da figura 5.18 é dada pela reta:

$$e_{\text{approx}}(t) = -1.21 \times 10^{-10} + 1.50 \times 10^{-11}t,$$

Aqui também podemos ver, pela aproximação linear, que a "velocidade" de crescimento do erro é reduzida uma vez que o coeficiente angular desta reta é da ordem de 10^{-11} .

5.7 Discussão dos resultados

A partir dos gráficos do deslocamento e da energia mostrados podemos observar que, nas condições em que foram descritas, o padrão de localização das maiores amplitudes de oscilações concorda com o padrão de localização da energia, ou seja, a região da cadeia em que ocorre uma localização das maiores amplitudes coincide com aquela em que ocorre a localização da energia. Outro ponto importante a ser observado e que foi evidenciado pelo gráfico da figura 5.6 é que tal localização ocorre na região (sítios 55-75) do promotor $T7A_1$ que é proporcionalmente rica em nucleotídeos com timina ou adenina [26].

Com relação aos resultados para as Sequências I, II e III verificamos que a localização da energia se dá em regiões onde há o maior número de nucleotídeos com Adenina ou Timina consecutivos e que quanto maior o número desses nucleotídeos consecutivos mais evidente é a localização. Na Sequência-III, da qual,

praticamente, excluímos estas seqüências consecutivas não houve a localização.

No teste de conservação da energia no sistema, pela projeção da aproximação linear do erro absoluto, temos que, por exemplo, para um tempo final de integração, $tf = 10.000$, o erro seria da ordem de 10^{-7} , ou seja, a energia se conserva até a sétima casa, o que é considerado satisfatório, uma vez que, para o sistema em questão, a localização foi observada em um tempo bem menor.



Conclusão

Além da considerável complexidade estrutural da molécula de DNA há uma variedade de parâmetros ambientais que podem influenciar a dinâmica desta molécula. Entretanto, considerando algumas simplificações concebemos modelos para a molécula de DNA. No caso específico do modelo tratado neste trabalho é importante citar que utilizamos um potencial de Morse para simular interações de ponte de hidrogênio, o que tem sido feito em outros trabalhos para simular estas interações em moléculas menores.

Assim, com este modelo bastante simplificado, que leva em conta a não-homogeneidade do código genético somente diferenciando o potencial de emparelhamento (ou pontes de hidrogênio) entre as bases A-T e C-G, obtivemos, para o caso específico da sequência do promotor T7A₁, sob determinadas condições, um resultado que indica a localização, tanto das maiores amplitudes de oscilação, quanto da energia, em uma região do promotor rica em nucleotídeos com bases adenina ou timina (o TATA-box). Coincidentemente esta região está localizada dentro do intervalo indicado como sendo aquele no qual a RNA-polimerase interage inicialmente com as fitas, ou seja, a região na qual é formada a bolha de transcrição. Os resultados obtidos para as sequências "aleatórias" I, II e III mostram que a localização ocorre, preferencialmente em regiões de sequência que

possuem nucleotídeos com adenina ou timina consecutivos. Desta forma, os resultados deste modelo simplificado, parece nos indicar que os pares de bases exercem um papel fundamental na sua dinâmica, pois pela maneira como estão associados além de contribuírem para estabilidade, atribuem à molécula propriedades que aparentemente são relevantes no processo dinâmico da transcrição. A sequência promotora apresenta amplitudes e energias significativamente maiores que das outras seqüências analisadas.

- [1] HAKILL, Y. M. D., ZHANG, B. H. *Effect of sequence and binding of DNA on transcription: analysis of the transcriptional regulation of DNA*. *J. Chem. Phys.*, v.76, p.3991-3992, 1982.
- [2] ENGLANDER, S. W., KACHURAVY, N. L., HENRIK, A. J., KRUMHOLTZ, J. A. and HOFER, A. *Study of the open state of DNA polymerase II by X-ray crystallography*. *Nature*, v.332, p.295-298, 1988.
- [3] SINGH, B. H. *DNA Structure and function*. Academic Press, San Diego, 1984.
- [4] KILBICK, W. *Properties of nucleic acid structure*. Springer-Verlag, New York, 1984.
- [5] YAKUSHEVICH, L. V. *Mathematical models of DNA*. John Wiley & Sons, New York, 1988.
- [6] VANDERHART, L. V. *Nucleic DNA structure*. John Wiley & Sons, New York, 1988.

