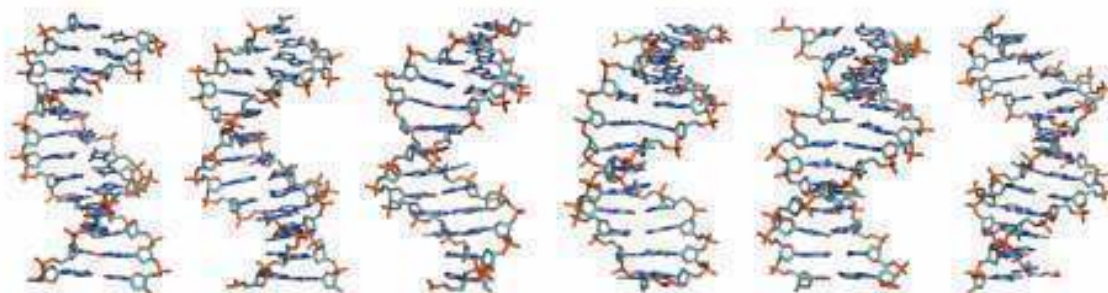


PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
FACULDADE DE BIOCÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

ANÁLISE POR SIMULAÇÕES PELA DINÂMICA MOLECULAR DO COMPLEXO  
PROTEÍNA-DNA DO DOMÍNIO ETS DA PROTEÍNA PDEF:  
FATORES INTRÍNSECOS AO DNA



**Ardala Breda**

ORIENTADOR: **Prof. Dr. Osmar Norberto de Souza**

DISSERTAÇÃO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO  
EM BIOLOGIA CELULAR E MOLECULAR - PUCRS COMO  
REQUISITO PARA A OBTENÇÃO DE TÍTULO DE MESTRE EM  
BIOLOGIA CELULAR E MOLECULAR.

Porto Alegre, Março de 2007

O presente trabalho foi realizado no período de Julho de 2005 a Fevereiro de 2007 no Laboratório de Bioinformática, Modelagem e Simulação de Biosistemas (LABIO/FACIN) da PUCRS, sob orientação do Professor Dr. Osmar Norberto de Souza.

Ardala Breda

“Careful observation is the only key to true and complete awareness”

– Karen deGroot

The D.H.A.R.M.A. Initiative

“Hey, ho, let’s go!”

– Ramones

## - Resumo -

A formação dos complexos proteína-DNA depende da complementaridade de cargas entre as cadeias laterais dos aminoácidos que compõem a proteína e os nucleotídeos que compõem a cadeia de DNA (reconhecimento direto) e do reconhecimento da estrutura, da conformação tridimensional, adotada ou capaz de ser adotada pela cadeia de DNA (reconhecimento indireto).

A proteína PDEF está relacionada com o controle da transcrição de genes responsáveis pelo desenvolvimento de tumores em tecidos sujeitos à regulação por hormônios sexuais; principalmente câncer de mama e de próstata. Assim como as demais proteínas da família ETS de fatores de transcrição, a PDEF reconhece e se liga à cadeia de DNA através do domínio ETS, que caracteriza esta família. As proteínas ETS apresentam maior dependência dos fatores de reconhecimento indiretos na formação dos complexos proteína-DNA. Ou seja, a conformação adotada pela cadeia de DNA tem maior influência sobre a afinidade de ligação da proteína do que a própria seqüência de nucleotídeos.

Ao contrário das demais proteínas ETS, que reconhecem com alta afinidade a seqüência 5'-GGAA-3', e apresentam baixa ou nenhuma afinidade pela seqüência 5'-GGAT-3', a proteína PDEF apresenta alta afinidade de ligação pela seqüência 5'-GGAT-3' e não forma complexos com cadeias de DNA que contenham o sítio de ligação 5'-GGAA-3'. A PDEF

apresenta ainda substituições únicas de aminoácidos na hélice de reconhecimento da cadeia de DNA.

Para avaliar os efeitos destas substituições de aminoácidos e das substituições de pares de base no sítio de ligação da proteína na conformação tridimensional da cadeia de DNA, e conseqüentemente, na afinidade de ligação entre proteína e DNA, foram realizadas análises das seqüências de aminoácidos das proteínas ETS e das interações observadas experimentalmente na interface proteína-DNA e simulações pela dinâmica molecular dos tridecâmeros de DNA que contêm as seqüências de alta e de baixa afinidade pela proteína PDEF (GGAT e GGAG, respectivamente) e do sistema controle sem afinidade de ligação, GGAA.

Os resultados das análises e das simulações pela dinâmica molecular indicam que as alterações na seqüência de aminoácidos da proteína, e a substituição dos grupamentos expostos no sulco maior da cadeia de DNA (padrão de reconhecimento direto, ou intermolecular), aliados a alterações na conformação da cadeia de DNA induzidas pelas substituições das bases na posição +4 do sítio de ligação (padrão de reconhecimento indireto ou intramolecular) parecem ser responsáveis pelo padrão de afinidade distinto da PDEF. No entanto, a não conservação dos resíduos que interagem com a cadeia de DNA, e o grande número de contatos não específicos com a cadeia fosfato-açúcar corroboram o padrão de leitura intramolecular como o principal mecanismo de reconhecimento da cadeia de DNA pelas proteínas ETS.

## - Abstract -

Protein-DNA complex assembling depends on amino acid side chains and nucleotides charge complementarity (direct readout mechanism) and on recognition by the protein of an already existing or induced three-dimensional DNA conformation (indirect readout mechanism).

The PDEF protein is a transcriptional factor of genes related with cancer development in tissues under sexual hormones regulation, mainly prostate and mammary epithelia. Similar to the other proteins of the ETS family, PDEF binds DNA sequences through its ETS domain, which characterizes this family of transcription factors.

Experimental studies indicate that ETS proteins are more dependable on indirect readout mechanism for complex formation, where the structure adopted by the DNA double helices has greater influence on the protein binding affinity than the nucleotide sequence.

While most ETS have high affinity for the nucleotide core sequence 5'-GGAA-3' and low affinity to 5'-GGAT-3', PDEF binds with high affinity the core sequence 5'-GGAT-3' and did not bind 5'-GGAA-3' sequences at all. PDEF has unique amino acid substitutions on the helix that binds DNA in the major groove.

In order to evaluate these amino acid substitutions consequences and the effect of base pair replacement on ETS binding site

on DNA's three-dimensional conformation, and consequently on protein-DNA affinity, we performed bioinformatics analyses on the ETS protein sequences and protein-DNA binding pattern analyses, as well as molecular dynamics simulation of 13 base-pairs DNA sequences with PDEF's high (GGAT), low (GGAG) and null (GGAA) affinity binding site.

Bioinformatics analyses and molecular dynamics simulations results indicate that the protein amino acids substitutions and the altered hydrogen bonds donor and acceptor exposed on the major groove of the DNA (direct readout mechanism), along with conformational changes induced by base pair replacement on protein binding site position +4 (indirect readout mechanism) seems to be responsible for PDEF distinct binding pattern. However, non-conserved residues that contact DNA and the greater number of unspecific protein side chain hydrogen bonds to the DNA's main chain phosphates point out to the intramolecular readout mechanism as the most important recognition mode for ETS proteins.

- Índice -

<b>Agradecimentos</b>	i
<b>Resumo</b>	iii
<b>Abstract</b>	v
<b>Lista de Figuras</b>	vii
<b>Lista de Tabelas</b>	x
<b>Lista de Abreviações</b>	xi
<b>Introdução</b>	01
1. Expressão Gênica	01
2. Controle de Expressão Gênica em Procariotos e Eucariotos	04
2.1 Fatores de Transcrição	08
3. Motivos Estruturais de Reconhecimento do DNA – Reconhecimento da Cadeia de DNA por Fatores de Transcrição	12
I. <i>Helix-turn-helix</i> (HTH)	12
II. <i>Zinc fingers</i>	13
III. Folhas $\beta$	13
IV. Zíper de Leucina	14
V. <i>Helix-loop-helix</i> (HLH)	15
4. A Família ETS de Fatores de Transcrição	16
5. PDEF ( <i>Prostate-derived ETS factor</i> )	22
5.1 PDEF e Oncogênese	27



---

6. A Estrutura do DNA	32
6.1 Parâmetros Conformacionais da Cadeia de DNA	38
<b>Hipótese de Trabalho</b>	45
<b>Objetivos</b>	49
<b>Materiais e Métodos</b>	50
1. Sistemas Biológicos Simulados Computacionalmente pelo Método da Dinâmica Molecular	50
1.1 Preparação dos Sistemas Moleculares	50
1.2 Controles das Simulações	53
Resumo dos Sistemas das Simulações pela DM	54
2. O método da DM	56
2.1 Protocolos de Simulações pela DM	57
2.2 Minimização de Energia	58
2.3 Aquecimento dos Sistemas	59
2.4 Fase de Produção	60
3. Análises das Simulações pela DM	61
4. Análises Bioinformática da Estrutura Primária do Domínio ETS das Proteínas ETS	64
<b>Resultados e Discussão</b>	66
1. Afinidade da Proteína PDEF pelo Sítio de Ligação da Cadeia de DNA	66
2. Simulações pela DM dos Tridecâmeros de DNA 1YO5, GGAT, GGAG e GGAA	83
2.1 Equilíbrio e Fase de Produção dos Sistemas	84
2.2 Comparações Entre os Sistemas - Influência da Substituição do Par de Bases +4 do Sítio de Ligação na Conformação dos Sistemas	91

3. Trajetória Dinâmica dos Sistemas 1YO5, GGAT, GGAG e GGAA	106
3.1 Cadeia Principal e Empacotamento do Açúcar	106
3.1.1 Ângulos Torcionais	106
Ângulos $\alpha$ , $\beta$ e $\gamma$	109
Ângulos $\epsilon$ e $\zeta$	113
Ângulos $\delta$ e $\chi$	115
3.1.2 Empacotamento do Açúcar	115
3.1.3 Distribuição dos Ângulos Torcionais e Empacotamento do Açúcar da Estrutura Experimental	118
3.2 Parâmetros Helicoidais	121
Deslocamento ao longo do eixo x (dx)	121
Deslocamento ao longo do eixo y (dy)	123
<i>Roll</i>	125
3.2.1 Curvatura Global	128
<b>Considerações Finais</b>	131
<b>Perspectivas</b>	137
<b>Referências Bibliográficas</b>	139
<b>Anexo I</b> – Arquivo de entrada para o módulo nucgen do pacote AMBER7	162
<b>Anexo II</b> – Arquivos de entrada para o módulo SANDER do pacote de programas AMBER7	165
<b>Anexo III</b> – Coordenadas conformacionais dos ângulos torcionais dos quatro tridecâmeros de DNA simulados	172
<b>Anexo IV</b> – Distribuições do ângulo de fase de pseudo-rotação para os quatro sistemas simulados	181
<b>Anexos V</b> – Parâmetros helicoidais dos quatro sistemas simulados envolvidos na determinação da estrutura intrínseca da cadeia de DNA	186

## Apêndices

- Artigo **“The mutation Y69Q on the PDEF protein effects its binding affinity and defines a new subfamily of ETS transcription factors”** - submetido à revista *Oncogene*, ISSN 09509232, classificação Qualis A, de circulação internacional, como um dos pré-requisitos para a obtenção do título de mestre pelo Programa de Pós-Graduação em Biologia Celular e Molecular da PUCRS.

## - Introdução -

### 1. Expressão Gênica

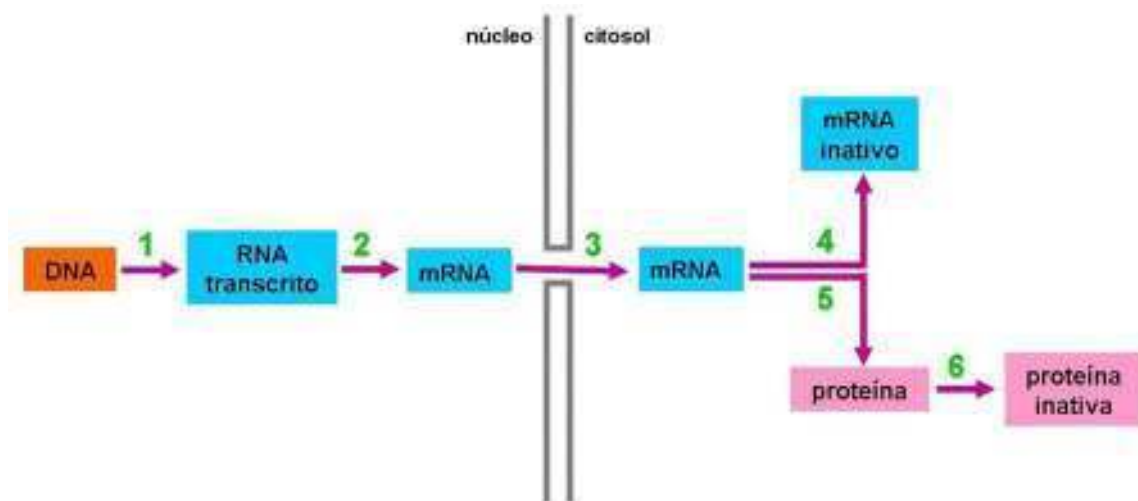
Os diversos tipos celulares que constituem um organismo diferem grandemente tanto em estrutura quanto em função, diferenças estas inicialmente atribuídas a deleções de genes no curso da diferenciação celular [1]. Avanços nas áreas da biologia molecular e genética permitiram o maior entendimento dos processos de diferenciação e especialização das células que compõem um organismo e a constatação da manutenção da integridade do genoma em cada uma destas células [1,2]; onde a diferenciação depende de modificações no padrão de expressão gênica e não de alterações na seqüência de nucleotídeos do genoma celular\*.

A expressão gênica pode ser controlada em diferentes pontos da via que engloba a transferência de informação do DNA até a proteína funcional [4]. Uma célula pode controlar as proteínas que expressa pela regulação quantitativa e qualitativa dos genes que transcreve (controle transcricional), por padrões diferenciados de processamento de moléculas de RNA, pela seleção de quais moléculas de RNA mensageiro (mRNA) serão exportadas ao citosol e qual será a sua localização, pela escolha de quais moléculas de mRNA presentes no citoplasma serão traduzidas em aminoácidos nos ribossomos, pela regulação da degradação das moléculas

---

\* Exceto na formação da região variável da cadeia pesada de anticorpos, onde ocorrem eventos de rearranjo gênico [3].

de mRNA, e por fim, pela ativação, inativação, degradação e compartimentalização de proteínas (Figura 1).



**Figura 1.** A expressão gênica de eucariotos pode estar sujeita ao controle transcricional (1), controle do processamento do RNA (2), do transporte e localização do mRNA no citosol (3), da degradação das moléculas de mRNA (4), da tradução (5) e da atividade protéica (6). A etapa 6 pode ter caráter reversível, através de ciclos de fosforilação e defosforilação e alosteria (adaptado de Alberts, *et al.*, 4º Ed., 2001).

A regulação pode ocorrer simultaneamente em mais de um dos pontos de controle para um mesmo gene; apesar da manutenção do genoma íntegro, apenas uma pequena fração do total de genes é expressa nas células, estando os demais genes inativos. A capacidade de regulação da expressão gênica confere vantagens adaptativas aos organismos, como por exemplo, a capacidade de modular o padrão de expressão de proteínas em resposta a estímulos externos, como as variações hormonais. Ou seja, esta capacidade de regulação permite que organismos procariotos e eucariotos se adaptem às variações do ambiente. Organismos eucarióticos

não apenas respondem às variações ambientais, como também alteram seu padrão de expressão gênica no decorrer do seu desenvolvimento. Em humanos, por exemplo, a passagem da infância para a adolescência e a fase adulta é acompanhada por modificações físicas e fisiológicas resultantes da variação da expressão gênica e, por conseqüência, das proteínas expressas. Mesmo após o organismo ter atingido a fase adulta, a regulação da expressão gênica permite a diferenciação de células para novas funções [5].

Como os processos de transcrição do RNA e de síntese de proteínas são dependentes de energia, as células poupam reservas energéticas sintetizando proteínas apenas quando necessárias; desta forma, a principal forma de controle da expressão gênica é o controle transcricional, a única forma de controle que evita de maneira efetiva a síntese supérflua de intermediários [4].

## 2. Controle de Expressão Gênica em Procariotos e Eucariotos

O controle da expressão gênica observada em organismos procariotos é significativamente mais simples que a de organismos eucariotos, uma vez que nestes organismos o empacotamento do seu material genético na forma de cromatina cria novas categorias de controle não observadas nos procariotos [6].

Procariotos são organismos unicelulares cujo genoma se constitui de um cromossomo circular, seu material genético não se encontra complexado a histonas, não é separado do conteúdo citoplasmático por envelope nuclear e seus genes não possuem íntrons [4].

A transcrição e tradução são processos simultâneos nestes organismos [4,5]. A meia vida das moléculas de mRNA é bastante curta, sendo degradadas em poucos minutos, assim, novas moléculas precisam ser constantemente transcritas para manter a síntese de proteínas. Por estas características, a regulação da transcrição - ou controle transcricional, é suficiente para a regulação dos níveis de proteínas da célula [5].

Os genes codificadores de proteínas, os genes estruturais, quando de funções correlatas, encontram-se agrupados seqüencialmente em unidades chamadas *operons*. Os genes de um *operon* são expressos de forma coordenada, em uma única cadeia de mRNA (mRNA policistrônico).

Em procariotos, o controle transcricional se dá através de **repressores**, proteínas regulatórias que evitam a ligação da RNA polimerase

ao promotor, interferindo na iniciação da transcrição (**controle negativo**); e por **ativadores**, pela estimulação da transcrição (**controle positivo**) [4,5,7].

Organismos eucariotos têm seu DNA organizado em nucleossomos, formando a cromatina; seus genes precisam estar na forma ativa para serem expressos. A estrutura de *operon* como em procariotos não é observada nestes organismos; genes que codificam proteínas de função correlata ou de uma mesma via estão geralmente localizados em cromossomos diferentes. Assim, cada gene possui seu próprio promotor e os processos de transcrição e de tradução são separados espacialmente pela compartimentalização celular, bem como temporalmente, uma vez que o mRNA só é traduzido depois de transladado para o citosol [8] (**Figura 1**).

A expressão gênica de eucariotos é regulada em múltiplos níveis (**Figura 1**). Uma vez que o gene esteja ativado pela remodelagem da cromatina, o principal mecanismo de regulação da expressão se dá na iniciação da transcrição na região promotora dos genes [9], este controle pode se dar pela regulação da formação do complexo de transcrição basal e pela sua ligação à região TATA *box* promotora do gene [5,10].

