

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS - PPGBM

**Modelagem Molecular e Imunodeteção de DNA Metiltransferases
2 de Drosofilídeos: Uma Abordagem Evolutiva da Enigmática
DNMT2**

Aluno: Gilberto Cavalheiro Vieira

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de **Mestre em Genética e Biologia Molecular**.

Orientadora: Dra. Vera Lúcia S. Valente Gaiesky

Co-orientadora: Dra. Marialva Sinigaglia

Colaborador: Dr. Gustavo Fioravanti Vieira

Porto Alegre
Março de 2015

Instituições e Fontes Financiadoras

Esta dissertação foi desenvolvida no Laboratório de *Drosophila* e Núcleo de Bioinformática da Imunogenética (NBLI) do Departamento de Genética da Universidade Federal Do Rio Grande Do Sul (UFRGS), Porto Alegre, RS

Apoio CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - Processo PRONEX 10/0028-7), PROPESQ-UFRGS (Pró-Reitoria de Pesquisa da Universidade do Rio Grande do Sul) e PPGBM (Programa de Pós-Graduação em Genética e Biologia Molecular).

“A ignorância gera confiança com mais frequência do que o conhecimento: são aqueles que sabem pouco, e não aqueles que sabem muito, que tão positivamente afirmam que esse ou aquele problema jamais será resolvido pela ciência.”

Charles Darwin

AGRADECIMENTOS

Primeiramente gostaria de agradecer meus amados pais, minha mãe, Dagmar, pela força e perseverança atávica que lhe forjou uma alma inabalável diante dos desafios e meu amado pai, Rivadávia, por sua inigualável fonte de ardor pelo trabalho diário e sua sanha pelo saber. Meu irmão, pessoa de personalidade única, um ser humano permeado de profundo conhecimento e que, mesmo mais jovem do que eu, é um exemplo de retidão científica e de aprofundamento acadêmico, além de um companheiro inestimável.

É claro que não poderia deixar de agradecer à minha companheira (esposa, namorada, amiga, parceira, redentora...) de todos os momentos - bons e ruins -, Camila. “Amore mio”, só tenho a te agradecer pelo exemplo de pessoa que você é, tanto na vida pessoal como na profissional.

Não é segredo a minha gratidão à minha orientadora Vera Valente, que por esses sete anos de LabDros depositou sua preciosa confiança em meu trabalho, sempre acreditando e incentivando meu crescimento não só no âmbito acadêmico, mas no humano também. Obrigado por me mostrar que ciência se faz com paixão, também.

Agradeço imensamente a minha co-orientadora Meg, por seu companheirismo e revigorante incentivo, mesmo naqueles momentos de dúvidas que se abatem diante de um desafio. Em especial a duas grandes pessoas do Núcleo de Bioinformática da Imunogenética: meu colaborador Gustavo, pelas “*brainstorms*” a respeito do assunto, que sempre foram de grande valia, e ao Maurício pelas “*hints*” and “*cheats*” dos programas, sem as quais o trabalho se uma missão quase impossível.

À Marícia pela amizade e palavras de apoio nos momentos complicados e por ter me feito admirar a epigenética, Lamarck, Jablonka, drosófilas... (risos).

À Maríndia pelas sempre entusiásticas parcerias e ideias de projetos e trabalhos, sempre apoiando e dando uma luz quando parecia não haver solução

para um determinado problema, além de ter papel fundamental na “ontogenia” do presente mestrado.

Ao Elmo por sua inexorável cordialidade e atenção dada frente às dúvidas recorrentes sobre o PPGBM.

Ao grande Dr. Leo Meira, do Departamento de Bioquímica, pela parceria e amizade firmada ao longo do trabalho.

Por fim, meus mais sinceros agradecimentos aos amigos e colegas de laboratório, Carol, Ju WG, Rebeca, Dani, Berê, Lucas (Avante Imortal Tricolor), Brenda, Paula e Natália pelos satisfatórios momentos compartilhados.