Referências Bibliográficas

- [1] WATSON, J. D., CRICK, F. H. C. A structure for deoxyribose nucleic acid. *Nature*, v.171, p.737-38, 1953.
- [2] BARKLEY, M.D., ZIMM, B.H. Theory of twisting and bending of chain macromolecules: analysis of the fluorescence depolarization of DNA. *J. Chem. Phys.*, v.70, p.2991-3007, 1979.
- [3] ENGLANDER, S.W., KALLENBACH, N.R., HEEGER, A.J., KRUMHANS, J.A. and LITWIN, A. Nature of the open state in long polynucleotide double helices: possibility of solitons excitations. *Proc. Natl. Acad. Sci. USA*, v.77, p.7222-26, 1980.
- [4] SINDEN, R. R., *DNA Structure and function*. Academic Press, San Diego, 1994.
- [5] SAENGER, W., *Principles of nucleic acid structure*. Springer-Verlag, New York, 1984.
- [6] YAKUSHEVICH, L. V., *Nonlinear Physics of DNA*. John Wiley & Sons, New York, 1998.
- [7] YAKUSHEVICH, L. V. Nonlinear DNA dynamics: a new model. *Phys. Lett. A*, v.136, p.413-17, 1989.

- [8] PEYRARD, M., BISHOP, A.R. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, v.62, p.2755-58, 1989.
- [9] DRIGO, E., RUGGEIRO, J. R. H-bond simulation in DNA using a harmonic oscillator isospectral Potential. *Phys. Rev. E*, v.56, (3), p.4486, 1997.
- [10] DRIGO, E., RUGGEIRO, J. R. Parameters describing H-bond in DNA. *Phys. Rev. A*, v.44, (12), p.8435-36, 1991.
- [11] GAETA, G. On a model of DNA torsion dynamics. *Phys. Lett. A*, v.143, p.227-32, 1990.
- [12] DAUXOIS, T. Dynamics of breathers modes in a nonlinear "helicoidal" model of DNA. *Phys. Lett. A*, v.159, p.390-95, 1991.
- [13] ZDRAVKOVIC, S. and STARIC, M. V. The impact of viscosity on the DNA dynamics. *Physica Scripta*, v.64, p.612-19, 2001.
- [14] GAETA, G., REISS, C., PEYRARD, M., and DAUXOIS, T. Simple models of non-linear DNA dynamics. *Del nuovo cimento*, v.17, p.1-48, 1994.
- [15] SCOTT, A. C., CHU, F. Y., and MAC LAUGHLIN, D.W. The soliton: a new concept in applied science. *Proc.IEEE*, v.61, p.1443-83, 1973.
- [16] KUDRITSKAYA, Z. G., DANILOV, V. I. Quantum mechanics study of bases interactions in various associates in atomic dipole approximation. *Theor. Biol.*, v.59, p.303-18, 1976.
- [17] GOULD, H., TOBOCHNIK, J. *An introduction to computer simulation: Applications to physical systems*. Addison-Wesley Publishing, 1996.
- [18] BUTKOV, E. *Mathematical physics*. Addison-Wesley Publishing, 1968.



- [19] REIGADA, R. One-dimensional arrays of oscillators energy localization in thermal equilibrium. *Journal of chemical*, v.111, n.4 p.1373-84, 1999.
- [20] SALERNO, M. Discrete model for DNA-promoter dynamics. *Physical Review A*, v.44, n08 p.5292-97, 1991.
- [21] SALERNO, M. Dynamical properties of DNA. *Physics Letters A*, v.167, p.49-53, 1992.
- [22] ALBERTS, B., BRAY, D., LEWIS, J., RAFF, M. ROBERTS, K., WATSON, J. D. *Molecular biology of the cell*. 3.ed., Garland Publishing, New York, 1994.
- [23] LEWIN, B., *Genes VI*. Oxiford Univercity Press, New York, 1997.
- [24] DREW, B., *Understanding dna*. Oxiford Univercity Press, New York, 1997.
- [25] WOLFRAM, S. *The mathematica book*. 4.ed., Cambridge University Press, New York, 1999.
- [26] FERREIRA, S. E., DRIGO E., RUGGEIRO, J. R. Modelo dinâmico para a molécula do DNA. In Anais do II Congresso Temático de Aplicações de Dinâmica, e Controle (DINCON 2003), J. M. Balthazar, M. Boaventura, G. N. Silva, M. Tsuchida, Eds., 2003, São José dos Campos.