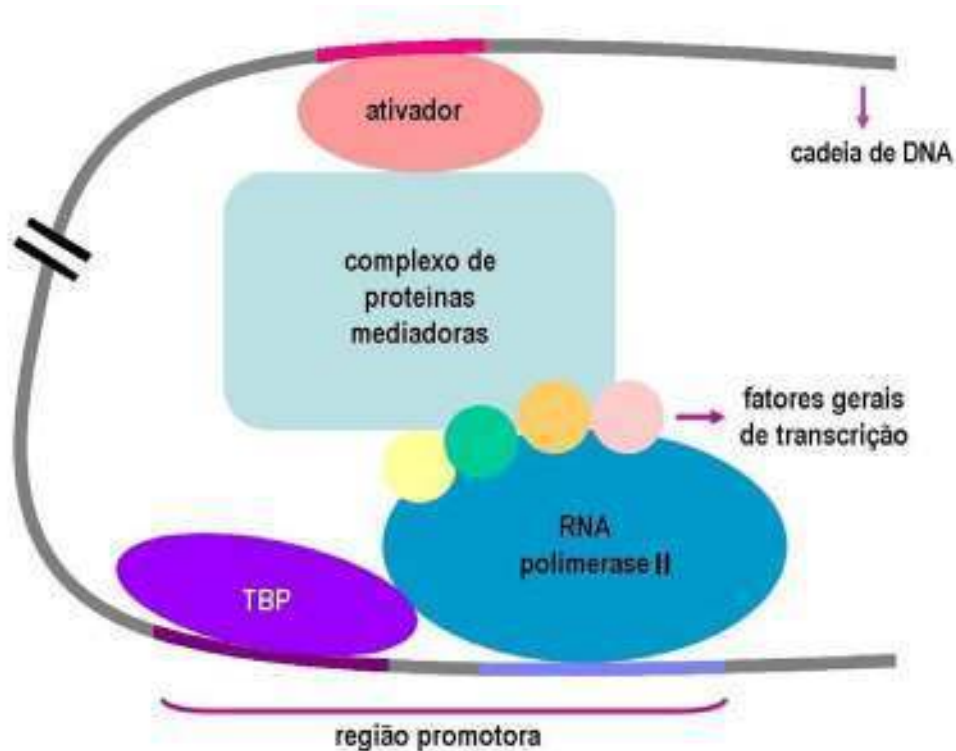
O complexo de transcrição basal é formado pela RNA polimerase II, a proteína de ligação à região TATA *box* - TBP (TATA *Box binding protein*) e outras proteínas chamadas fatores gerais de transcrição [6,9]. Fatores de transcrição adicionais são necessários para todos os genes e ligam-se na porção *upstream* adjacente à região promotora dos



genes que regulam; sua função é aumentar a frequência de transcrição gênica e são necessários para o funcionamento adequado dos promotores [11,12,13].

A região controladora dos genes contém ainda seqüências de DNA ditas regulatórias, específicas para cada gene e capazes de aumentar sua transcrição em até 1000 vezes ou mais, razão pela qual são denominadas *enhancers*, podendo estar localizadas a centenas de nucleotídeos de distância da região promotora do gene. **Fatores de transcrição** específicos para cada gene ligam-se a estas seqüências regulatórias e interagem com proteínas mediadoras chamadas co-ativadoras que, por sua vez, induzem a formação de um *loop* na cadeia de DNA, possibilitando a sua interação com o complexo de transcrição basal, induzindo alterações conformacionais no complexo para a sua forma ativa, no sítio de iniciação da região promotora do gene que regulam [6,8,9,14] (**Figura 2**).

As principais diferenças no controle da expressão gênica de eucariotos e procariotos podem ser resumidas em três pontos principais, em razão da maior complexidade da organização do material genético dos organismos eucariotos. Primeiro, o empacotamento do DNA de eucariotos em cromatina permite novos níveis de regulação não disponíveis em procariotos.



**Figura 2.** Representação esquemática do controle transcricional mediado por fatores de transcrição (ativador), que atuam sobre o complexo de transcrição basal (RNA polimerase II, fatores gerais de transcrição e TBP) através das proteínas mediadoras. Proteínas regulatórias ou fatores de transcrição ligam-se diretamente à cadeia de DNA, nas seqüências regulatórias da cadeia, podendo ser tanto ativadoras quanto repressoras da transcrição gênica. O ativador (em rosa) liga-se a proteínas mediadoras (co-ativadoras ou co-repressoras - em verde claro) que por sua vez atuam sobre o complexo de transcrição basal, induzindo sua alteração para a conformação ativa capaz de reconhecer a região promotora do gene (a proteína TBP liga-se à porção TATA *box* do promotor) iniciando o processo de transcrição. A linha interrompida indica que a porção da cadeia de DNA em *loop* pode ser bastante longa e de comprimento variável (adaptado de Smith *et al.*, 2ª Ed., 2004).

Segundo, as proteínas regulatórias de eucariotos podem atuar sobre a região promotora de genes localizados a grandes distâncias no genoma, e um mesmo promotor pode sofrer influência simultânea de múltiplos fatores de transcrição.

Terceiro, a RNA polimerase II não é capaz de iniciar o processo de transcrição sozinha, sua atividade é dependente dos fatores gerais de transcrição e da TBP, que devem estar arranados corretamente no complexo de transcrição basal, o que aumenta as possibilidades de controle e regulação do processo de transcrição em organismos eucariotos, onde os fatores de transcrição são hábeis a atuar em diferentes etapas do processo de iniciação da transcrição gênica.

## 2.1 Fatores de Transcrição

As proteínas regulatórias que se ligam diretamente a seqüências de DNA são chamadas de **fatores de transcrição** e não devem ser confundidas com os fatores gerais de transcrição do complexo de transcrição basal. Estas proteínas são também chamadas de ativadoras, transativadoras, indutoras, repressoras ou receptores nucleares. Estima-se que de 5 a 10% dos 30000 genes humanos codifiquem fatores de transcrição [5].

Além do sítio de ligação ao DNA, as proteínas regulatórias geralmente possuem sítios de ligação a proteínas mediadoras (co-ativadoras, co-repressoras ou TAFS - *TATA binding protein associated factor*) [6,12]. As proteínas mediadoras não se ligam diretamente à cadeia de DNA, interagindo com componentes do complexo de transcrição basal,

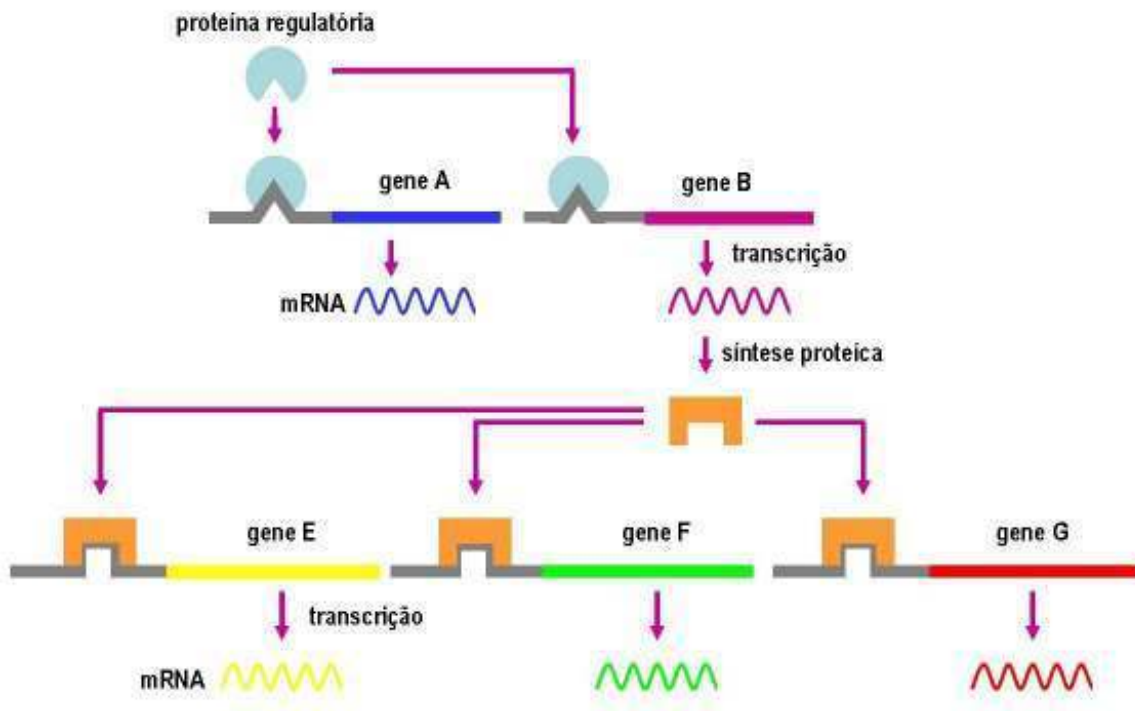
mediando sua ligação à região promotora; podem ser específicas ou ligar-se a diferentes fatores de transcrição [5,6,8,9].

Quando estão formadas as ligações apropriadas entre o fator de transcrição, o co-ativador e o complexo de transcrição basal, a taxa de transcrição do gene regulado é aumentada\reduzida. Fatores de transcrição inibitórios, ou repressores podem atuar inibindo a ligação de ativadores à seqüência regulatória, ou ainda pela ligação de um co-repressor que inibe a formação do complexo de transcrição basal. Alguns repressores ligam-se diretamente ao complexo de transcrição basal, inibindo-o, outros atuam através do recrutamento de complexos de remodelagem da cromatina em heterocromatina ou ainda, no recrutamento de histonas deacetilases [4,5,8,9,12].

Como precisam interagir com as proteínas mediadoras para estimular a transcrição gênica, os fatores de transcrição têm sua atividade regulada pela disponibilidade destas proteínas. O aumento ou diminuição da síntese dos co-ativadores, co-repressores ou TAFS influem sobre a taxa de transcrição. Também a quantidade de fatores de transcrição sintetizada e\ou disponível afeta o processo de transcrição [5,6], assim como a ligação de ligantes estimulatórios ou inibitórios e a atividade antagonista de diferentes fatores de transcrição sobre um mesmo gene [4,6,9].

Um mesmo fator de transcrição pode induzir e ativar a transcrição de genes diferentes caso estes genes apresentem a mesma seqüência regulatória [9], num efeito conhecido como cascata

transcricional (**Figura 3**). Um exemplo de cascata de transcrição é o processo de diferenciação de adipócitos, onde um estímulo inicial (insulina) induz a síntese de fatores de transcrição específicos que atuam em cascata para a expressão de genes exclusivos de células adiposas [5,15].



**Figura 3.** Ativação de um conjunto de genes a partir de um único ativador. Os genes A e B e os genes E, F, e G possuem seqüências regulatórias comuns entre si. O produto do gene B é um fator de transcrição que estimulará a síntese do segundo conjunto de genes (adaptado de Smith *et al.*, 2<sup>a</sup> Ed., 2004).

Além da sua atividade sobre a formação do complexo de transcrição basal e sua ligação à cadeia de DNA, os fatores de transcrição podem promover a transcrição gênica pelo rearranjo da cromatina nas seqüências regulatórias ou regiões promotoras do gene através do recrutamento de histonas acetilases e complexos de remodelagem de

cromatina ATP-dependentes [4,8,11], garantindo maior acessibilidade à cadeia de DNA, facilitando o reconhecimento e ligação de fatores de transcrição adicionais e do próprio complexo de transcrição basal.

A localização das seqüências regulatórias distante da região promotora do gene abre a possibilidade de um mesmo gene estar sujeito à ação de múltiplos fatores de transcrição, sendo o controle transcricional o produto final da ação conjunta em sinergia destes fatores [6,9].

Cada gene apresenta características peculiares na ativação da sua transcrição, muito embora, para dar início ao evento de transcrição, um fator de transcrição específico (uma proteína regulatória) deve inicialmente reconhecer e liga-se à seqüência regulatória da cadeia de DNA para desencadear todo o processo subsequente de transcrição gênica [6,8,12].

### 3. Motivos Estruturais de Reconhecimento do DNA – Reconhecimento da Cadeia de DNA por Fatores de Transcrição

As proteínas regulatórias reconhecem e se ligam a seqüências de DNA através da complementaridade ao seu sítio de ligação, através de pontes de hidrogênio, ligações iônicas e contatos hidrofóbicos [4]. Apesar de cada contato ser fraco individualmente, o conjunto deles torna a ligação proteína–DNA de alta afinidade e especificidade.

Apesar de cada interação proteína–DNA ser específica, os métodos experimentais de determinação de estruturas por cristalografia por difração de raios X e ressonância magnética nuclear, permitiram a identificação e classificação dos padrões de interações e reconhecimento da cadeia de DNA pelos fatores de transcrição em cinco grupos principais [4,16,17]. Cada um destes grupos comporta um conjunto de proteínas cujas estruturas mantêm o motivo de reconhecimento e ligação à cadeia de DNA embora possam apresentar variações na organização e composição da sua estrutura terciária.

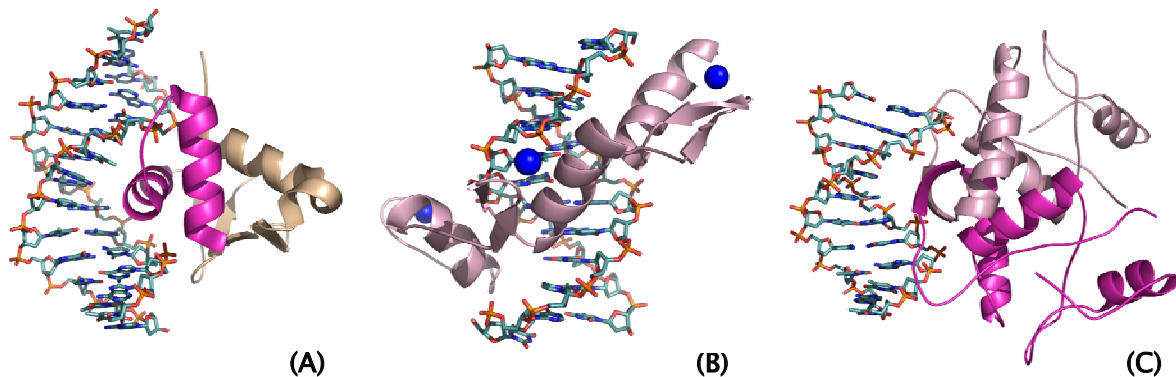
**1. Helix–turn–helix (HTH):** Foi o primeiro motivo estrutural de reconhecimento de seqüências de DNA a ser descrito, identificado inicialmente em bactérias, é encontrado em procariotos e eucariotos. É formado por duas hélices  $\alpha$  conectadas por uma volta (*turn*); a hélice da porção carbóxi terminal (C–terminal) é denominada hélice de reconhecimento, por ser a responsável pela interação direta com o DNA, no

sulco maior da cadeia (**Figura 4A**). A variação dos aminoácidos que compõem a hélice de reconhecimento confere especificidade ao HTH, motivo estrutural que é observado em diversos fatores de transcrição, incluindo as proteínas de homeodomínio [16,18].

**II. Zinc Fingers:** As diferentes estruturas agrupadas sob o nome de “*zinc fingers*” têm em comum íons zinco como componente estrutural. Desde a sua descoberta em *Xenopus*, diversos motivos de ligação contendo íons zinco foram descobertos e identificados como “*zinc fingers*”, sendo suas estruturas tão diversas quanto suas funções [19,20]. A nomenclatura do grupo é derivada da representação esquemática reproduzida na primeira descrição deste motivo de reconhecimento. Este motivo estrutural é formado por uma hélice  $\alpha$  e uma folha  $\beta$ , unidos por um íon zinco, geralmente em repetições ao longo da proteína, envolvendo a cadeia de DNA, ligando-se aos seus sulcos maiores através de suas hélices  $\alpha$  (**Figura 4B**).

**III. Folhas  $\beta$ :** Apesar de as hélices  $\alpha$  serem o motivo de reconhecimento de seqüências de DNA mais comum, estruturas de reconhecimento formadas por folhas  $\beta$  também são conhecidos (**Figura 4C**). Assim como nos grupos descritos acima, os aminoácidos que formam as folhas  $\beta$  e que interagem com a cadeia de DNA são responsáveis pela especificidade do reconhecimento e ligação às seqüências regulatórias [4,5,16].



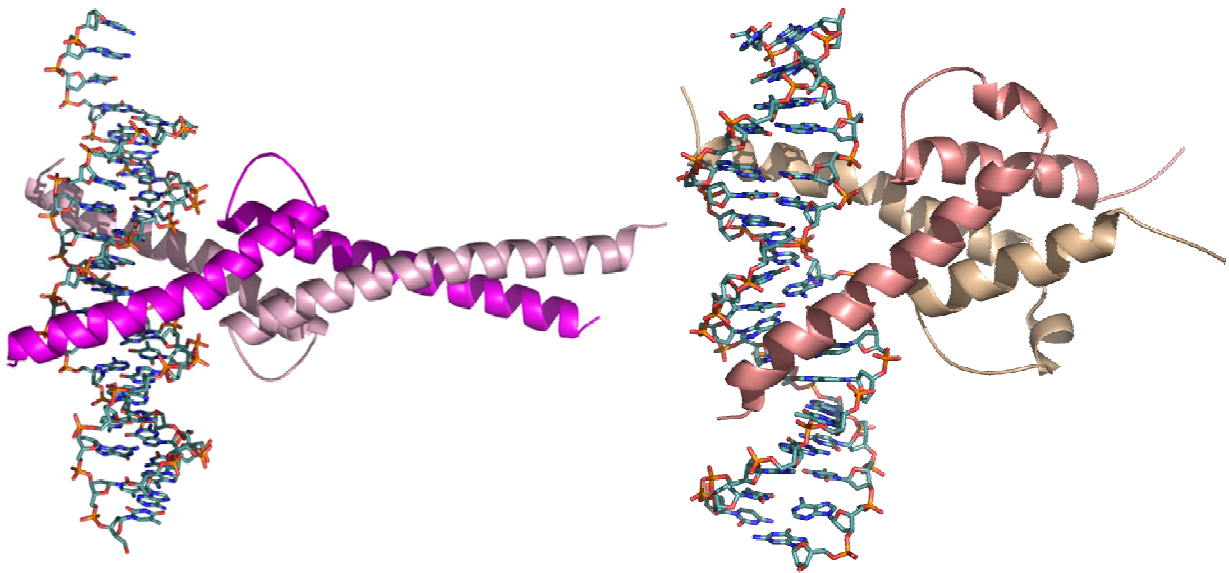


**Figura 4.** Exemplo de motivos estruturais de reconhecimento da cadeia de DNA. **(A)** Motivo HTH. A proteína mostrada é o fator de transcrição PDEF (*prostate-derived ETS factor*, código PDB: 1YO5), com sua seqüência de DNA complementar, o domínio HTH da proteína está destacado em magenta, a hélice de reconhecimento (horizontal) reconhece o padrão de pares de bases expostos no sulco maior da cadeia de DNA. **(B)** *Zinc finger* (código PDB: 1A1G). Cada uma das três repetições do motivo de reconhecimento é composta por uma hélice  $\alpha$ , uma folha  $\beta$  e um íon zinco (esferas azuis). Cada uma das hélices reconhece seqüências específicas da cadeia de DNA nos seus sulcos maiores, enlaçando a dupla hélice. **(C)** Estrutura do repressor met, responsável pelo controle da transcrição das enzimas relacionadas com a síntese de metionina de *E. coli* (código PDB: 1CMA). A folha  $\beta$  de reconhecimento da seqüência regulatória é formada por uma fita  $\beta$  de cada uma das subunidades da proteína (coloridas em rosa e magenta). Em (A), (B) e (C), a topologia da proteína está representada no modelo de fitas, suas cadeias laterais não estão representadas para maior clareza da figura. Cadeias de DNA representada no modelo de palitos.