Lista de Abreviaturas

1DCT – PDB do cristal da enzima DNMT2 de *Haemophilus influenzae*

1G55 – PDB do cristal da enzima DNMT2 humana

1MHT – PDB do cristal da enzima DNMT2 de *Haemophilus haemolyticus*

3QV2 – PDB do cristal da enzima DNMT2 de *Entamoeba histolytica*

4DKJ – PDB do cristal da enzima DNMT2 de *Mycoplasma penetrans*

4H0N – PDB do cristal da enzima DNMT2 de *Spodoptera frugiperda*

AdoMet – S-Adenosilmetionina

AIC – Critério de Informação de Akaike

APBS – Adaptive Poisson-Boltzmann Solver

BLAST – Basic Local Alignment Search Tool

BSA – Bovine serum albumin

DNMT – DNA methyltransferase

ECL – Enhanced chemiluminescence

EhmetH – Metiltransferase de *E. histolytica*

GTR – Modelo evolutivo *General Time Reversible*

HaeIII – Metiltransferase de *H. influenzae*

HhaI – Metiltransferase de *H. haemolyticus*

JTT – Modelo evolutivo Jones-Taylor-Thornton

LG – Modelo evolutivo Le Gascuel

M.Mpel – Metiltransferase de *M. penetrans*

m4C-MTase – DNA metiltransferase com atividade de metilação do N4 de citosinas

m5C-MTase – DNA metiltransferase com atividade de metilação do C5 de citosinas

m6A-MTase – DNA metiltransferase com atividade de metilação do N6 de adeninas

MDBs – Proteínas de ligação ao DNA metilado

mIT, wIT, MmIT, HsIT – Modelos de *Drosophila melanogaster*, *D. willistoni*, *Mus musculus* e *Homo sapiens* com a ferramenta I-TASSER

mMD, wMD, MmMD, HsMD – Modelos de *Drosophila melanogaster*, *D. willistoni*, *Mus musculus* e *Homo sapiens* com a ferramenta MODELLER

mRB, wRB, MmRB, HsRB – Modelos de *Drosophila melanogaster*, *D. willistoni*, *Mus musculus* e *Homo sapiens* com a ferramenta ROBETTA

MSRE – Endonucleases de Restrição Sensíveis à Metilação

mSW, wSW, MmSW, HsSW - Modelos de *Drosophila melanogaster*, *D. willistoni*, *Mus musculus* e *Homo sapiens* com a ferramenta SWISS-MODEL