**IV. Zíper de Leucina:** Motivos em hélice  $\alpha$  que apresentam resíduos de leucina a cada sete aminoácidos, voltados todos para a mesma face da hélice. Duas hélices se associam formando dímeros através de interações hidrofóbicas entre seus resíduos de leucina alinhados (**Figura 5A**). Ao contrário dos demais motivos de reconhecimento, no caso dos zíperes de leucina as subunidades que se associam não precisam ser idênticas, podendo haver tanto estruturas de homodímeros quanto de heterodímeros. A formação de heterodímeros confere mais um grau de

especificidade de ligação possível para os fatores de transcrição, já que duas seqüências regulatórias distintas podem ser reconhecidas simultaneamente [4,16].

**V. *Helix-loop-helix* (HLH):** Estruturalmente similares aos zíperes de leucina, também se associam em homo e heterodímeros, o que influi na sua especificidade de ligação (**Figura 5B**). Não devem ser confundidos com o motivo HTH. O motivo HLH consiste em duas hélices unidas por uma alça (*loop*), cuja flexibilidade permite a formação do dímero e sua ligação à cadeia de DNA. Mais de 240 fatores de transcrição HLH já foram identificados em organismos desde *Saccharomyces cerevisiae* a *Homo sapiens* [21,22,23].



**Figura 5.** (A) Zíper de leucina. Fator de transcrição SREBP (*sterol regulatory element binding protein*, código PDB: 1AM9). (B) Homodímero HLH da proteína ativadora do sistema fosfato de *S. cerevisiae* (código PDB: 1A0A). Em (A) e (B), As regiões de reconhecimento de cada hélice ligam-se a sulcos maiores adjacentes na cadeia de DNA. Subunidade do homodímero coloridas em (A) rosa e magenta e (B) salmão e amarelo. Proteínas e cadeias de DNA representadas como na Figura 4.

#### 4. A Família ETS de Fatores de Transcrição

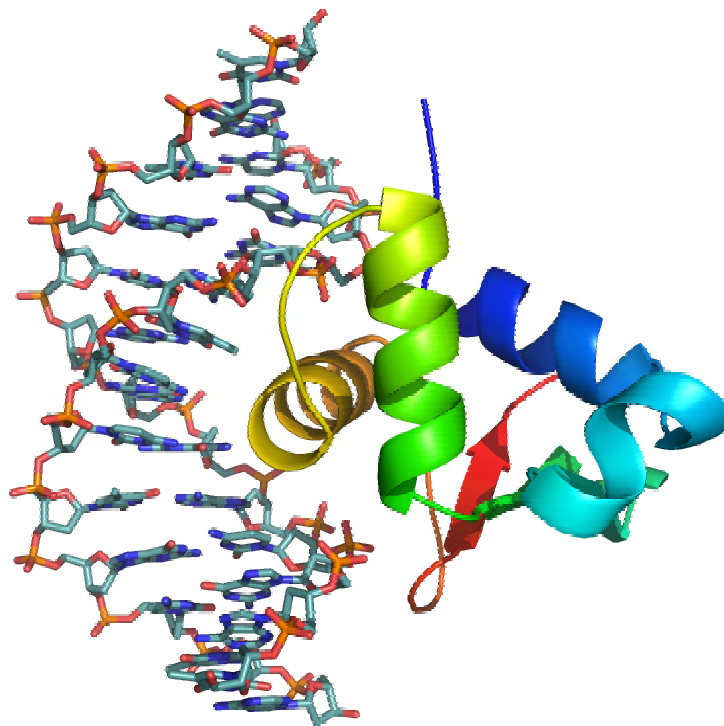
Em 1983, foi identificado o vírus da eritroblastose aviária E26, cujo genoma é composto por seqüências de genes característicos de retrovírus e por dois oncogenes celulares, *myb* e um segundo gene até então desconhecido, designado *ets* (*e26 transformation-specific sequence*) [24]. A identificação da proteína vEts-1 levou à descoberta de proteínas homólogas em múltiplas espécies, agrupadas na família ETS de fatores de transcrição. As proteínas que fazem parte desta família foram identificadas e agrupadas com base na homologia com o domínio de ligação ao DNA da proteína Ets-1, chamado domínio ETS [24,25].

O domínio ETS é um subgrupo do motivo de reconhecimento da cadeia de DNA HTH, denominado *winged helix-turn-helix*, wHTH, constituído de 85 aminoácidos, composto por duas hélices  $\alpha$  conectadas por uma volta e por duas folhas  $\beta$  conectadas por uma alça (esta alça se destaca da estrutura principal como uma asa - *wing*, o que originou a nomenclatura deste motivo estrutural - **Figura 6**). Estudos estruturais mostraram uma grande conservação no arranjo deste domínio, com as três hélices  $\alpha$  e quatro fitas  $\beta$  arranjadas na seguinte ordem:  $\alpha 1-\beta 1-\beta 2-\alpha 2-\alpha 3-\beta 3-\beta 4$  [26]. O principal ponto de interação com a cadeia de DNA é a volta formada entre as hélices 2 e 3, a hélice número 3 ( $\alpha 3$ ), e a alça entre as fitas  $\beta 3$  e  $\beta 4$  [25].

As proteínas ETS reconhecem e se ligam à seqüência de nucleotídeos 5'-GGA(A/T)-3'. Os quatro nucleotídeos localizados nas

posições 5' e 3' desta seqüência central também são responsáveis pela especificidade de ligação de cada um dos membros da família ETS [25], bem como as proteínas mediadoras ou co-ativadoras [6].

Apenas os organismos metazoários (a linhagem eucariótica, exceto fungos, leveduras e plantas) apresentam fatores de transcrição da família ETS, cuja função parece estar relacionada a cascatas regulatórias exclusivas a estes organismos [25,27].



**Figura 6.** A mesma estrutura mostrada na Figura 4A (código PDB: 1YO5), PDEF colorida por sucessão de estrutura secundária, da extremidade N-terminal (azul) para C-terminal (vermelho). Além da hélice 3 (hélice horizontal), característica do motivo HTH, também a volta que conecta as hélice 2 e 3 e a alça entre as fitas  $\beta$ 3 e  $\beta$ 4 fazem contato com a cadeia de DNA. Cadeia de DNA representada no modelo de palitos, átomos de fósforo em laranja, carbono em ciano, oxigênio em vermelho e nitrogênio em azul. Estrutura da proteína na representação de fitas. Os átomos de hidrogênio da cadeia de DNA e as cadeias laterais da proteína foram suprimidos para maior clareza da figura.

Os fatores de transcrição ETS atuam como receptores nucleares de cascatas de transdução de sinais, geralmente funcionando em colaboração com outros fatores de transcrição, ligados ou não à cadeia de DNA, com co-reguladores, histona acetilases e proteínas fosfatases [26,28,29,30,31]. Os mecanismos de sinalização variam significativamente entre os membros da família, havendo tanto ativadores quanto repressores do processo de transcrição [29,32], com funções tão diversas quanto o desenvolvimento hematopoiético e neuronal em mamíferos, ocular em *Drosophila* e diferenciação sexual em *Caenorhabditis elegans* [25].

O padrão de distribuição temporal e espacial destas proteínas e de seus co-reguladores em embriões e em organismos adultos contribui para a especificidade de seus efeitos no controle transcricional [33].

As proteínas desta família são classificadas em subfamílias de acordo com a similaridade de suas seqüências e outros domínios conservados em suas estruturas terciárias, o que se reflete nas suas funções biológicas e moleculares (**Tabela 1**).

Um dos domínios que também é altamente conservado entre os membros da família ETS é o domínio *pointed* (PNT) descrito inicialmente em *Drosophila melanogaster*, composto por um domínio HLH que atua como sítio de interações proteína-proteína, como local de oligomerização na formação de homo e heterodímeros, como ponto de ligação de inibidores da função das proteínas ETS e sítio de fosforilação pela proteína ERK2, na cascata de ativação das MAP cinases (MAPK) [25,26,28,30,31].

**Tabela 1.** Proteínas que integram a família ETS, identificadas na Classe *Mammalia*.

Proteína de domínio ETS	Domínio PNT	Ativador ou Repressor	Função biológica
<i>Subfamília Ets</i>			
Ets-1	Sim	A	Diferenciação de células T NK*
Ets-2	Sim	A	Desenvolvimento de tecidos extra-embrionários, remodelagem da matriz extracelular
Pointed	Sim	A	Múltiplos papéis no desenvolvimento
<i>Subfamília TCF</i>			
Elk-1	Não	A/R	Diferenciação neuronal
SAP-1	Não	A	Diferenciação de células T
SAP-2/Net	Não	R/A	Formação de vasos linfáticos
Lin-1	Não	R (?**)	Diferenciação sexual em <i>C. elegans</i>
<i>Subfamília Erg</i>			
Fli-1	Sim	A	Diferenciação de megacariócitos, angiogênese
Erg	Sim	A	
<i>Subfamília PEA3</i>			
PEA3	Não	A	Diferenciação neuronal
ERM	Não	A	Diferenciação neuronal (?) e muscular
ER81	Não	A	Formação de conexões sensoriais
<i>Subfamília GABP</i>			
GABP $\alpha$	Sim	A	Atuação na sinapse muscular
<i>Subfamília Elf</i>			
Elf-1	Não	A	?
NERF-1/-2	Não	A/R	?
MEF	Não	A	?
<i>Subfamília Spi</i>			
PU.1	Não	A	Diferenciação linfóide e mielóide
SpiB	Não	A	Observado em células B
SpiC	Não	A	?
<i>Subfamília Yan</i>			
TEL	Sim	R	Angiogênese no saco vitelínico, hematopoiese em adultos
TEL2	Sim	R	
Yan	Sim	R	Múltiplos papéis no desenvolvimento
<i>Subfamília Erf</i>			
ERF	Não	R	?
PE-1	Não	?	?
<i>Outras ETS</i>			
ESE-1/-2/-3	Sim	A/R	Expressão epitélio específica
PDEF	Sim	A/R	
ER71	Não	?	

\* Células *natural killer*.

\*\* Função ainda não identificada ou confirmada experimentalmente.

Dados originais e referências para cada subfamília em [25], números de acesso no *GenBank* em [34].

As proteínas ETS estão relacionadas a processos biológicos distintos, sem haver ainda uma função que abranja todos os membros da família. Funções tão diversas como crescimento, indução e inibição de apoptose, desenvolvimento, diferenciação, regulação da senescência, hematopoiese, angiogênese, transformação oncogênica entre outras são decorrência da divergência das seqüências destas proteínas nos seus demais domínios estruturais, o que contribui para diferentes padrões de expressão, propriedades de regulação da transcrição e por fim, diferentes papéis biológicos [26], assim como sua interação com outros fatores de transcrição não relacionados, o que origina o chamado controle combinatório, propriedade característica dos membros da família ETS [33].

Algumas proteínas ETS são tecido-específicas, e estão envolvidas no crescimento e diferenciação das células do tecido em que são expressas através da ativação da transcrição de genes codificantes de fatores de crescimento característicos de cada tecido [26].

A sinalização mediada por fatores de crescimento e seus receptores modulam a atividade de diversas proteínas ETS pela via MAPK através de fosforilações que podem afetar a sua capacidade de ligação à cadeia de DNA, e a sua capacidade de associação com outras proteínas regulatórias ou co-regulatórias, bem como a sua localização celular [26,28,31].

Subfamílias de proteínas ETS estão envolvidas na regulação da expressão de genes relacionados com o controle do ciclo celular,

oncogenes e genes supressores de tumor [26]. A desregulação ou expressão inapropriada destas proteínas parecem ter papel crucial no desencadeamento de alguns tipos de câncer [25,34,35], estando correlacionadas com altas taxas de invasividade e metástase [26,36,37].

Estudos com ratos *knockout* apontam uma nova etapa no estudo destes fatores de transcrição, onde conhecer e caracterizar cada um dos membros da família ETS, seus mecanismos moleculares e funções biológicas têm tido maior relevância que a procura pelo fator comum entre seus membros [36]. Uma vez que estas proteínas estão envolvidas em processos de crescimento celular, desenvolvimento, diferenciação e apoptose, o entendimento preciso do seu funcionamento pode auxiliar no controle de várias doenças humanas incluindo desordens imunológicas, doenças vasculares e o desenvolvimento de tumores malignos.

Uma vez conhecidos os mecanismos de ação dos membros da família ETS, a etapa seguinte deve ser a integração deste conhecimento na análise do impacto funcional de múltiplas proteínas ETS expressas em um mesmo tipo celular e sua capacidade de regular a expressão de genes em comum, com diferentes magnitudes e propriedades (ativador\repressor) e ainda padrões de expressões diferentes (constitutiva\transiente), assim como a relação com outros fatores de transcrição e seu sinergismo no controle transcricional [36].



## 5. PDEF (*Prostate-derived ETS factor*)

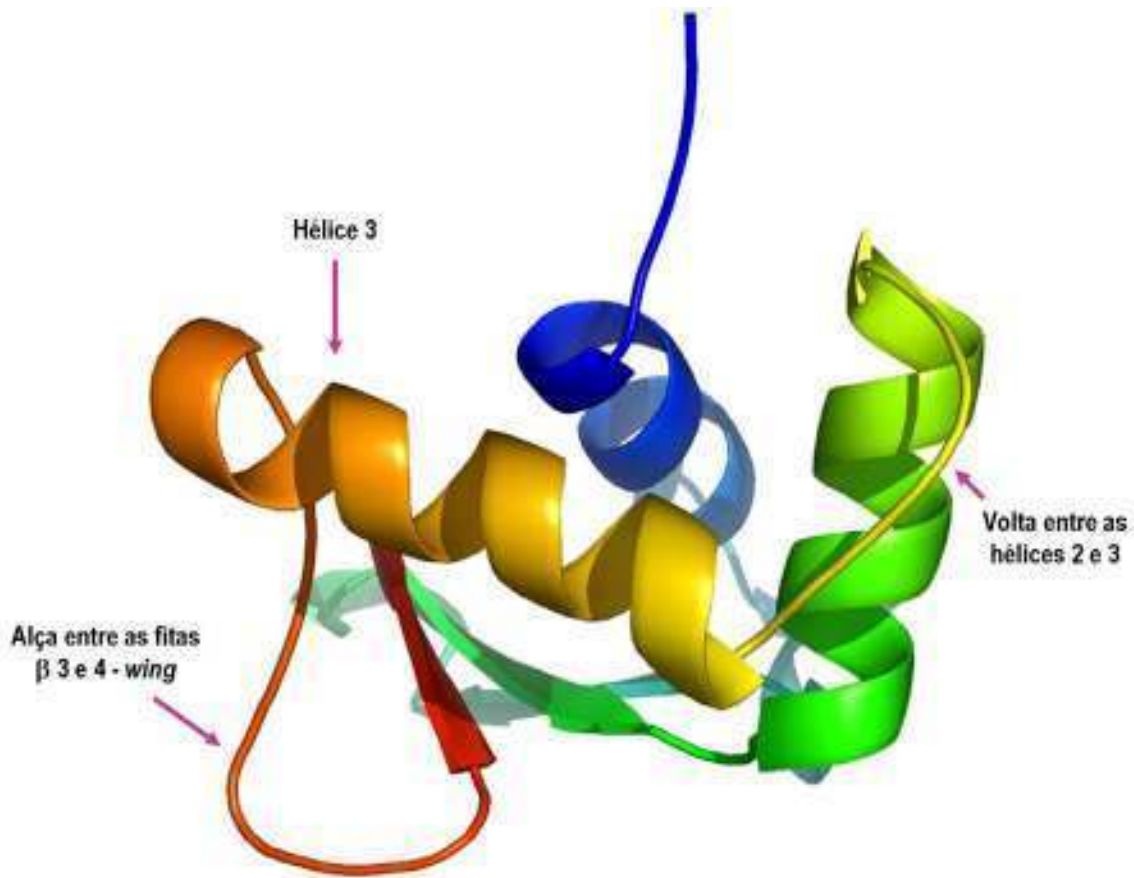
No ano de 2000 foi identificado e caracterizado um novo fator de transcrição específico de células epiteliais da próstata, denominado PDEF (*prostate-derived ETS factor* - número de acesso do *GenBank*: NM\_012391), mais tarde também observado, porém em menores concentrações, no epitélio dos ovários, glândulas mamárias e salivares; todos tecidos sujeitos à regulação por androgênios ou esteróides [38].

A proteína PDEF (**Figura 7**), composta por 335 aminoácidos, integra a família ETS de fatores de transcrição, constituindo uma nova subfamília da qual era até então único membro para a Classe *Mammalia* (**Tabela 1**) [39]. A proteína homóloga da PDEF, mPSE (*mouse prostate specific ets* - PSE corresponde à nomenclatura inicialmente atribuída à proteína PDEF) é observada no tecido epitelial da próstata em camundongos, com 98.2% de identidade entre as seqüências de aminoácidos (número de acesso no *GenBank*: AB019436) [40].

A proteína D-Ets4 de *Drosophila* (número de acesso no *GenBank*: AF071538 - proteína expressa nas células polares, relacionada com a diferenciação das células germinais) também apresenta similaridade significativa com PDEF, com 42% de identidade entre as seqüências de aminoácidos [41].

O alinhamento do domínio ETS na porção C-terminal da PDEF e de outras proteínas ETS indica alta identidade, 75%, 76% e 98.8% respectivamente, apenas com as proteínas D-Ets4 de *Drosophila*, SpEts4 de

*Strongylocentrotus purpuratus* (ouriço do mar púrpura), proteína envolvida no estabelecimento do eixo corporal (número de acesso no *GenBank*: NP\_999699) [38], e com a proteína mPSE [40], o que aponta que estas proteínas podem ser agrupadas como uma nova subfamília dos fatores ETS de transcrição [40].



**Figura 7.** Estrutura do domínio ETS da proteína PDEF (código PDB: 1YO5), composta por 85 aminoácidos. A estrutura secundária da proteína, representada no modelo de fitas, está colorida do azul para o vermelho, da extremidade amino terminal (N-terminal) para C-terminal. Os pontos de contato com a cadeia de DNA, característicos do motivo de reconhecimento wHTH, estão destacados na figura.

A proteína PDEF apresenta o domínio PNT característico da família ETS [25,26]; entretanto, ao contrário das demais proteínas da família onde o domínio PNT está localizado na porção N-terminal da proteína e o domínio ETS na porção C-terminal [26], na PDEF ambos ocupam posições adjacentes na porção C-terminal da cadeia, entre os nucleotídeos 142-210 e 248-331 para os domínios PNT e ETS, respectivamente [38].

Além destes domínios conservados, a PDEF também apresenta em sua seqüência de aminoácidos potenciais sítios de fosforilação por MAPK e tirosina cinases [38].

O domínio ETS da PDEF reconhece e se liga à seqüência de nucleotídeos 5'-GGAT-3' com alta afinidade, sendo a alta afinidade por esta seqüência específica uma característica peculiar da PDEF. Demais membros da família ETS têm preferência pela seqüência de nucleotídeos 5'-GGAA-3', havendo algumas proteínas que reconhecem a variante 5'-GGAT-3', apresentando porém baixa afinidade de ligação. Esta diferença de afinidade de ligação resulta no reconhecimento de seqüências regulatórias que não são sítios de ligação para as demais proteínas ETS [38,39].

A PDEF apresenta várias substituições de aminoácidos na hélice 3, a hélice de reconhecimento da cadeia de DNA [25], assim como na volta e na alça que caracteristicamente interagem também com a cadeia de DNA no motivo estrutural de reconhecimento WHTH; estas substituições

podem explicar a diferença de especificidade de ligação da PDEF em relação à família ETS [39]. Diferentes estudos indicam a presença de resíduos conservados na hélice 3 como responsáveis pelo reconhecimento e ligação aos três primeiros nucleotídeos do sítio de ligação (5'-GGA-3') [33,42] e alterações na afinidade de ligação das proteínas ETS observadas em decorrência da substituição de apenas um resíduo de aminoácido na seqüência desta hélice [42,43], bem como em aminoácidos localizados na porção N-terminal da proteína, cuja substituição afeta a conformação do domínio ETS [44,45].

Dados experimentais obtidos a partir das estruturas cristalográficas e ensaios de afinidade de ligação apontam, porém o mecanismo de reconhecimento intramolecular como componente principal na formação do complexo proteína-DNA para proteínas da família ETS, onde estas proteínas reconhecem uma conformação específica da cadeia de DNA [46,47], conformação esta dependente da sua seqüência de pares de bases [48,49].

A cadeia de DNA nestas estruturas apresenta curvaturas entre 11 e 28° no complexo formado, também são observadas pontes de hidrogênio entre resíduos conservados da terceira hélice e a seqüência de ligação de alta afinidade no sulco maior da cadeia de DNA, e contatos diretos e mediados por moléculas de água entre os resíduos da volta entre as hélices 2 e 3 e da alça entre as fitas 3 e 4, com os fosfatos dos sulcos menores adjacentes ao sulco maior onde se localiza a hélice de

reconhecimento [47]. Os contatos observados, além dos formados entre base e aminoácido, ressaltam a importância dos pares de bases nas porções 5' e 3' da seqüência de ligação de alta afinidade e da conformação da cadeia de DNA, dois fatores relacionados entre si [48,49].

É possível que a combinação dos contatos com os átomos de fosfato da cadeia de DNA, as variações na seqüência primária de aminoácidos e variações sutis na conformação do domínio ETS, bem como a flexibilidade conformacional do DNA, contribuam para a preferência por sítios de ligação específicos com maior ou menor afinidade [39].

## 5.1 PDEF e Oncogênese

O câncer de próstata é o mais freqüente em homens no mundo e no Brasil, com risco estimado na região Sul de 68 casos a cada 100.000 homens e de 81.92/100.000 homens no Estado do Rio Grande do Sul (RS) [50]; é uma das causas mais freqüentes de morte em homens de terceira idade, e a falta de uma terapia efetiva decorre em parte da falta de conhecimento acerca dos mecanismos moleculares envolvidos no desenvolvimento e progressão da doença [51].

A expressão da PDEF no epitélio da próstata indica um papel ativo no desenvolvimento da glândula e possivelmente também no desenvolvimento de câncer. Estudos iniciais apontavam a possível relação entre o aumento da expressão de PDEF com o desenvolvimento de câncer, pois foram encontrados sítios de ligação de domínios ETS na região promotora do gene *psa*, codificante do antígeno específico da próstata PSA (*prostate specific antigen*) [38], proteína utilizada no diagnóstico do câncer de próstata.

A expressão da PDEF em epitélio da próstata estaria relacionada com a expressão de PSA de forma hormônio independente e também hormônio dependente através da sua ligação e co-regulação do receptor de androgênio AR (*androgen receptor*). Ensaios *in vitro* mostraram um aumento em 57 vezes da taxa de transcrição deste gene na presença de androgênio, AR e PDEF; e 11 vezes na presença apenas de PDEF,

indicando uma provável ação na expressão androgênio independente de PSA nos casos avançados de câncer de próstata [38].

Estudos posteriores demonstraram, porém, a ausência da proteína PDEF em células cancerígenas do epitélio da próstata, e a sua expressão apenas nas células saudáveis deste tecido [48]. Nas células que desenvolviam processos cancerígenos foram observadas grandes quantidades de mRNA da proteína PDEF apesar da ausência da proteína, indicando que a PDEF só é traduzida nas células saudáveis e que a regulação da sua expressão é pós-transcricional (**Figura 1**) [52].

Assim como a perda da expressão de PDEF observada em linhagens celulares em casos avançados de câncer de próstata, uma expressão reduzida da proteína PDEF em células invasivas de câncer de mama e sua ausência em linhagens celulares cancerígenas derivadas de tecido mamário também é descrita, acompanhada do aumento na quantidade de mRNA presente [53,54].

A indução da expressão normal de PDEF inibe a invasão, migração e crescimento destas células, indicando que a perda do fator de transcrição PDEF leva a uma alteração na expressão de genes relacionados com o controle e progressão do câncer de mama, possuindo papel no potencial de metástase destas células [53].

O câncer de mama é o segundo tipo de câncer mais freqüente no mundo e o primeiro entre as mulheres, embora seja o câncer maligno de diagnóstico mais comum [55]. Sua estimativa de risco é de 69 casos a

cada 100 mil mulheres para a região Sul do Brasil e de 88.81/100.000 no Estado do RS [50]. Sua mortalidade é atribuída à alta taxa de metástase e a ausência de resposta à quimioterapia e radioterapia [55].

Quando normalmente expressa, a PDEF inibe a transcrição do gene *uPA* e ativa a transcrição do gene *maspin*. *uPA* é um marcador de células mamárias em metástase e é responsável pela degradação da membrana basal; *maspin* é um gene supressor de tumor com propriedades anti-metástase quando expresso em células mamárias cancerígenas e invasivas [36,53]. A perda da PDEF resulta num aumento da transcrição do gene *uPA* e diminuição do *maspin*, efeitos que podem contribuir para a invasividade e metástase [36].

A indução da expressão de PDEF em linhagens celulares cancerígenas inibe ainda o crescimento celular; mantendo as células no estágio G<sub>0</sub>-G<sub>1</sub>, através da inibição da ciclina A [53].

Chama atenção ainda o fato de apesar da perda de PDEF; a quantidade de mRNA sintetizada a partir do seu gene é aumentada nas células mamárias cancerígenas [53,54], assim como nas células de próstata cancerígenas [52], corroborando a hipótese de um mecanismo de controle pós-transcricional ou ainda, pós-traducional (**Figura 1**) para este fator de transcrição.

Estudos posteriores porém, identificaram aumento na expressão também da proteína PDEF em tumores de mama [56] quando comparada a sua expressão em tecido normal; assim como o aumento da



quantidade de mRNA observada, de acordo com estudos anteriores [52,53,54].

Estudos mais recentes sobre a atividade da proteína PDEF apontam para a sua regulação pela via MAPK, através da proteína fosfatase ERK, que fosforila resíduos específicos de treonina localizados no domínio PNT, e que esta ativação via fosforilação da PDEF leva a um aumento de 10 a 30 vezes na mobilidade celular [57], o que sugere que a proteína PDEF está de fato relacionada com o aumento da expressão do gene *uPA*, e que a ativação da via MAPK/ERK colabora com a indução da invasão celular mediada pela PDEF [57].

A co-relação entre a repressão do gene *uPA* e a proteína PDEF foi sugerida inicialmente com estudos comparativos entre as seqüências da proteína PDEF e outras proteínas ETS cuja ação sabidamente é de repressão da expressão deste gene [36]. É sabido que a ação destes fatores de transcrição é dependente do contexto celular, e que podem atuar tanto como ativadores quanto repressores em função do controle combinatório [26,28,29]. É possível então, que apesar da similaridade entre as seqüências de aminoácidos do domínio ETS e PNT da proteína PDEF com outras proteínas que reprimem a expressão do gene *uPA*, ela exerça um controle positivo sobre a sua transcrição pela associação com outros fatores de transcrição ou proteínas regulatórias.

A atribuição da atividade de indutor do gene *maspin* foi inferida da mesma forma que a atividade de repressor do gene *uPA*;

entretanto, estudos mais recentes indicam a ausência de relação entre a expressão da PDEF e do gene *maspin*, cuja transcrição seria unicamente dependente da presença de p63 [58].

O papel da PDEF na ativação e expressão da PSA foi corroborado pela identificação do gene NKX-3.1 e da sua interação com a proteína PDEF. A perda da atividade de NKX-3.1 está relacionada com o desenvolvimento de câncer de próstata, e a sua expressão evita a função de ativador da transcrição do promotor PSA pela PDEF [59,60]. Assim, a perda da atividade deste gene liberaria o controle negativo sobre a PDEF, levando ao aumento da transcrição de PSA observado no câncer de próstata.

## 6. A Estrutura do DNA

Estudos realizados por Rosalind Franklin, James Watson e Francis Crick em 1953 [61,62,63] revelaram não apenas a estrutura em dupla hélice da cadeia de DNA como também a existência de dois tipos de cadeias, DNA A e DNA B, com características estruturais particulares. Posteriormente, em 1972, uma terceira forma de DNA foi também descrita, o DNA Z (Tabela 2) [64].

**Tabela 2.** Características estruturais principais das diferentes conformações da cadeia de DNA (1 nm = 10 Å). Adaptado de Bates & Maxwell, 1993.

Conformação da cadeia de DNA	Orientação da hélice	Pares de base/volta	Diâmetro da hélice (nm)	Sulco maior	Sulco menor
B	Horária	10	~2.0	Grande e profundo	Pequeno e profundo
A	Horária	11	~2.6	Pequeno e profundo	Grande e raso
Z	Anti-horária	12	~1.8	Raso	Pequeno e profundo

O DNA B equivale à forma mais comum adotada pela cadeia de DNA e corresponde à forma observada com maior frequência *in vivo*. Sua estrutura foi determinada experimentalmente por cristalografia por difração de raios X de fibras de DNA em condições de 92% de umidade. A característica que o difere das demais formas é a orientação quase perpendicular das bases em relação ao eixo da hélice e os sulcos maior e menor da cadeia, com acesso facilitado às bases pelo sulco maior [64]. Seus parâmetros estruturais (Tabela 2) definem a estrutura canônica da molécula de DNA, embora a molécula observada *in vivo* não corresponda a

uma molécula de DNA B perfeita, uma vez que tanto a seqüência de nucleotídeos quanto as condições do meio e a formação de complexos proteína-DNA afetam a sua estrutura local [48,49,65].

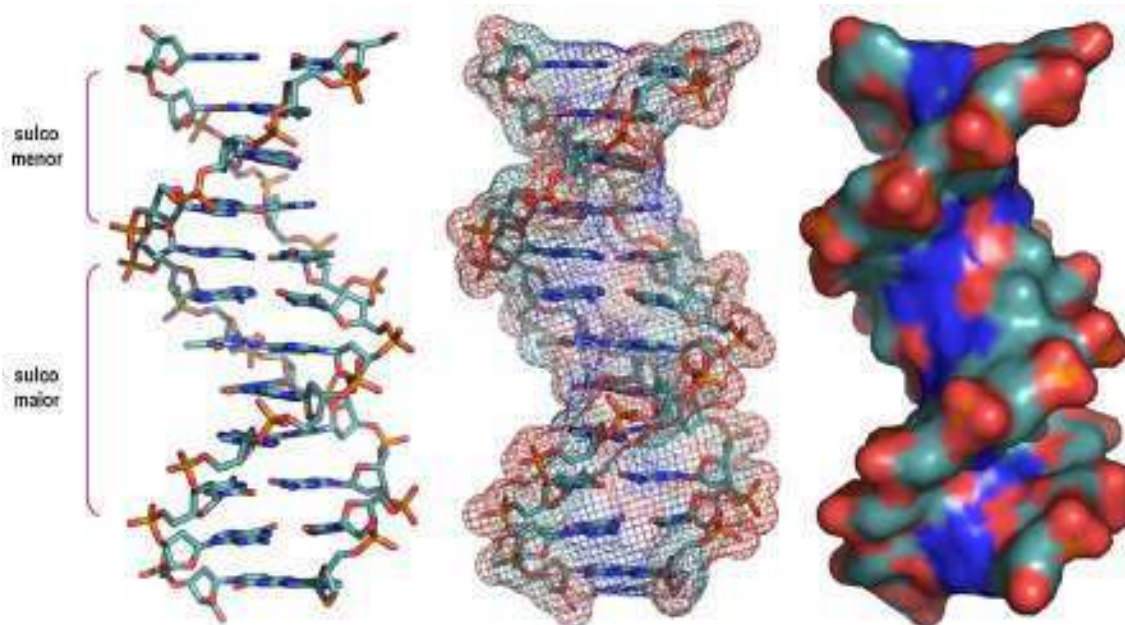
A forma DNA A foi identificada em estudos de cristalografia por difração de raios X de fibras de DNA sob condições de baixa umidade (75%), e equivale à forma B desidratada. Seus sulcos não são tão profundos quanto os da forma DNA B e seus pares de base não apresentam disposição perpendicular. Seqüências de bases homopurinas ou homopirimidinas (poli-G e poli-C, por exemplo) apresentam características estruturais da hélice de DNA A [66]. Assume-se que na molécula de DNA B observada *in vivo*, algumas de suas regiões estejam na forma DNA A em decorrência da sua seqüência de nucleotídeos [66].

A cadeia de DNA Z é a única de orientação anti-horária, e apresenta conformação em *zig-zag* da cadeia principal de açúcar-fosfato. Esta forma é observada em seqüências onde há alternância de bases púricas e pirimídicas e é favorecida por condições de alta salinidade e em regiões de superenrolamento da cadeia de DNA, atuando na manutenção da sua integridade e como provável elemento regulatório no controle da expressão gênica [66,67,68]. A forma DNA Z é dinâmica, sendo formada como resultado de processos fisiológicos (transcrição gênica, por exemplo), depois relaxando e retornando à forma DNA B [68,69].

Outras variações da estrutura da cadeia de DNA podem se formar sob condições especiais, como DNA C, DNA D e DNA T [66], com

parâmetros estruturais característicos. A ocorrência de diferentes formas da cadeia de DNA ilustra a sua plasticidade estrutural [64].

A ligação assimétrica das bases aos anéis de ribose da cadeia principal fosfato-açúcar confere a característica assimetria da cadeia de DNA em sulco maior e menor (**Figura 8**), essencial para os processos de replicação, transcrição e controle da expressão gênica, pois todos dependem do reconhecimento de seqüências específicas da cadeia de DNA por proteínas, reconhecimento este facilitado pela maior exposição das bases no sulco maior da cadeia. Algumas proteínas, entretanto reconhecem e se ligam ao sulco menor da cadeia de DNA, como por exemplo, a proteína TBP [46].