MTases – Metiltransferases

PCNA – Antígeno Nuclear de Proliferação Celular

PDB – Banco de dados de proteínas

PIPSA – Análises de Similaridades de Propriedades das Interações de Proteína

ProCys – Dipeptídeo Prolina Cisteína

RMN – Ressonância Magnética Nuclear

RMSD – Desvio da raiz quadrada média

RMSF – Flutuação da raiz quadrada média

sfDNMT2 – Metiltransferase de *S. frugiperda*

TBS – Solução salina tamponada com Tris

TBS-T – Solução salina tamponada com Tris e Triton

TEs – Elementos Transponíveis

TRD – Região de reconhecimento da sequência-alvo a ser metilada

UniProt – Banco de dados *Universal Protein Resource*

Lista de Figuras

- Figura 1.** Representação esquemática do mecanismo de metilação de citosina por meio de DNMT. Num primeiro momento há um ataque nucleofílico na posição 6 da citosina por parte da enzima (Cys81 em *MHal*, por exemplo). Esse ataque acaba por ativar a posição 5 para um ataque eletrofílico pelo cofator (AdoMet), quando o grupamento metil é transferido para a citosina.27
- Figura 2.** Ilustração esquemática de DNMTs conhecidas de vertebrados. A região em preto representa o domínio catalítico C-terminal conservado (imagem adaptada de Mund *et al.*, 2004).28
- Figura 3.** Conservação evolutiva de DNMT2 em diversos reinos. Destacados com asterisco vermelho os grupos onde a DNMT2 é a única metiltransferase até o momento descrita (imagem adaptada de Schaefer & Lyko, 2010).31
- Figura 4.** Níveis de metilação de citosinas ao longo dos estágios de desenvolvimento em *Drosophila melanogaster* (imagem modificada de Lyko *et al.*, 2000).34
- Figura 5.** Sequências proteicas de DNMT2 humana (391 aminoácidos) e *M. musculus* (415 aminoácidos) traduzida a partir da junção dos éxons. Destacado em cinza os resíduos pertencentes às regiões não resolvidas na estrutura cristalográfica da DNMT2 humana (PDB 1G55).46
- Figura 6.** Sequências proteicas de DNMT2 de *D. melanogaster* (345 aminoácidos) e *D. willistoni* (341 aminoácidos) traduzida a partir da junção dos éxons.47
- Figura 7.** Fluxograma apresentando as etapas desenvolvidas no presente trabalho.48
- Figura 8.** Estruturas secundárias preditas a partir das sequências de DNMT2 de *D. melanogaster* (esquerda) e *D. willistoni* (direita) provenientes da ferramenta PSIPRED. Representação: Folha β (setas amarelas), hélices α (cilindros rosa) e alças (linha preta). Barras na parte superior das estruturas preditas indicam o grau de confiança da predição.59

Figura 9. Representação da estrutura secundária predita pelo programa PSIPRED, a partir da sequência de DNMT2 de humanos (esquerda) e de *M. musculus* (direita). Folha β (setas amarela), hélices α (cilindros rosa) e alças (linha preta). Acima da estrutura predita encontra-se o grau de confiança da predição.

.....60

Figura 10. Representação em *ribbon* das estruturas 3D de DNMT2. Estruturas alfa-hélices (vermelho), folhas-beta (amarelo) e alças (verde). A alça do domínio catalítico está colorida em cinza e os átomos do dipeptídeo ProCys do sítio ativo representados como esferas (carbonos em cinza, enxofre em amarelo e nitrogênio em azul). **(A)** *H. haemolyticus* (PDB 1MHT); **(B)** *H. influenzae* (PDB 1DCT); **(C)** *M. penetrans* (PDB 4DKJ); **(D)** *E. histolytica* (PDB 3QV2); **(E)** *D. willistoni* (wRB1); **(F)** *D. melanogaster* (mMD2), em ciano a alça contendo a inserção dos resíduos 267-281; **(G)** *S. frugiperda* (PDB 4H0N); **(H)** *M. musculus* (MmSW); e **(I)** Humanos (HsRB1), em ciano região não conservada referente aos resíduos 189 – 247 presentes apenas em DNMT2 de mamíferos.64

Figura 11. Sequência de oito motivos conservados presentes no domínio catalítico das estruturas analisadas.....66

Figura 12. Sobreposição do domínio catalítico das nove DNMT2 analisadas. Cisteína catalítica colorida em verde, o motivo conservado ENV e a arginina do motivo V, que são cruciais para a atividade catalítica pra enzima, estão coloridos em vermelho. Os demais resíduos conservados e/ou envolvidos com interações diretas tanto com o cofator, como o substrato, estão coloridos em ciano. O cofator SAM está colorido em azul.67

Figura 13. Representação das estruturas terciárias das TRDs das DNMT2 analisadas. Os resíduos descritos como aqueles que interagem direta ou indiretamente com o substrato (DNA/tRNA), tendo a função de reconhecimento da sequência alvo a ser metilada, estão representados como esferas vermelhas.... 69

Figura 14. Distribuição das frequências de aminoácidos quanto às suas propriedades nos modelos de DNMT2. Todas as frequências estão expressas em percentual.70

Figura 15. Distribuição das frequências de aminoácidos das DNMTs. Todas as frequências expressas em percentual.....71