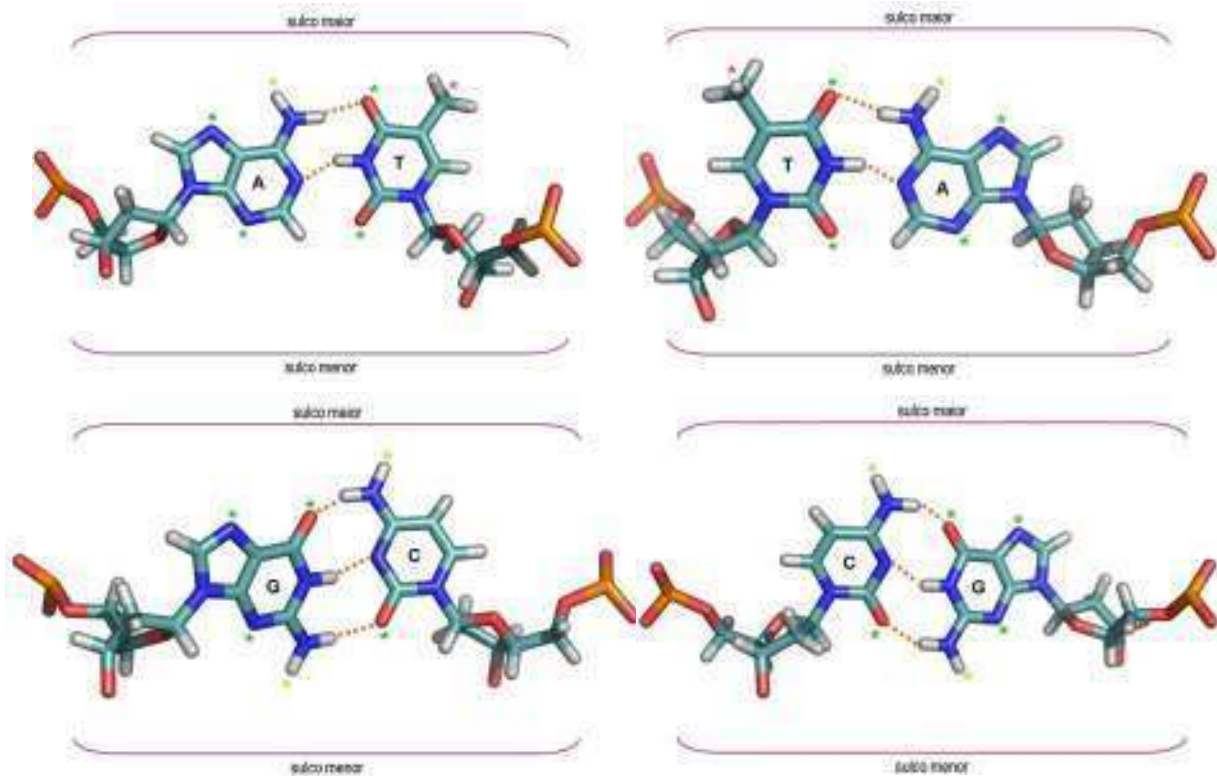
**Figura 8.** Sulco maior e menor da cadeia de DNA B (estrutura canônica). A mesma cadeia é mostrada em três representações diferentes, no modelo de palitos onde se pode notar a disposição quase perpendicular das bases em relação ao eixo da hélice; e em duas representações da superfície da cadeia, no modelo *mesh* e de esferas, onde a diferença entre os sulcos é melhor visualizada. Nas três representações os átomos de carbono estão em azul claro, nitrogênio em azul escuro, oxigênio em vermelho, fósforo em laranja. Átomos de hidrogênio suprimidos para maior clareza da figura (código PDB: 114D).

As proteínas regulatórias reconhecem e se ligam a seqüências específicas de nucleotídeos na cadeia de DNA. Inicialmente especulou-se ser necessário um acesso direto às pontes de hidrogênio entre as bases para o reconhecimento das seqüências, porém, a superfície da dupla hélice da cadeia de DNA provê informação suficiente para a diferenciação e reconhecimento dos pares de base pelos fatores de transcrição, sem ser necessário romper a sua estrutura [4].

Cada par de bases possui uma superfície característica com padrão único de doador de hidrogênio, acceptor de hidrogênio e grupamentos hidrofóbicos que permite a sua identificação tanto no sulco maior quanto no sulco menor da cadeia de DNA (**Figura 9**) [65].

No entanto, apenas a superfície exposta no sulco maior da cadeia é que permite a diferenciação clara entre os pares de bases (**Figura 10**), sendo por esta razão a superfície onde geralmente se ligam as proteínas regulatórias [16].

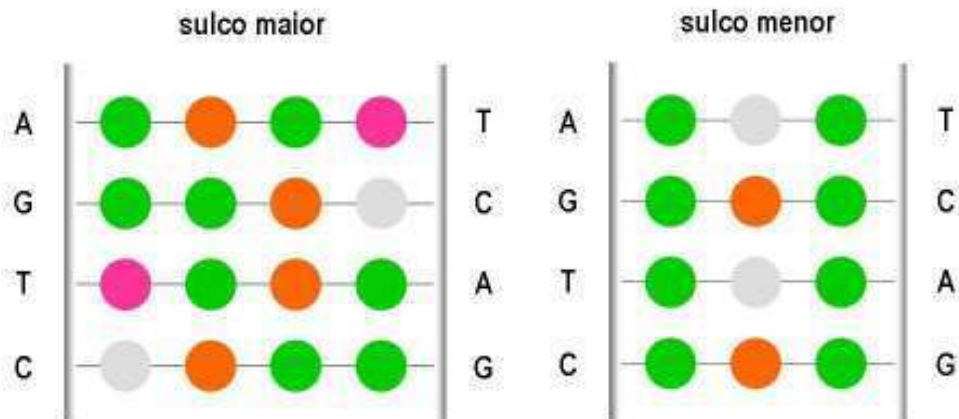
Os padrões de distribuição de grupamentos aceptores e doadores de hidrogênio são utilizados no reconhecimento da cadeia de DNA pelos fatores de transcrição; padrões estruturais da própria cadeia de DNA determinados pela sua seqüência de nucleotídeos, que distorcem os parâmetros da hélice “clássica” (**Tabela 2**, DNA B), também são importantes na identificação de seqüências regulatórias [16,46,49,70].



**Figura 9.** Os diferentes pares de bases da cadeia de DNA podem ser reconhecidos pelas suas superfícies desiguais. Podem-se observar ainda as diferenças no mesmo par de bases entre a superfície exposta na face do sulco maior e sulco menor da dupla hélice do DNA. As faces do sulco maior e menor da cadeia de DNA estão indicadas para cada par de bases. As bases estão representadas no modelo de palitos e coloridas como na Figura 6. Os grupos potenciais doadores de hidrogênio estão assinalados em amarelo, potenciais aceptores de hidrogênio em verde e grupamentos metila (CH<sub>3</sub>) hidrofóbicos em rosa. Os átomos de hidrogênio ligados a átomos de carbono e não disponíveis para a formação de pontes de hidrogênio não estão destacados. Pontes de hidrogênio representadas pelas linhas pontilhadas. A= adenina, T= timina, C= citosina e G= guanina.

O reconhecimento da cadeia de DNA por proteínas regulatórias diretamente através de contatos atômicos é chamado de padrão de leitura ou de reconhecimento intermolecular. Já o reconhecimento indireto por meio de propriedades conformacionais da cadeia de DNA é chamado de padrão de leitura ou reconhecimento intramolecular. Da mesma forma que mutações na seqüência de aminoácidos da proteína afetam a

especificidade de ligação, alterações na conformação ou na flexibilidade da cadeia de DNA também o fazem [46].



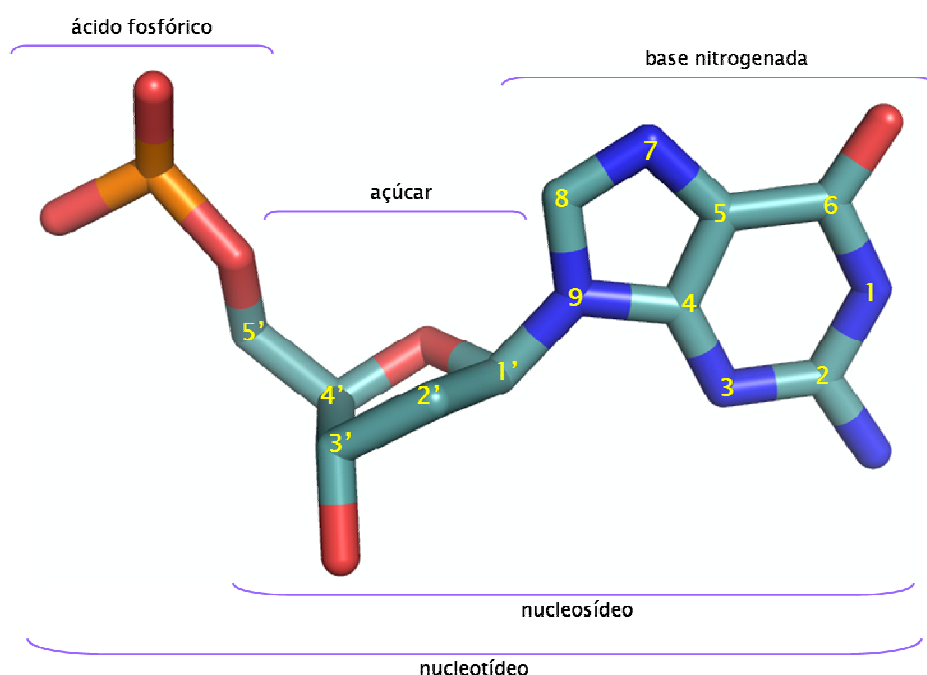
**Figura 10.** Representação esquemática dos grupamentos expostos nos sulcos da cadeia de DNA mostrados na Figura 9. Fatores de transcrição podem reconhecer os grupamentos expostos nos sulco maior e menor da cadeia de DNA. Cada par de bases possui um padrão característico de grupamentos que permite sua identificação e diferenciação no sulco maior da dupla hélice. Já o padrão de grupamentos expostos no sulco menor não permite a diferenciação entre pares AT e TA, GC e CG. Círculos verdes indicam grupamentos potenciais aceptores de hidrogênio, círculos laranja indicam potenciais doadores de hidrogênio, círculos rosa indicam os grupamentos hidrofóbicos e os círculos brancos indicam os átomos de hidrogênio ligados a átomos de carbono e não disponíveis para a formação de pontes de hidrogênio. A= adenina, T= timina, C= citosina e G= guanina (adaptado de Branden & Tooze, 2ª Ed., 1999).

Os padrões de reconhecimento intermolecular e intramolecular contribuem para a especificidade de ligação na formação dos complexos proteína-DNA em proporções que variam para cada complexo, não parecendo haver associação entre motivos estruturais de reconhecimento da cadeia de DNA ou estruturas secundárias de proteínas e preferência por um ou outro mecanismo de reconhecimento [46].



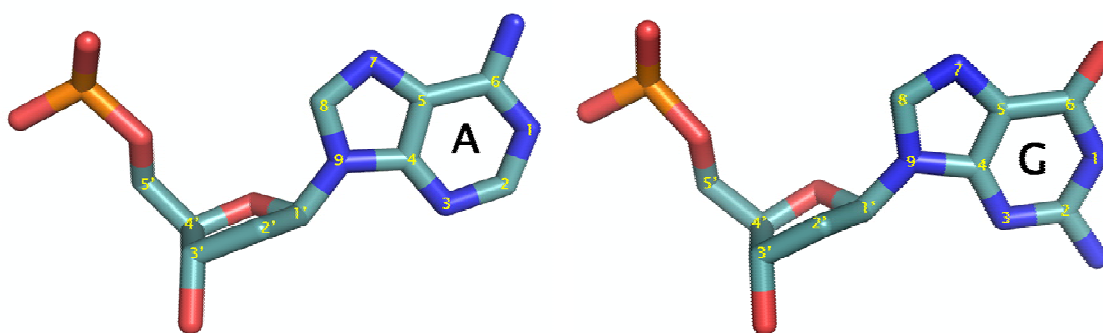
## 6.1 Parâmetros Conformacionais da Cadeia de DNA

A cadeia de DNA é formada por nucleotídeos unidos por ligações fosfodiéster. Cada nucleotídeo é composto por uma base nitrogenada, um açúcar e um grupamento fosfato, unidos por ligações covalentes (Figura 11).

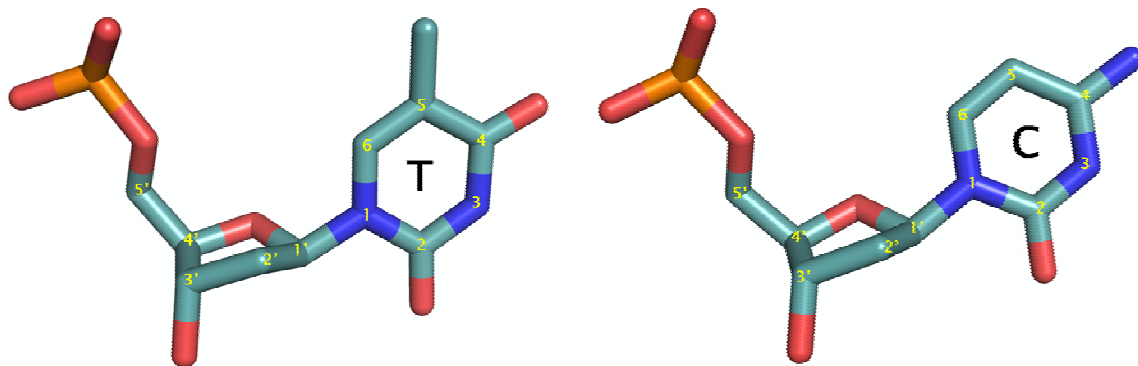


**Figura 11.** Estrutura de um nucleotídeo, unidade básica da cadeia de DNA. A ligação glicosídica entre o açúcar e a base nitrogenada forma o nucleosídeo. A adição de um ácido fosfórico ao carbono C5' do açúcar forma o nucleotídeo. O açúcar observado nas moléculas de DNA ( $\beta$ -D-desoxirribose) apresenta um hidrogênio ligado ao carbono C2' do anel; em moléculas de RNA existe um grupamento hidroxila ligado a este átomo. O nucleotídeo mostrado é a desoxiguanosina. Estrutura representada no modelo de palitos. Átomos de carbono em ciano, nitrogênio em azul, oxigênio em vermelho e fósforo em laranja. Os átomos de hidrogênio foram omitidos para maior clareza da figura. Numeração padrão dos átomos da base e açúcar em amarelo.

As bases nitrogenadas de ocorrência mais comum são divididas em bases púricas, cuja estrutura é formada por dois anéis – adenina (A) e guanina (G) (**Figura 12**); e bases pirimídicas, com apenas um anel em sua estrutura – timina (T) e citosina (C) (**Figura 13**). O açúcar observado nas cadeias de DNA é a  $\beta$ -D-desoxirribose, que apresenta um hidrogênio ligado ao carbono C2' do anel da ribose, o que o diferencia do açúcar observado nas moléculas de RNA, uma  $\beta$ -D-ribose, que apresenta um grupamento hidroxila nesta mesma posição (**Figura 11**). A ligação glicosídica entre a base nitrogenada e o anel da ribose se dá através do carbono C1' do açúcar e o nitrogênio N1 para as pirimidinas e N9 para as bases púricas [71].



**Figura 12.** Bases nitrogenadas púricas: adenina (A – à esquerda) e guanina (G – à direita). Os nucleotídeos estão representados e coloridos como na Figura 11. Existem ligações duplas entre os átomos C2 e N3, C4 e C5, N7 e C8 para adenina e guanina e ainda uma ligação dupla adicional entre os átomos N1 e C6 na adenina. A numeração dos átomos de cada um dos nucleotídeos está indicada em amarelo.



**Figura 13.** Bases nitrogenadas pirimídicas: timina (T - à esquerda) e citosina (C - à direita). Os nucleotídeos estão representados e coloridos como na Figura 11. Existem ligações duplas entre os átomos N1 e C2, C5 e C6 para timina e citosina e ainda uma ligação dupla adicional entre os átomos N3 e C4 na citosina. A numeração dos átomos de cada um dos nucleotídeos está indicada em amarelo.

A estrutura formada pela base nitrogenada e pelo açúcar conectados pela ligação glicosídica é chamada de nucleosídeo (**Figura 11**). Quando um ácido fosfórico é esterificado com um grupo hidroxila da porção açúcar de um nucleosídeo, forma-se um nucleotídeo [71]; sendo os nucleotídeos 5' (ligação do ácido fosfórico ao carbono C5' do açúcar) os de ocorrência mais comum.

A formação da estrutura em dupla hélice da cadeia de DNA se dá pela ligação fosfodiéster entre nucleotídeos adjacentes, e pelo pareamento das bases complementares em cada uma das cadeias da hélice.

A ligação entre os nucleotídeos adjacentes ocorre entre o átomo de fósforo e os grupos hidroxila dos carbonos C3' e C5' destes nucleotídeos, originando uma ligação 3',5'-fosfodiéster. A numeração dos

nucleotídeos de cada uma das cadeias é feita a partir da extremidade 5' que apresenta um grupo fosfato, até a extremidade 3' que apresenta uma hidroxila livre.

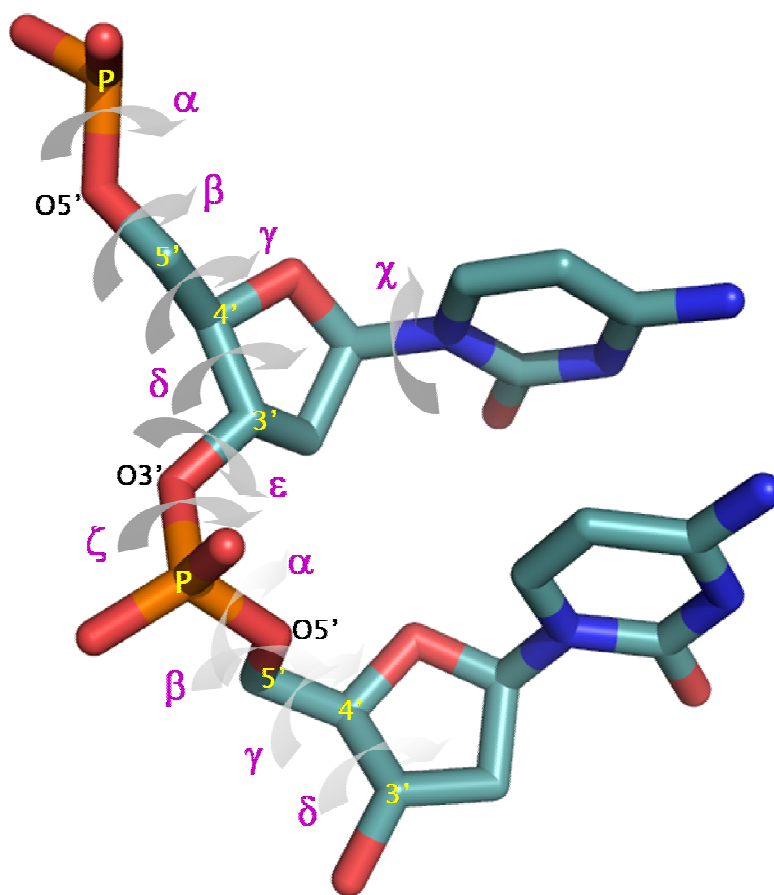
O pareamento das bases complementares ocorre por pontes de hidrogênio entre os grupos eletrofílicos doadores e aceptores de hidrogênio de cada base. O pareamento ideal descrito por Watson e Crick [61] ocorre entre uma base púrica e uma base pirimídica, mais especificamente entre A e T e entre C e G, com duas e três pontes de hidrogênio respectivamente (**Figura 9**). As cadeias de DNA que compõem a dupla hélice são complementares e antiparalelas.

A cadeia principal fosfato-açúcar da molécula de DNA apresenta seis ângulos conformacionais (**Figura 14**), o que confere às cadeias polinucleotídicas uma maior flexibilidade quando comparadas com as cadeias polipeptídicas, pelo maior número de ângulos rotacionais.