Figura 16. Representação do potencial eletrostático de superfície das estruturas de DNMT2. As superfícies estão coloridas de acordo com o potencial eletrostático, em azul para potencial positivo ($5 kT$), vermelho para negativo ($-5 kT$) e branco para neutro, onde k representa a constante de Boltzmann e T , a temperatura. As estruturas de DNMT2 humana e de *M. musculus* estão representadas sem a alça de ligação dos domínios I e II (resíduos 189 – 247) das MTase2 de mamíferos. A TRD (seta verde) localiza-se na região mais central da molécula, com característica predominantemente de potencial eletrostático positivo.....74

Figura 17. Epograma da análise por PIPSA (*Protein Interaction Property Similarity Analysis*). As estruturas estão agrupadas segundo a relação de similaridade dos potenciais eletrostáticos a partir da matriz de distância elaborada pelo programa. Acima à esquerda encontra-se a chave de cores, indicando a relação de similaridade, quanto menor o valor, maior similaridade entre as estruturas.....76

Figura 18. Análise filogenética pelo método de máxima verossimilhança das sequências de aminoácidos. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas, sendo tomada como representativa da história evolutiva dos taxos analisados. Os ramos que apresentaram *bootstrap* menor do que 75 foram condensados. O modelo evolutivo utilizado foi o de Le Gascuel (LG) com distribuição gama discreta com 4 categorias. (A) árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; (B) árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2.....78

Figura 19. Análise filogenética pelo método de máxima verossimilhança das sequências nucleotídicas. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas. Os ramos que apresentaram *bootstrap* menor do que 75 foram condensados. O modelo evolutivo utilizado foi o de Le Gascuel (LG) com distribuição gama discreta com 4 categorias. (A) árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; (B) árvore consenso

utilizando informações de *fitting* das estruturas 3D das DNMT2..... 79

Figura 20. Inferência filogenética bayesiana das DNMT2 das sequências de aminoácidos. A árvore foi gerada utilizando o modelo evolutivo JTT com distribuição gama. Os ramos que apresentaram *bootstrap* menor do que 75 foram condensados. **(A)** árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; **(B)** árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2. 80

Figura 21. Inferência filogenética bayesiana das sequências nucleotídicas. Os ramos que apresentaram *bootstrap* menor do que 75 foram condensados. O modelo evolutivo utilizado foi o GTR (*general time reversible*) com taxa de sítios invariáveis (I) e distribuição gama (G). **(A)** árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; **(B)** árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2..... 80

Figura 22. Desvio médio quadrático (RMSD) com todos os átomos das estruturas 3D de DNMT2 durante a simulação de dinâmica molecular, no intervalo de 2,5ns à 50 ns – que corresponde ao tempo de término da fase de minimização de energia e termalização das estruturas e início da fase de produção..... 83

Figura 23. Representação *ribbon* das estruturas DNMT2 de *D. melanogaster* e humana. Os diferentes domínios estão representados em diferentes cores: domínio catalítico (verde), alça contendo cisteína catalítica (vermelho), domínio II contendo a TRD (margenta), alça de ligação dos domínios (cinza), segmento não resolvido na estrutura cristalográfica 1G55, DNMT2 humana (ciano), motivo conservado CFT (amarelo)..... 84

Figura 24. Comparação entre o RMSD dos domínios catalíticos e TRDs das DNMT2, no tempo de 2,5 ns à 50 ns. Linha contínua representando o RMSD do domínio catalítico e linha pontilhada, TRD. Região de preenchimento entre os gráficos destacando as diferenças entre os dois domínios. Na análise não foi levado em conta a alça de ligação dos domínios catalíticos e TRDs..... 86

Figura 25. RMSF dos modelos wRB1 (*D. williston*), mMD2 (*D. melanogaster*),

MmSW (*M. musculus*) e HsRB1 (humana) durante a simulação de dinâmica molecular. As estruturas secundárias das estruturas iniciais de cada modelo estão representadas abaixo de cada gráfico correspondente e coloridas de acordo com as representações da figura 24..... 87