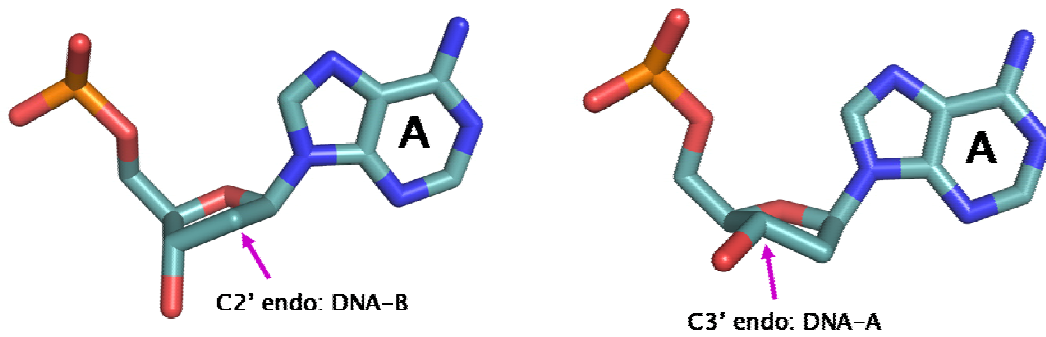
A disposição planar dos átomos que compõem o anel do açúcar também afeta a conformação da cadeia de DNA (denominado empacotamento do açúcar) e caracteriza formas distintas de moléculas de DNA [72]. A conformação C2'-endo é típica de moléculas de DNA B, enquanto a conformação C3'-endo é observada em moléculas de DNA A (**Figura 15**) [73].

Os parâmetros conformacionais empregados na definição da geometria da cadeia de DNA foram normalizados e definidos inicialmente em 1989 [48] e atualizados em 2001 [49]. Tais parâmetros compreendem

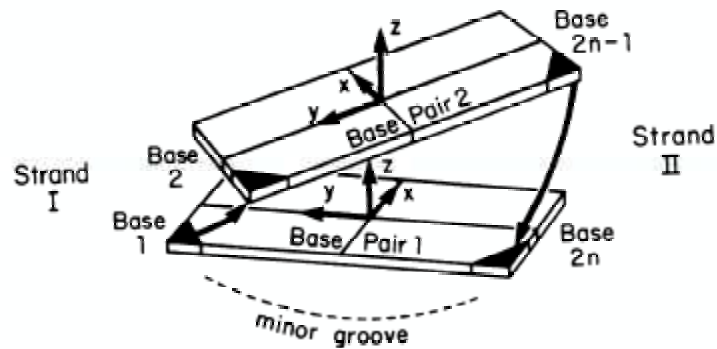
ângulos e distâncias entre as bases de um par (AT, TA, CG e GC) e entre pares de bases adjacentes na dupla hélice (Figuras 16 e 17) e são todos dependentes da seqüência de nucleotídeos, além de apresentarem correlações entre si [49].



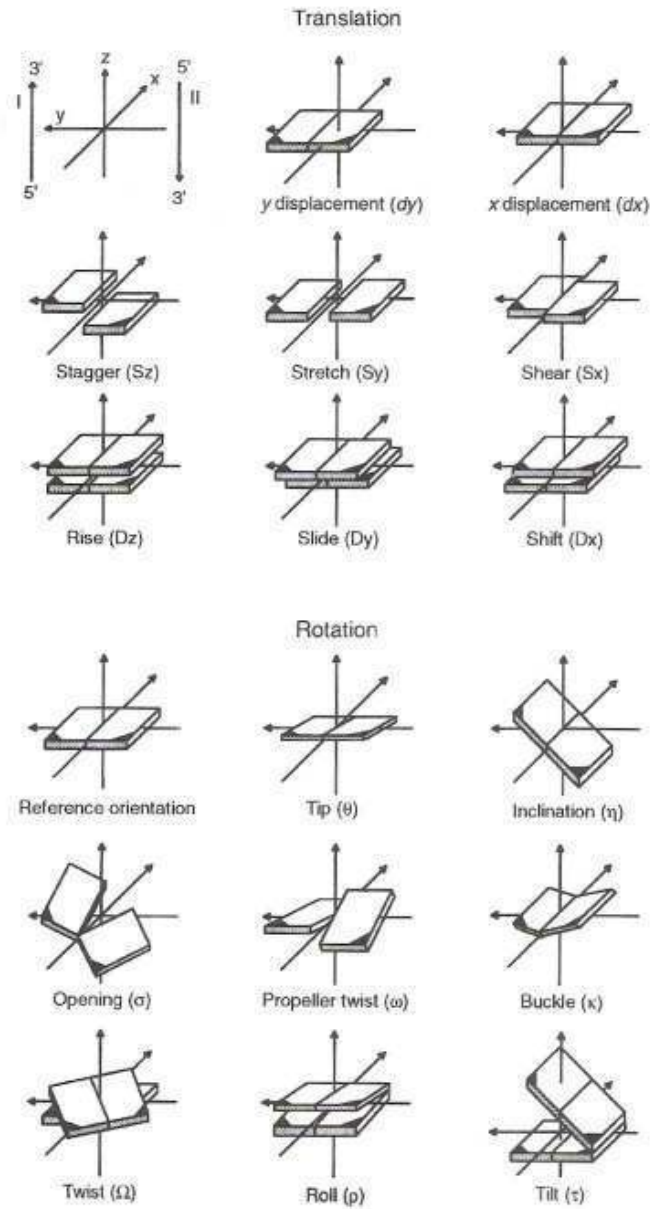
**Figura 14.** A conformação do esqueleto fosfato-açúcar em uma cadeia de DNA é definida pelos ângulos rotacionais  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  e  $\zeta$ . A orientação da base nitrogenada em relação ao açúcar é dada pelo ângulo  $\chi$ . Estrutura da cadeia de DNA representada e colorida como na Figura 11. Cada um dos ângulos está indicado por setas em cinza. Em cinza mais claro está indicada a repetição dos ângulos diedros no nucleotídeo adjacente.



**Figura 15.** Disposição dos átomos do açúcar no plano do açúcar para um nucleotídeo de adenina. À direita, conformação C2'-endo tipicamente observada na forma DNA B. À esquerda, o arranjo dos átomos do açúcar característico da forma DNA A, C3'-endo.



**Figura 16.** Definição dos eixos x, y e z para os pares de bases da cadeia de DNA. A partir do sulco menor da cadeia, eixo x na direção do sulco maior, eixo y na direção da base para a cadeia fosfato-açúcar e eixo z na direção do eixo da hélice. Adaptado de Dickerson, *et al. JMB.* 205:787-791, 1989.



**Figura 17.** Principais parâmetros conformacionais utilizados na descrição da geometria da cadeia de DNA. Eixos x, y e z definidos conforme a Figura 16. Demais parâmetros conformacionais e figuras originais na referência [48].

## - Considerações Finais -

As funções biológicas das cadeias de DNA e a sua estrutura são dependentes da sua seqüência de nucleotídeos [49,117]. Uma nova forma de codificação de informação, além da seqüência de trincas de bases traduzida em aminoácidos, está sendo reconhecida, onde a conformação adotada pela dupla hélice também tem papel nas suas funções [130].

O uso de simulações pela dinâmica molecular permite a caracterização da estrutura atômica e do movimento de macromoléculas em sistemas aquosos e, em última análise, a quantificação das diferentes forças que atuam em um sistema biológico como uma cadeia de DNA [131], provendo uma descrição dos eventos dinâmicos ao longo de uma trajetória em um nível não acessado pelos métodos clássicos de estudos de estruturas de macromoléculas biológicas (cristalografia por difração de raios X e RMN).

Foram realizadas simulações pela DM de quatro sistemas contendo tridecâmeros de DNA representando os sítios de ligação de alta afinidade (**1YO5**, **GGAT**) [38], de baixa afinidade (**GGAG**) [74] e sistema controle (**GGAA**), sem afinidade de ligação pela proteína PDEF [38]. Cada sistema foi neutralizado pela adição de contra-íons e solvatados de forma a manter as condições de soluções diluídas [83], com uma camada de água



suficientemente espessa para evitar a interação entre as imagens das cadeias de DNA nas condições periódicas de contorno [84,85].

Cada um dos sistemas foi aquecido lentamente [86,89] até a temperatura de 298.16K e pressão de 1 atm, onde foi equilibrado até atingir a fase de produção. O tempo de simulação foi de 5ns para os quatro sistemas estudados.

Estudos realizados em cadeias de cadeia de DNA B compostas por 12 pares de bases indicam que os parâmetros helicoidais e conformacionais das cadeias de DNA apresentam um tempo de relaxação inferior a 500ps [132], e que um tempo de simulação igual a dez vezes o tempo de relaxação destas variáveis deve ser suficiente na ausência de sub-estados no decorrer de trajetória dinâmica [133,134]. A escala de tempo de 5ns utilizada, além de ter sido suficiente para que os sistemas entrassem em fase de produção, também satisfaz a premissa acima.

Os resultados destas simulações pela DM permitem um maior entendimento dos fatores intrínsecos à cadeia de DNA relacionados com a formação dos complexos proteína-DNA para a proteína PDEF, membro da família ETS de fatores de transcrição.

A análise bioinformática da seqüência primária das proteínas ETS indicam que o padrão de distribuição dos grupamentos aceptores e doadores de hidrogênio e dos grupamentos hidrofóbicos da cadeia de DNA e a sua complementaridade com a seqüência de aminoácidos da hélice 3 da proteína PDEF e a presença apenas nesta subfamília do resíduo

de glutamina na posição 69 parecem ser um dos motivos da diferença de afinidade observada entre a PDEF e as demais proteínas ETS.

A substituição observada nas proteínas da subfamília PDEF, do resíduo de tirosina por uma glutamina na posição 69 (Y69Q), um resíduo sabidamente essencial para o reconhecimento e ligação à seqüência 5'-GGAA-3' [33,42] pode ter induzido a alteração da afinidade destas proteínas por seqüências promotoras distintas, levando ao reconhecimento com alta afinidade de seqüências que anteriormente apresentavam apenas baixa afinidade de ligação para as proteínas ETS.

Aliado às alterações na seqüência de aminoácidos da proteína PDEF; as alterações na conformação da cadeia de DNA induzidas pelas substituições das bases na posição +4 do sítio de ligação também são responsáveis pelo padrão de afinidade distinto observado na subfamília PDEF.

As estruturas das cadeias de DNA simuladas são mais similares (têm menor valor de RMSD) à estrutura experimental (1YO5.pdb) que às formas canônicas DNA B e DNA A; e mesmo livres em solução, apresentam uma curvatura intrínseca à sua seqüência de nucleotídeos, responsável pelo padrão indireto de reconhecimento. Estudos da afinidade de ligação da proteína Ets-1 pelas seqüências 5'-GGAA-3' e 5'-GGAG-3' também relatam a presença da curvatura intrínseca das cadeias de DNA reconhecidas pela proteína ETS, favorecendo o seu reconhecimento, cuja amplitude de curvatura depende da seqüência de nucleotídeos [74,78].

A substituição da base na posição +4 do sítio de ligação de alta afinidade (**1YO5** e **GGAT**) de uma timina para uma guanina, ou seja, o sítio de baixa afinidade (**GGAG**), leva a uma maior flexibilidade da cadeia de DNA e maior dependência da seqüência de nucleotídeos adjacentes, conferidas pelos dímeros formados nas posições +3/+4 e +4/+5 (os dímeros que distinguem cada um dos sistemas).

O sistema controle, **GGAA**, apresenta valores mais elevados de RMSD na comparação dos pares de bases que compõem o sítio de ligação da proteína à cadeia de DNA, provavelmente pela presença de duas timinas adjacentes nas posições +3' e +4', que provocariam um rearranjo na disposição das bases adjacentes. Em contraste, tanto os sítios de ligação dos sistemas de alta afinidade quanto o sítio de ligação do sistema controle, têm sua conformação estabilizada pela presença de dímeros rígidos nas posições +3/+4; o que pode facilitar o reconhecimento da conformação local da cadeia de DNA destes sistemas pelas proteínas ETS.

Todos os sistemas simulados possuem ainda purinas nas posições adjacentes aos dímeros +1/+2, +2/+3 e +3/+4 do sítio de ligação, que aumentam a sua estabilidade; e o dímero +4/+5, na porção final do sítio de ligação, apresenta flexibilidade conformacional necessária para acomodar as distorções induzidas pela proteína na formação do complexo proteína-DNA.

Os sistemas de baixa afinidade e sem afinidade de ligação para a PDEF (**GGAG** e **GGAA**) apresentam maior número de bases com transições

entre as conformações BI e BII, incluindo as bases do sítio de ligação; ao contrário da seqüência reconhecida com alta afinidade (1Y05 e GGAT).

**GGAG** apresenta ainda maior variação nos ângulos  $\alpha$ ,  $\beta$  e  $\gamma$ , que juntamente com as transições BI→BII podem explicar os valores de RMSD mais elevados em relação à estrutura experimental e às demais estruturas médias e a menor afinidade de ligação da proteína PDEF, como conseqüência da maior flexibilidade deste sistema.

A observação de alterações conformacionais nos ângulos  $\epsilon$  e  $\zeta$  e as conformações atípicas dos ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  observadas no sistema **GGAA**, além da sua menor curvatura, podem estar relacionadas com a ausência de afinidade da proteína PDEF por esta seqüência. Posto que o reconhecimento intramolecular é o principal componente na formação dos complexos proteína-DNA para as proteínas ETS [46,47], a conformação adotada pelo sistema **GGAA** pode não ser ideal para o reconhecimento e formação do complexo.

A comparação dos parâmetros helicoidais dos sistemas livres em solução, com a cadeia de DNA da estrutura experimental, indica que a formação do complexo proteína-DNA induz um desdobramento da dupla hélice, inicialmente curvada em direção ao sulco maior, para a acomodação das cadeias laterais da hélice 3 da PDEF no interior do sulco maior da cadeia de DNA.

Por fim, a substituição do quarto par de bases do sítio de ligação 5'-GGA(T/G/A)-3' influencia o reconhecimento da proteína PDEF

tanto pela substituição dos grupamentos expostos no sulco maior da cadeia de DNA (padrão de reconhecimento direto, ou intermolecular), quanto pelas alterações induzidas na conformação da hélice (padrão de reconhecimento indireto ou intramolecular).

A não conservação dos resíduos de aminoácidos que interagem com a cadeia de DNA, e o grande número de contatos com a cadeia fosfato-açúcar corroboram o padrão de leitura intramolecular como o principal mecanismo de reconhecimento da cadeia de DNA pelas proteínas ETS, conforme proposto por Szymczyna e Arrowsmith [47].

Um tempo maior de simulação pela DM pode ser necessário para a convergência completa da curvatura em todos os sistemas; no entanto, os resultados aqui apresentados indicam que a conformação (curvatura) da cadeia de DNA e os contatos entre os resíduos de aminoácidos das proteínas ETS, conservados ou não, com os fosfatos da cadeia de DNA (contatos não específicos) são responsáveis pelo reconhecimento indireto na formação destes complexos, e a estabilidade de ligação; ou seja, a “regulação fina” da formação dos complexos é mediada pelo reconhecimento direto e formação de pontes de hidrogênio específicas com as bases nas posições +4 e +4’ do sítio de ligação.

## - Perspectivas -

Na preparação dos sistemas estudados, as coordenadas iniciais para os átomos que compõem a cadeia de DNA foram determinadas conforme as formas canônicas DNA B e DNA A (sistemas GGAT, GGAG e GGAA) ou a partir da estrutura experimental 1yo5.pdb (sistema 1YO5).

As moléculas de água foram então adicionadas em posições aleatórias no interior da caixa periódica estabelecida, e suas posições foram relaxadas lentamente.

As coordenadas das águas cristalográficas não foram consideradas, nem mesmo as coordenadas das moléculas presentes na interface proteína-DNA. No entanto, moléculas de água medeiam pontes de hidrogênio entre proteínas e cadeias de DNA (“pontes de água”) e podem ter um papel importante na especificidade e na afinidade de ligação [135]; além de apresentarem uma alta estabilidade nestas interfaces [136,137].

A realização de simulações pela DM com as coordenadas iniciais utilizadas em cada um dos quatro sistemas, levando em consideração as coordenadas das águas cristalográficas na interface proteína-DNA que medeiam interações entre a proteína e a cadeia de DNA permitirá a avaliação do tempo de residência destas moléculas ao longo da trajetória dinâmica dos sistemas simulados e do seu papel na formação do

complexo, principalmente da molécula de água WAT34 (de acordo com a estrutura experimental 1yo5.pdb), envolvida na formação das pontes de hidrogênio entre o resíduo Q69 e as bases nas posições +3 e +4 do sítio de ligação.

Considerando apenas a molécula de DNA livre em solução, a influência das seqüências flanqueadoras do sítio de ligação nas posições 5' e 3' podem ser inferidas pela alteração sistemática destes nucleotídeos. Dados experimentais [47] indicam que apenas as substituições de nucleotídeos que afetam a conformação da cadeia de DNA têm influência sobre a sua afinidade de ligação pelas proteínas ETS. Substituições sistemáticas de nucleotídeos nas posições flanqueadoras permitirão avaliar os efeitos na conformação tridimensional da hélice de DNA.

Simulações pela DM do domínio ETS da proteína PDEF contendo o resíduo de glutamina na posição 69 e do domínio de ligação de outra proteína ETS que apresente um resíduo de tirosina nesta posição, com as estruturas dos tridecâmeros de DNA contendo os sítios de ligação de alta afinidade, baixa afinidade e sistema controle (em relação à proteína PDEF) obtidas como resultado deste trabalho possibilitarão um maior entendimento do efeito da substituição Q69Y na dinâmica da formação dos contatos entre proteína e DNA e conseqüentemente na afinidade de ligação dos complexos.

**- Referências Bibliográficas -**

[1] Gurdon, JB. **Transplanted nuclei and cell differentiation.** *Scientific American.* 219(6):24-35, 1968.

[2] Gurdon, JB; Byrne, JA & Simonsson, S. **Nuclear reprogramming and stem cell creation.** *Proceedings of National Academy of Science.* 100(S1):11819-11822, 2003.

[3] Collins, AM; Sewell, WA & Edwards, MR. **Immunoglobulin gene rearrangement, repertoire diversity, and the allergic response.** *Pharmacology & therapeutics.* 100:157-170, 2003.

[4] Alberts, B; Johnson, A; Lewis, J; Raff, M; Roberts, K & Walter P. **Molecular Biology of the Cell.** 4ª Edição. Garland Publishing, New York. 2001.

[5] Smith, C; Marks, A & Lieberman, M. **Mark's Basic Medical Biochemistry - A Clinical Approach.** 2º Edição. Lippincott Williams & Wilkins. 2004.

[6] Kornberg, RD. **Mediator and mechanism of transcriptional activation.** *TRENDS in Biochemical Sciences.* 30(5):235-239, 2005.

[7] Borukhov, S & Lee, J. **RNA polymerase structure and function at lac operon.** *Comptes Rendus Biologies.* 328:576-587, 2005.

[8] Malik, S & Roeder, RG. **Dynamic regulation of pol II transcription by the mammalian mediator complex.** *TRENDS in Biochemical Sciences.* 30(5):256-263, 2005.

[9] Conaway, RC; Sato, S; Tomomori-Sato, C; Yao, T & Conaway, JW. **The mammalian mediator complex and its role in transcriptional regulation.** *TRENDS in Biochemical Sciences.* 30(5):250-255, 2005.



- [10] Burley, SK. **The TATA box binding protein.** *Current Opinion in Structural Biology.* 6(1):69–75, 1996.
- [11] Sims, RJ III; Mandal, SS & Reinberg, D. **Recent highlights of RNA-polymerase-II-mediated transcription.** *Current Opinion in Cell Biology.* 16:263–271, 2004.
- [12] Hahn, S. **Structure and mechanism of the RNA polymerase II transcription machinery.** *Nature Structural and Molecular Biology.* 11(5):394–403, 2004.
- [13] Cramer, P. **RNA polymerase II structure: from core to functional complexes.** *Current opinion in Genetics and Development.* 14:218–226, 2004.
- [14] Chadick, JZ & Asturias, FJ. **Structure of eukaryotic mediator complexes.** *TRENDS in Biochemical Sciences.* 30(5):264–271, 2005.
- [15] Bost, F; Aouadi, M, Caron, B & Binétruy. **The role of MAPKs in adipocyte differentiation and obesity.** *Biochimie.* 87:51–56, 2004.
- [16] Branden, C & Tooze, J. **Introduction to Protein Structure.** 2<sup>a</sup> Ed. Garland Publishing, New York. 1999.
- [17] Marmorstein, R & Fitzgerald, MX. **Modulation of DNA-binding domains for sequence-specific DNA recognition.** *Gene.* 304:1–12, 2003.
- [18] Aravvind, L; Anantharaman, V; Balaji, S; Babu, MM & Iyer, LM. **The many faces of the helix–turn–helix domain: transcription regulation and beyond.** *FEMS Microbiological Reviews.* 29:231–262, 2005.
- [19] Laity, JH; Lee, BM & Wright, PE. **Zinc finger proteins: new insights into structural and functional diversity.** *Current Opinion in Structural Biology.* 11:39–46, 2001.

- [20] Wolfe, SA; Nekludova, L & Pabo, Co. **DNA recognition by Cys2His2 zinc finger proteins.** *Annu. Rev. Biophys. Biomol. Struct.* 3:183–212, 1999.
- [21] Jones, S. **An overview of the basic helix–loop–helix proteins.** *Genome Biology.* 5:226.1–226.6, 2004.
- [22] Massari, ME & Murre, C. **Helix–loop–helix proteins: regulators of transcription in eucaryotic organisms.** *Molecular and cellular Biology.* 20(2):429–440, 2000.
- [23] Atchley, WR & Fitch WM. **A natural classification of the basic helix–loop–helix class of transcription factors.** *Proceedings of National Academy of Science.* 94:5172–5176, 1997.
- [24] Blair, DG & Athanasiou, M. **Ets and retroviruses – transduction and activation of members of the Ets oncogene family in viral oncogenesis.** *Oncogene,* 19:6472–6481, 2000.
- [25] Sharrocks, AD. **The ETS–domain transcription factor family.** *Nature Reviews – Molecular Cell Biology.* 2:827–837, 2001.
- [26] Oikawa, T & Yamada, T. **Molecular biology of the ETS family of transcription factors.** *Gene.* 303:11–34, 2003.
- [27] Sementchenko, VI & Watson, DK. **Ets target genes: past, present and future.** *Oncogene,* 19:6533–6548, 2000.
- [28] Wasylyk, B; Hagman, J & Gutierrez–Hartmann, A. **ETS Transcription factors: nuclear effectors of the Ras–MAP–kinase signaling pathway.** *TIBS.* 23:213–216, 1998.
- [29] Li, R; Pei, H & Watson, DK. **Regulation of the Ets function by protein–protein interactions.** *Oncogene,* 19:6514–6523, 2000.

[30] Mackereth, CD; Schärpf, M; Gentile, LN, MacIntosh, SE, Slupsky, CM & McIntosh, LP. **Diversity in structure and function of the Ets family PNT domains.** *Journal of Molecular Biology*, 342:1249–1264, 2004.

[31] Seidel, JJ & Graves, BJ. **A ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors.** *Genes & Development*, 16:127–137, 2002.

[32] Mavrothalassitis, G & Ghysdael, J. **Proteins of the ETS family with transcriptional repressor activity.** *Oncogene*, 19:6524–6532, 2000.

[33] Verger, A & Duterque-Coquillaud, M. **When ETS transcription factors meet their partners.** *BioEssays*. 24:362–370, 2002.

[34] Seth, A & Watson, DK. **ETS transcription factors and their emerging roles in human cancer.** *European Journal of Cancer*, 41:2462–2478, 2005.

[35] Dittmer, J & Nordheim, A. **Ets transcription factors and human disease.** *Biochimica et Biophysica Acta*, 1377:F1–F11, 1998.

[36] Hsu, T; Trojanowska, M & Watson, DK. **ETS proteins in biological control and cancer.** *Journal of Cellular Biochemistry*. 91:896–903, 2004.

[37] Oikawa, T. **ETS transcription factors: Possible targets for cancer therapy.** *Cancer Sciences*. 95(8):626–633, 2004.

[38] Oettgen, P; Finger, E; Sun, Z; Akbarali, Y; Thamrongsak U; Boltax J; Grall, F; Dube, A; Weiss, A; Brown, L; Quinn, G; Kas, K; Endress G; Kunsch, C & Libermann TA. **PDEF, a novel prostate epithelium-specific ETS transcription factor, interacts with the androgen receptor and activates prostate-specific antigen gene expression.** *The Journal of Biological Chemistry*. 275(2):1216–1225, 2000.

[39] Wang, Y; Feng, L; Said, M; Balderman, S; Fayazi, Z; Liu, Y; Ghosh, D & Gulick, A. **Analysis of the 2.0 Å crystal structure of the protein–DNA**

complex of the human PDEF ETS domain bound to the prostate specific antigen regulatory site. *Biochemistry*. 44(19):7095–7106, 2005.

[40] Yamada, N; Tamai, Y; Miyamoto, H & Nozaki, M. Cloning and expression of the mouse *Pse* gene encoding a novel Ets family member. *Gene*, 241:267–274, 2000.

[41] Hsouna A; Watson, DK & Hsu, T. Developmental expression pattern of D-ets4, the *Drosophila* homologue of human PDEF. *Gene Expression Patterns*. 5:285–289, 2004.

[42] Bosselut, R; Levin, J; Adjadj, E & Ghysdael, J. A single amino-acid substitution in the Ets domain alters core DNA binding specificity of Ets1 to that of the related transcription factor Elf1 and E74. *Nucleic Acids Research*, 21(22):5184–5191, 1993.

[43] Pio, F; Assa-Munt, N; Yguerabide, J & Maki, RA. Mutants of ETS domain PU.1 and GGAA/T recognition: Free energies and kinetics. *Protein Science*, 8:2098–2109, 1999.

[44] Shore, P; Whitmarsh, A; Bhaskaran, R; Davis, RJ; Waltho, J & Sharrocks, AD. Determinants of the DNA-binding specificity of ETS domain transcription factors. *Molecular and Cellular Biology*, 16(7):3338–3349, 1996.

[45] Mo, Y; Vaessen, B; Johnston, K & Marmorstein, R. Structure of the Elk1–DNA complex reveals how DNA-distal residues affects ETS domain recognition of DNA. *Nature Structural Biology*, 7(4):292–297, 2000.

[46] Gromiha, MM; Siebers, JG; Selvaraj, S; Kono, H & Sarai, A. Intermolecular and intramolecular readout mechanisms in protein–DNA recognition. *Journal of Molecular Biology*. 337:285–294, 2004.

[47] Szymczyzna, BR & Arrowsmith, CH. DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein–DNA

**recognition.** *The Journal of Biological Chemistry.* 275(37):28363–28370, 2000.

[48] Dickerson, RE *et al.* **Definitions and nomenclature of nucleic acid structure parameters.** *Journal of Molecular Biology.* 205:787–791, 1989.

[49] Olson, WK *et al.* **A standard reference frame for the description of nucleic acid base-pair geometry.** *Journal of Molecular Biology.* 313:229–237, 2001.

[50] Instituto Nacional do Câncer. Coordenação de prevenção e Vigilância. **Estimativa 2006: Incidência de câncer no Brasil.** Ministério da Saúde, Secretaria de Atenção à Saúde. 94 p. Rio de Janeiro: INCA, 2005.

[51] Small, EJ. **Advances in prostate cancer.** *Current Opinion in Oncology.* 11(3):226–235, 1999.

[52] Nozawa, M; Yomogida, K; Kanno, N; Nonomura, N; Miki, T; Okuyama, A; Nishimune Y & Nozaki, M. **Prostate-specific transcription factor hPSE is translated only in normal prostate epithelial cells.** *Cancer Research.* 60:1348–1352, 2000.

[53] Feldman, RJ; Sementchenko, VI; Gayed, M; Fraig, MM & Watson, DK. **PDEF expression in human breast cancer is correlated with invasive potential and altered gene expression.** *Cancer Research.* 63:4626–4631, 2003.

[54] Ghadersohi, A & Sood, AK. **Prostate Epithelium-derived ETS transcription factor mRNA is overexpressed in human breast tumor and is a candidate breast tumor marker and a breast tumor antigen.** *Clinical Cancer Research.* 7:2731–2738, 2001.

[55] Debies, MT & Welch, DR. **Genetics basis of human breast cancer metastasis.** *Journal of Mammary Gland Biology and Neoplasia.* 6(4):441–451, 2001.

- [56] Galang, CK; Muller, WJ; Foos, G & Oshima, RG. **Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors.** *The Journal of Biological Chemistry*, 279(12):11281–11292, 2004.
- [57] Gunawardane, RN; Sgroi, DC; Wrobel, CN; Koh, E; Daley, GQ & Brugge, JS. **Novel role for PDEF in epithelial cell migration and invasion.** *Cancer Research*, 65(24):11572–11580, 2005.
- [58] Kim, S; Han, J, Kim, J & Park, Chaehwa. **Maspin expression is transactivated by p63 and is critical for the modulation of lung cancer progression.** *Cancer research*, 64:6900–6905, 2004.
- [59] Chen, H; Nandi, AK; Li, X & Bieberich, CJ. **NKX–3.1 interacts with prostate–derived Ets factor and regulates the activity of the *PSA* promoter.** *Cancer Research*, 62:338–340, 2002.
- [60] Chen, H & Bieberich, CJ. **Structural and functional analysis of domains mediating interaction between NKX–3.1 and PDEF.** *Journal of Cellular Biochemistry*, 94:168–177, 2005.
- [61] Watson, JD & Crick, FHC. **Molecular structure of nucleic acids – A structure for deoxyribose nucleic acids.** *Nature*. 171(4356):737–741, 1953.
- [62] Watson, JD & Crick, FHC. **Genetical implications of the structure of deoxyribonucleic acid.** *Nature*. 171(4361):964–967, 1953.
- [63] Franklin, RE & Gosling RG. **Evidence for 2–chain helix in crystalline structure of sodium deoxyribonucleate.** *Nature*. 172(4369):156–157, 1953.
- [64] Bates, AD & Maxwell A. **DNA topology.** Oxford University Press, 1993.
- [65] Cheatham III, TE & Kollman, PA. **Molecular dynamics simulation of nucleic acids.** *Annu. Rev. Phys. Chem.* 51:435–471, 2000.
- [66] Sinden, RR. **DNA Structure and Function.** Academic Press, 1994.

- [67] Sinden, RR. **DNA twists and flip.** *Nature.* 437:1097–1098, 2005.
- [68] Ha, SC; Lowenhaupt, K; Rich, A; Kim, YG & Kim, KK. **Crystal structure of a junction between B–DNA and Z–DNA reveals two extruded bases.** *Nature.* 437:1183–1186, 2005.
- [69] Rich, A & Zhang, S. **Z–DNA: the long road to biological function.** *Nature Reviews in Genetics.* 4:566–572, 2003.
- [70] Norberto de Souza, O & Ornstein, RL. **Inherent DNA curvature and flexibility correlate with TATA Box functionality.** *Biopolymers.* 46:403–415, 1998.
- [71] Campbell, MK. **Bioquímica.** Capítulo 7 – Ácidos nucleicos: de que modo a estrutura transfere informação. 3ª edição. Editora Artmed, Porto Alegre, 2000.
- [72] Altona, C & Sundaralingam, M. **Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation.** *Journal of American Chemical Society.* 94(23):8205–8212, 1972.
- [73] Bloomfield, VA; Crothers, DM & Tinoco Jr., I. **Nucleic acids – structures, properties, and functions.** University Science Books, Sausalito, California, EUA, 2000.
- [74] Obika, S; Reddy, SY & Bruice, TC. **Sequence–specific DNA binding of Ets–1 transcription factor: Molecular dynamics study on the ETS domain–DNA complexes.** *Journal of Molecular Biology.* 331:345–359, 2003.
- [75] Perez–Martín, J; Rojo, F & De Lorenzo V. **Promoters responsive to DNA bending: a common theme in prokaryotic gene expression.** *Microbiological Reviews.* 58:268–290, 1994.

[76] Starr, DB; Hoopes, BC & Hawley. **DNA bending is an important component of site-specific recognition by the TATA binding protein.** *Journal of Molecular Biology.* 250:434-446, 1995.

[77] Giudice E & Lavery, R. **Simulations of nucleic acids and their complexes.** *Acc. Chem. Res.* 35:350-357, 2002.

[78] Reddy, SY; Obika, S & Bruice, TC. **Conformations and dynamics of Ets-1 ETS domain-DNA complexes.** *PNAS.* 100(26):15475-15480, 2003.

[79] Strahs, D; Barrash, D, Qian, XL & Schlick, T. **Sequence-dependent solution structure and motions of 13 TATA/TBP (TATA-box binding protein) complexes.** *Biopolymers.* 69:216-243, 2003.

[80] Qian, XL, Strahs, D & Schlick, T. **Dynamic simulation of 13 TATA variants refine kinetic hypotheses of sequence/activity relationships.** *Journal of Molecular Biology.* 308:681-703, 2001.

[81] Case, DA; Pearlman, DA; Caldwell, JW; Cheatham III, TE; Wang, J; Ross, WS; Simmerling, CL; Darden, TA; Merz, KM; Stanton, RV; Cheng, AL; Vincent, JJ; Crowley, M; Tsui, V; Gohlke, H; Radmer, RJ; Duan, Y; Pitera, J; Massova, I; Seibel, GL; Singh, UC; Weiner, PK and Kollman, PA. **AMBER 7.** University of California, San Francisco, 2002.

[82] Jorgensen, WL; Chandrasekhar, J & Madura, JD. **Comparison of simple potential functions for simulating liquid water.** *J. Chme. Phys.* 79(2):926-935, 1983.

[83] Tinoco, I; Sauer, K; Wang JC & Puglisi, JD. **Physical Chemistry - Principles and applications in biological sciences.** Ed. Prentice Hall, 4<sup>a</sup> edição, 2002.

[84] Norberto de Souza, O & Ornstein, RL. **Effect of periodic Box size on aqueous molecular dynamics simulation of a DNA dodecamer with Particle-Mesh Ewald method.** *Biophysical Journal.* 72:2395-2397, 1997.



[85] van Gunsteren, WF & Berendsen, HJC. **Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry.** *Angew. Chem. Int. Ed. Engl.*, 29: 992–1023, 1990.

[86] Norberto de Souza, O & Ornstein, RL **Molecular dynamics simulations of a protein–protein dimer: particle–mesh Ewald electrostatic model yields far superior results to standard cutoff model.** *Journal of Biomolecular Structure & Dynamics*, 16: 1205–1217, 1999.

[87] Cornell, WD; Cieplak, P; Bayly, CI; Gould, IR; Merz Jr, KM; Ferguson, DM; Spellmeyer, DC; Fox, T; Caldwell, JW; Kollman, PA. **A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.** *J. Am. Chem. Soc.* 117:5179–5197, 1995.

[88] van Gunsteren, WF & Mark, AE. **On the interpretation of biochemical data by molecular dynamics computer simulation.** *Eur. J. Biochem.*, 204:947–961, 1992.

[89] Schroeder, EK; Basso, LA; Santos, DS & Norberto de Souza, O. **Molecular dynamics simulation studies of the wild-type, I21V, and I16T mutants of Isoniazid-resistant *Mycobacterium tuberculosis* enoyl reductase (InhA) in complex with NADH: toward the understanding of NADH–InhA different Affinities.** *Biophysical Journal*. 89:876–884, 2005.

[90] Van Gunsteren, WF & Berendsen, HJC. **Algorithms for macromolecular dynamics and constraint dynamics.** *Molecular Physics*. 34:1311–1327, 1977.

[91] Darden, T; York, D & Pedersen, L. **Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.** *J. Chem. Phys.* 98(12):10089–10092, 1993.

[92] Essmann, U; Perera, L; Berkowitz, ML; Darden, T; Lee, H & Pedersen, LG. **A smooth particle mesh Ewald method.** *J. Chem. Phys.* 103(19):8577–8593, 1995.