Figura 26. Análise de raio de giro das regiões catalíticas e TRDs das DNMT2 durante a simulação de dinâmica molecular, no intervalo de 2,5ns à 50 ns..... 89

Figura 27. Epograma da análise por PIPSA (*Protein Interaction Property Similarity Analysis*) das 23 DNMT2 de drosofilídeos. Os modelos mMD2 e wRB1 destacados em vermelho. As estruturas estão agrupadas segundo a relação de similaridade dos potenciais eletrostáticos a partir da matriz de distância elaborada pelo programa. Acima à esquerda encontra-se a chave de cores, indicando a relação de similaridade, quanto menor o valor, maior similaridade entre as estruturas. 94

Figura 28. Potencial eletrostático de superfície das estruturas de DNMT2 de representantes de diferentes grupos de drosofilídeos. As superfícies estão coloridas de acordo com o potencial eletrostático, em azul para potencial positivo ($5 kT$), vermelho para negativo ($-5 kT$) e branco para neutro, onde k representa a constante de Boltzmann e T , a temperatura..... 95

Figura 29. Análise filogenética pelo método de máxima verossimilhança das sequências de aminoácidos de DNMT2 de diferentes espécies de drosofilídeos. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas. O modelo evolutivo utilizado foi o de Le Gascuel (LG) com distribuição gama discreta com 4 categorias..... 99

Figura 30. Inferência filogenética bayesiana das sequências de aminoácidos de DNMT2 das 22 espécies de drosofilídeos com genoma depositado no banco de dados Flybase, mais *D. tropicalis*. A árvore foi gerada utilizando o modelo evolutivo JTT com distribuição gama..... 100

Figura 31. Análise filogenética pelo método de máxima verossimilhança das sequências de nucleotídeos de *Dnmt2*. A árvore consenso foi inferida a partir de

um *bootstrap* de 10.000 réplicas. O modelo evolutivo utilizado foi o *General Time Reversible* (GTR) com distribuição gama discreta e alguns sítios invariáveis.....101

Figura 32. Inferência filogenética bayesiana das sequências de nucleotídeos de *Dnmt2* das 22 espécies de drosofilídeos com genoma depositado no banco de dados Flybase, mais *D. tropicalis*. A árvore foi gerada utilizando o modelo evolutivo *General Time Reversible* (GTR) com distribuição gama discreta e alguns sítios invariáveis.....102

Figura 33. Sequência completa da *Dnmt2* de humanos com a sequência imunógena em destaque.104

Figura 34. Padrões de expressão da DNMT2 em espécies do grupo *willistoni* de *Drosophila* na fase de embrião. Marcador indicando o peso molecular esperado para DNMT2 de drosofilídeos (entre 40 e 48 kDa, aproximadamente).....105

Figura 35. Padrões de expressão da DNMT2 em espécies do grupo *willistoni* de *Drosophila* na fase de pupa. Marcador indicando o peso molecular esperado para DNMT2 de drosofilídeos (entre 40 e 48 kDa, aproximadamente).....106

Lista de Tabelas

- Tabela 1.** Modelos cristalográficos de DNMT2 depositados no banco de dados Protein Data Bank utilizados para análise e escolha da estrutura molde para o processo de modelagem por homologia e estudos evolutivos..... 57
- Tabela 2.** Resultados das validações dos modelos obtidos das estruturas de DNMT2. A primeira linha indica as ferramentas utilizadas e a segunda os principais parâmetros analisados. NP – Não permitido; GP – Generosamente permitido; AP – Adicionalmente permitido. Em cinza: modelos selecionados como preditivos. Modelos que não foram aprovados no Verify 3D não foram validados (N.A.)..... 62
- Tabela 3.** RMSD referente comparação pareada dos alinhamentos estruturais dos modelos de Dnmt2. Diagonal inferior RMSD da proteína total. Diagonal superior, RMSD do domínio catalítico..... 73
- Tabela 4.** Valores de RMSD as estruturas entre os períodos de simulação de dinâmica molecular de 5 ns à 50 ns..... 84
- Tabela 5.** Sequências homólogas de Dnmt2 obtidas através da ferramenta BLASTn. A sequência de Dnmt2 de *D. melanogaster* foi utilizada como *query*....90
- Tabela 6.** Estimativa da divergência evolutiva dos aminoácidos entre grupos (diagonal inferior) e de nucleotídeos (diagonal superior). Todas as posições que continham *gap* foram eliminadas da análise..... 96
- Tabela 7.** Estimativa de divergência evolutiva das sequências de aminoácidos (diagonal inferior) e das sequências de nucleotídeos (diagonal superior). Todas as posições que continham *gaps* foram eliminadas da análise. Maiores divergências destacadas em vermelho e menores, em azul..... 97
- Tabela 8.** Desvio de uso de códons da sequência codificante de Dnmt2 de *D. melanogaster* em comparação às 22 espécies de drosofilídeos analisadas. O asterisco indica preferência de códon idêntica entre as espécies. As menores relações encontradas no uso de códons estão destacadas em negrito..... 98

Sumário

RESUMO	19
ABSTRACT	21
CAPÍTULO I - Epigenética e o Fenômeno de Metilação	
1. Introdução Geral	24
1.1.1 Epigenética	24
1.1.2 O fenômeno de metilação	25
1.1.3 As metiltransferases	26
1.1.4 A DNMT1: mantenedora das marcações epigenéticas	28
1.1.5 DNMT3a e DNMT3b: metilação <i>de novo</i>	30
1.1.6 A DNMT2: sua ancestralidade comum e sua ampla distribuição	31
1.1.7 O fenômeno de metilação em drosofilídeos	34
1.1.8 Sobre a evolução das proteínas: estrutura e função	36
1.1.9 Bioinformática Estrutural: um campo interdisciplinar para resolução de estruturas e sistemas de proteínas alvo	37
1.1.10 Modelagem comparativa por homologia: MODELLER e SWISS-MODEL	38
1.1.11 Modelagem molecular por métodos mistos: I-TASSER e Robetta	40
1.1.12 Validação dos modelos	41
1.2. Justificativa	42
1.3. Objetivos	43

1.4 Objetivo geral	43
1.5 Objetivos específicos	43

CAPÍTULO II Modelagem Molecular & Análises Evolutivas

2.1 Metodologia	46
2.1.1 Predição e resolução das estruturas tridimensionais de DNMT2 humana, <i>Mus musculus</i> , <i>Drosophila melanogaster</i> e <i>Drosophila willistoni</i>	46
2.1.1.1 Modelos DNMT2 I-TASSER	49
2.1.1.2 Modelos DNMT2 MODELLER	49
2.1.1.3 Modelos DNMT2 SWISS-MODEL	50
2.1.1.4 Modelos DNMT2 ROBETTA	51
2.1.1.5 Validação dos modelos	51
2.1.1.6 Simulação de Dinâmica Molecular	51
2.1.1.7 Análise dos dados da Dinâmica Molecular	53
2.1.1.8 Análise do potencial eletrostático de superfície.....	53
2.1.1.9 Análises das relações evolutivas estruturais de DNMT2	54
2.2 Resultados e Discussão	56
2.2.1 Busca de estruturas cristalográficas	56
2.2.2 Validações e seleção dos modelos desenvolvidos	61
2.2.3 Análises estruturais, físico-químicas e evolutivas das DNMT2	65
2.2.4 Dinâmica molecular: Análises do comportamento das DNMT2 em meio fisiológico	81

CAPÍTULO III - Modelagem Molecular por Homologia & Análises Evolutivas das DNMT2 de drosofilídeos