- [93] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. **Molecular dynamics with coupling to an external bath.** *J. Chem. Phys.*, 81 (8):3684–3690, 1984.
- [94] Guex, N & Peitsch, MC. **SWISS\_MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling.** *Electrophoresis*. 18:2714–2723, 1997.
- [95] DeLano, WL. **The PyMOL Molecular Graphics System.** DeLano Scientific, San Carlos, CA, USA, 2002. [<http://www.pymol.org>]
- [96] Humphrey, W; Dalke, A & Schulten, K. **VMD – Visual Molecular Dynamics.** *J. Mol. Graphics*. 14, 33–38, 1996.
- [97] Lavery, R & Sklenar, H. **The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids.** *Journal of Biomolecular Structure and Dynamics*. 6(1):63–91, 1988.
- [98] Lavery, R & Sklenar, H. **Defining the structure of irregular nucleic acids: conventions and principles.** *J. Biomol. Struct. Dynam.* 6:655–667, 1989.
- [99] Norberto de Souza, O. **Intrinsic curvature of DNA: a molecular dynamics study.** Tese de Doutorado. University of London, 220 p. 1994.
- [100] Norberto de Souza, O & Goodfellow, JM. **The intrinsic curvature of a 51 bp K-DNA fragment of *Leishmania tarentolae*: a molecular model.** *Journal of Biomolecular Structure and Dynamics*. 15:905–930, 1998.
- [101] Lu, XJ & Olson WK. **Resolving the discrepancies among nucleic acid conformational analyses.** *Journal of Molecular Biology*. 285:1563–1575, 1999.
- [102] Origin® 6.1, 2000. OriginLab Corporation. <http://www.originlab.com>

[103] Luscombe, NM; Laskowski, RA & Thornton, JM. **NUCPLOT: a program to generate schematic diagrams of protein–nucleic acid interactions.** *Nucleic Acids Research*, 25(24):4940–4945, 1997.

[104] Kumar, S; Tamura, K; & Nei M. **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Briefings in Bioinformatics*. 5:150–163, 2004.

[105] Benson, DA; Karsch–Mizrachi, I; Lipman, DJ; Ostell, J & Wheeler, DL. **GenBank.** *Nucleic Acids Research*. 34:D16–D20, 2006.

[106] Kabash, W & Sander, C. **Dictionary of protein secondary structure.** *Biopolymers*. 22:2577–2637, 1983.

[107] Thompson, JD; Higgins, DG & Gibson, TJ. **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position–specific gap penalties and weight matrix choice.** *Nucleic Acids Res*. 22(22): 4673–4680, 1994.

[108] Thompson, JD; Gibson, TJ; Plewniak, F; Jeanmougin, F & Higgins, DG. **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Research*, 25:4876–4882, 1997.

[109] Page, RDM. **TreeView: An application to display phylogenetic trees on personal computers.** *Computer Applications in the Biosciences*. 12:1423–1425, 1996.

[110] Altschul, SF; Gish, W; Miller, W; Myers, EW & Lipman, DJ. **Basic local alignment search tool.** *Journal of Molecular Biology*. 215:403–410, 1990.

[111] Laudet, V; Hänni, C; Stéhelin, D & Duterrque–Coquillard, M. **Molecular phylogeny of the ETS gene.** *Oncogene*. 18:1351–1359, 1999.

[112] Garvie, CW & Wolberger, C. **Recognition of specific DNA sequences.** *Molecular Cell*. 8:937–946, 2001.

- [113] Jayaram, B & Jain, T. **The role of water in protein–DNA recognition.** *Annu. Rev. Biophys. Biomol. Struct.* 33:343–361, 2004.
- [114] Kono, H & Sarai, A. **Structure–based prediction of DNA target sites by regulatory proteins.** *Proteins.* 35:114–131, 1999.
- [115] Sarai, A & Kono, H. **Protein–DNA recognition patterns and predictions.** *Annu. Rev. Biophys. Biomol. Struct.* 34:379–398, 2005.
- [116] Pabo, CO & Nekludova, L. **Geometric analysis and comparison of protein–DNA interfaces: Why is there no simple code for recognition?** *Journal of Molecular Biology.* 301:597–624, 2000.
- [117] Olson, WK; Gorin, AA; Lu, XJ; Hock, LM & Zhurkin, VB. **DNA sequence–dependent deformability deduced from protein–DNA crystal complexes.** *Proc. Natl. Acad. Sci. USA.* 95:11163–11168, 1998.
- [118] Cheatham III, TE. **Simulation and modeling of the nucleic acid structure, dynamics and interactions.** *Current Opinion in Structural Biology.* 14:360–367, 2004.
- [119] Norberg, J & Nilsson, L. **Molecular dynamics applied to nucleic acids.** *Acc. Chem. Res.* 35:465–472, 2002.
- [120] Packer, MJ; Dauncey, MP & Hunter, CA. **Sequence–dependent DNA structure: dinucleotide conformational maps.** *Journal of Molecular Biology.* 295:71–83, 2000.
- [121] Packer, MJ; Dauncey, MP & Hunter, CA. **Sequence–dependent DNA structure: tetranucleotide conformational maps.** *Journal of Molecular Biology.* 295:85–103, 2000.
- [122] Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. Biochemical nomenclature and related documents ® IUPAC/IUBMB, 2<sup>a</sup> Ed. Portland Press; páginas 115–121, 1982.

[123] Nomenclature of organic chemistry, section E: stereochemistry. Recommendations 1974. IUPAC Commission on Nomenclature of Organic Chemistry (CNOC). *Pure Appl. Chem.*, 45:11–30, 1976.

[124] Várnai, P; Djuranovic, D; Lavery, R & Hartmann, B.  **$\alpha/\gamma$  transitions in the B-DNA backbone.** *Nucleic Acids Research*. 30(24)5398–5406, 2002.

[125] Fratini, AV; Kopka, ML; Drew, HR & Dickerson, RE. **Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATT<sup>Br</sup>CGCG.** *The Journal of Biological Chemistry*. 257:14686–14707, 1982.

[126] Gorin, AA; Zhurkin, VB & Olson, WK. **B-DNA twisting correlates with base-pair morphology.** *Journal of Molecular Biology*. 247:34–48, 1995.

[127] Cheatham III, TE & Kollman, PA. **Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution.** *Journal of Molecular Biology*. 259:434–444, 1996.

[128] Lu, XJ; Shakked, Z & Olson, WK. **A-form conformational motifs in ligand-bound DNA structures.** *Journal of Molecular Biology*. 300:819–840, 2000.

[129] Crothers, DM. **DNA curvature and deformation in protein-DNA complexes: A step in the right direction.** *Proceedings of National Academy of Sciences USA*. 95:15163–15165, 1998.

[130] Pearson, H. **Genetic information: Codes and enigmas.** *Nature*. 444:259–261, 2006.

[131] Beveridge, DL & McConnell, KJ. **Nucleic acids: theory and computer simulation.** *Current Opinion in Structural Biology*. 10:182–196, 2000.

[132] Ponomarev, SY; Thayer, KM & Beveridge, DL. **Ion motions in molecular dynamics simulations on DNA.** *Proceedings of National Academy of Science*. 101(41):14771–14775, 2004.

[133] Beveridge, DL; Barreiro, G; Byun, KS; Case, DA; Cheatham III, TE; Dixit, SB; Giudice, E; Lankas, F; Lavery, R; Maddocks, JH; Osman, R; Seibert, E; Sklenar, H; Stoll, G; Thayer, KM; Varnai, P & Young, MA. **Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps.** *Biophysical Journal*. 87:3799–3813, 2004.

[134] Dixit, SB; Beveridge, DL; Case, DA; Cheatham III, TE; Giudice, E; Lankas, F; Lavery, R; Maddocks, JH; Osman, R; Sklenar, H; Thayer, KM & Varnai, P. **Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II. Sequence context effects on the dynamical structure of the 10 unique dinucleotide steps.** *Biophysical Journal*. 89:3721–3740, 2005.

[135] Tsui, V; Radhakrishnan, I; Wright, PE & Case, DA. **NMR and molecular dynamics studies of the hydration of a zinc finger–DNA complex.** *Journal of Molecular Biology*. 302:1101–1117, 2000.

[136] Sen, S & Nilsson, L. **Structure, interaction, dynamics and solvent effects on the DNA–EcoRI complex in aqueous solution from molecular dynamics simulation.** *Biophysical Journal*. 77:1782–1800, 1999.

[137] Suenaga, A; Yatsu, C; Komeiji, Y; Uebayasi, M; Meguri, T & Yamato, I. **Molecular dynamics simulation of trp–repressor/operator complex: analysis of hydrogen bond patterns of protein–DNA interaction.** *Journal of Molecular Structure*. 526:209–218, 2000.

## - Anexo I -

Arquivo de entrada para o módulo nucgen do pacote de programas AMBER7 [81] para a montagem dos sistemas a partir da suas seqüências de nucleotídeos:

### Sistema 02

Conformação canônica DNA B

**nuc.in:**

```
    NUC    1
D
T5  A    G    C    A    G    G    A    T    G    T    G    T3

    NUC    2
D
A5  C    A    C    A    T    C    C    T    G    C    T    A3

END
$ABDNA
```

Conformação canônica DNA A

**nuc.in:**

```
    NUC    1
D
T5  A    G    C    A    G    G    A    T    G    T    G    T3

    NUC    2
D
A5  C    A    C    A    T    C    C    T    G    C    T    A3

END
$ADNA
```

**Sistema 03:**

**Conformação canônica DNA B**

**nuc.in:**

```
      NUC   1
D
T5   A   G   C   A   G   G   A   G   G   T   G   T3

      NUC   2
D
A5   C   A   C   C   T   C   C   T   G   C   T   A3

END
$ABDNA
```

**Conformação canônica DNA A**

**nuc.in:**

```
      NUC   1
D
T5   A   G   C   A   G   G   A   G   G   T   G   T3

      NUC   2
D
A5   C   A   C   C   T   C   C   T   G   C   T   A3

END
$ADNA
```

**Sistema 04:**

**Conformação canônica DNA B**

**nuc.in:**

```
      NUC   1
D
T5   A   G   C   A   G   G   A   A   G   T   G   T3

      NUC   2
D
A5   C   A   C   T   T   C   C   T   G   C   T   A3

END
```



\$ABDNA

## Conformação canônica DNA A

**nuc.in:**

```
      NUC   1  
D  
T5  A   G   C   A   G   G   A   A   G   T   G   T3
```

```
      NUC   2  
D  
A5  C   A   C   T   T   C   C   T   G   C   T   A3
```

```
END  
$ADNA
```

## - Anexo II -

Arquivos de entrada para o módulo SANDER do pacote de programas AMBER7 [81] para as simulações pela dinâmica molecular<sup>1</sup>:

### mini\_00.in:

```
# Belly Minimization.NTF=NTC=2.
&cntrl
  imin=1,
  ntp=10,
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=1, ntr=0, cut=9.0,
  ntc=2, tol=0.00001,
  maxcyc=100, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,
  drms=0.0001,
&end
&ewald
&end
Belly dynamics. Solvent and ions atoms allowed to move
ATOM 825 18938
END
END
```

### mdcp\_00.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating solvent (WAT + Na+) at 298.16K.
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  ntxo=1, ntp=50,
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0,
  ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=1, ntr=0, cut=9.0,
  nstlim=10000, nscm=0, t=0.0, dt=0.002,
  temp0=298.16, tempi=10.0, ig=555333, heat=0.0001, ntt=1,
  dtemp=10.0, tautp=1.0, vlimit=10.0,
```

---

<sup>1</sup> Foram utilizados os valores *default* para as variáveis NBFLAG e SKINNB nos arquivos de entrada para o módulo SANDER, que determinam a atualização da lista das interações entre os átomos não ligados a cada vez que um átomo ultrapassasse em 1.0Å o raio de corte estabelecido (9Å).

```
    ntp=1, pres0=1.0, comp=44.6, taup=2.0,  
    ntc=2, tol=0.00001,  
  &end  
  &ewald  
  &end  
  Belly dynamics. Solvent and ions atoms allowed to move  
ATOM 825 18938  
END  
END
```

### mini\_01.in:

```
# Restrained Minimization.NTF=NTC=2.NTR=1.  
&cntrl  
  imin=1,  
  ntp=10,  
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,  
  ibelly=0, ntr=1, cut=9.0,  
  ntc=2, tol=0.00001,  
  maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,  
  drms=0.0001,  
  &end  
  &ewald  
  &end  
  Restraining DNA  
  1000.000  
ATOM 1 824  
END  
END
```

### mini\_02.in:

```
# Restrained Minimization.NTF=NTC=2.NTR=1.  
&cntrl  
  imin=1,  
  ntp=10,  
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,  
  ibelly=0, ntr=1, cut=9.0,  
  ntc=2, tol=0.00001,  
  maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,  
  drms=0.0001,  
  &end  
  &ewald  
  &end  
  Restraining DNA  
  500.000  
ATOM 1 824  
END  
END
```

### mini\_03.in:

```
# Restrained Minimization.NTF=NTC=2.NTR=1.
&cntrl
  imin=1,
  ntp=10,
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=1, cut=9.0,
  ntc=2, tol=0.00001,
  maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,
  drms=0.0001,
&end
&ewald
&end
Restraining DNA
250.000
ATOM 1 824
END
END
```

### mini\_04.in:

```
# Restrained Minimization.NTF=NTC=2.NTR=1.
&cntrl
  imin=1,
  ntp=10,
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=1, cut=9.0,
  ntc=2, tol=0.00001,
  maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,
  drms=0.0001,
&end
&ewald
&end
Restraining DNA
100.000
ATOM 1 824
END
END
```

### mini\_05.in:

```
# Restrained Minimization.NTF=NTC=2.NTR=1.
&cntrl
  imin=1,
  ntp=10,
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=1, cut=9.0,
  ntc=2, tol=0.00001,
  maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,
  drms=0.0001,
&end
```

```
&ewald
&end
Restraining DNA
50.000
ATOM 1 824
END
END
```

### mini\_06.in:

```
# Restrained Minimization.NTF=NTC=2.NTR=1.
&cntrl
  imin=1,
  ntp=10,
  ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=1, cut=9.0,
  ntc=2, tol=0.00001,
  maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,
  drms=0.0001,
&end
&ewald
&end
Restraining DNA
10.000
ATOM 1 824
END
END
```

### mini\_07.in:

```
# Unrestrained Minimization.NTF=NTC=2.
&cntrl  imin=1,
        ntp=10,
        ntf=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
        ibelly=0, ntr=0, cut=9.0,
        ntc=2, tol=0.00001,
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5,
        drms=0.0001,
&end
&ewald
&end
```

### mcdp\_01.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating the system from 10K to 50K.
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  ntso=1, ntp=50,
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0,
```

```
ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,  
ibelly=0, ntr=0, cut=9.0,  
nstlim=10000, nscm=0, t=0.0, dt=0.002,  
temp0=50.0, tempi=10.0, ig=555333, heat=0.0001, ntt=1,  
dtemp=10.0, tautp=1.0, vlimit=10.0,  
ntp=1, pres0=1.0, comp=44.6, taup=2.0,  
ntc=2, tol=0.00001,  
&end  
&ewald  
&end
```

### mcdp\_02.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating the system from 50K to 100K.  
&cntrl  
  imin=0, ntx=1, irect=0, ntrx=1,  
  ntxo=1, ntp=50,  
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,  
  ioutfm=0, ntwprt=0,  
  ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,  
  ibelly=0, ntr=0, cut=9.0,  
  nstlim=10000, nscm=0, t=0.0, dt=0.002,  
  temp0=100.0, tempi=50.0, ig=555333, heat=0.0001, ntt=1,  
  dtemp=10.0, tautp=0.2, vlimit=10.0,  
  ntp=1, pres0=1.0, comp=44.6, taup=0.2,  
  ntc=2, tol=0.00001,  
&end  
&ewald  
&end
```

### mcdp\_03.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating the system from 100K to 150K.  
&cntrl  
  imin=0, ntx=1, irect=0, ntrx=1,  
  ntxo=1, ntp=50,  
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,  
  ioutfm=0, ntwprt=0,  
  ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,  
  ibelly=0, ntr=0, cut=9.0,  
  nstlim=10000, nscm=0, t=0.0, dt=0.002,  
  temp0=150.0, tempi=100.0, ig=555333, heat=0.0001, ntt=1,  
  dtemp=10.0, tautp=0.2, vlimit=10.0,  
  ntp=1, pres0=1.0, comp=44.6, taup=0.2,  
  ntc=2, tol=0.00001,  
&end  
&ewald  
&end
```

### mcdp\_04.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating the system from 150K to 200K.
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  nt xo=1, ntp r=50,
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0,
  ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=0, cut=9.0,
  nstlim=10000, nscm=0, t=0.0, dt=0.002,
  temp0=200.0, tempi=150.0, ig=555333, heat=0.000001, ntt=1,
  dtemp=10.0, tautp=0.2, vlimit=10.0,
  ntp=1, pres0=1.0, comp=44.6, taup=0.2,
  ntc=2, tol=0.00001,
&end
&ewald
&end
```

### mcdp\_05.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating the system from 200K to 250K.
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  nt xo=1, ntp r=50,
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0,
  ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=0, cut=9.0,
  nstlim=10000, nscm=0, t=0.0, dt=0.002,
  temp0=250.0, tempi=200.0, ig=555333, heat=0.000001, ntt=1,
  dtemp=10.0, tautp=0.2, vlimit=10.0,
  ntp=1, pres0=1.0, comp=44.6, taup=0.2,
  ntc=2, tol=0.00001,
&end
&ewald
&end
```

### mcdp\_06.in:

```
# 20ps NPT. NTF+NTC=2. Equilibrating the system from 250K to 298.16K.
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  nt xo=1, ntp r=50,
  ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0,
  ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,
  ibelly=0, ntr=0, cut=9.0,
  nstlim=10000, nscm=0, t=0.0, dt=0.002,
  temp0=298.16, tempi=250.0, ig=555333, heat=0.000001, ntt=1,
  dtemp=10.0, tautp=0.2, vlimit=10.0,
  ntp=1, pres0=1.0, comp=44.6, taup=0.2,
```

```
    ntc=2, tol=0.00001,  
&end  
&ewald  
&end
```

### **mdcp\_prod\_01.in:**

```
# 100ps NTB=2. NTT=1. TAUTP=2.0. NTF+NTC=2.  
&cntrl  
    imin=0, ntx=5, irect=1, ntrx=1,  
    nt xo=1, ntp r=50,  
    ntwr=250, iwrap=0, ntwx=250, ntwv=0, ntwe=0,  
    ioutfm=0, ntwprt=0,  
    ntf=2, ntb=2, dielc=1.0, igb=0, scnb=2.0, scee=1.2,  
    ibelly=0, ntr=0, cut=9.0,  
    nstlim=50000, nscm=0, t=0.0, dt=0.002,  
    temp0=298.16, tempi=298.16, ig=555333, heat=0.0001, ntt=1,  
    dtemp=10.0, tautp=2.0, vlimit=10.0,  
    ntp=1, pres0=1.0, comp=44.6, taup=0.2,  
    ntc=2, tol=0.00001,  
&end  
&ewald  
&end
```