3.1 Metodologia	91
3.1.1 Busca de sequências homólogas no banco de dados Flybase e modelagem molecular por homologia das DNMT2 de drosofilídeos	91
3.2 Resultados e Discussão	92
3.2.1 As DNMT2 de drosofilídeos são conservadas em sequência e estrutura, mas divergem em suas propriedades	93

CAPÍTULO IV - Imunodeteção da enzima DNMT2 em drosofilídeos do subgrupo willistoni

4.1 Metodologia	104
4.1.2 Western blot	104
4.2 Resultados e Discussão	105

CAPÍTULO V

5.1 Discussão Geral	109
5.2 Conclusão	116
5.3 Perspectivas	118
5.4 Referências	119

RESUMO

A metilação do DNA genômico é um dos principais mecanismos de regulação epigenética nos organismos. Dentre as diferentes classes de DNA MTase, as m5C-MTase são as que se distribuem amplamente de procariotos a eucariotos. Em vertebrados existem três diferentes famílias: DNMT1, DNMT2 e DNMT3a e 3b. A DNMT1 possui atividade junto ao DNA hemimetilado. As DNMT3a e 3b são responsáveis pela metilação *de novo*. Já a subfamília DNMT2 possui seus sítios catalíticos altamente conservados, desde procariotos até eucariotos, possuindo propriedades que permitem executar funções tanto de metilação *de novo*, assim como de manutenção de metilação. Além disso, as enzimas da família DNMT2 podem atuar metilando citosinas genômicas ou de tRNAs. Em mamíferos, invertebrados e plantas a DNMT2 é classificada, prioritariamente *in vivo* como uma tRNA MTase. Entretanto, já foi descrita atividade de DNA MTase por parte dessas enzimas, mesmo que em baixos níveis. O que se discute são as atividades preferenciais da DNMT2 e os mecanismos que modulam sua atividade, pois se reconhece que em organismos que não possuem as MTases canônicas (DNMT1 e DNMT3), mas apresentam metilação em seu genoma, é a DNMT2 que atua como MTase em ambos os substratos. Espécies de *Drosophila* são conhecidas como de *Dnmt2-only*, justamente por possuírem apenas a DNMT2 na função de metilação de citosinas. A importância de seu papel no âmbito ecológico e evolutivo nesse grupo de espécies se reflete na presença de fenômenos peculiares, como a metilação sexo-específica presente em espécies do subgrupo *willistoni* de *Drosophila*, descrito por nosso grupo de pesquisa. No presente trabalho realizou-se a modelagem das enzimas DNMT2 de *D. melanogaster*, *D. willistoni* e *Mus musculus* com diferentes metodologias. A partir desses modelos realizaram-se análises comparativas com as estruturas cristalográficas de DNMT2 depositadas no banco de dados PDB, com o objetivo de estabelecer as relações evolutivas e funcionais entre as diferentes enzimas. Adicionalmente, de posse de modelos de DNMT2 - de validada qualidade - de duas espécies pertencentes a diferentes grupos evolutivos de drosofilídeos, realizaram-se estudos de

caracterização evolutiva e estrutural das 22 espécies que tiveram seus genomas sequenciados e depositados no bando de dados Flybase, somando-se a essas a sequência de DNMT2 de *Drosophila tropicalis* (subgrupo *willistoni*), sequenciado por nosso grupo de pesquisa. Os resultados das análises evolutivas e estruturais sugerem propriedades diferenciais entre as DNMT2 de espécies do subgrupo *willistoni* em relação às demais. Estes resultados indicam que mesmo em espécies que possuem relações evolutivas próximas possam ocorrer mecanismos adaptativos que estabeleçam gradações na afinidade das DNMT2 por diferentes substratos, sem que para isso ocorram drásticas mudanças na arquitetura da enzima.

ABSTRACT

The methylation of genomic DNA is a major mechanism of epigenetic regulation in organisms. Among the different classes of DNA MTase the M5C-MTase are as widely distributed in prokaryotes to eukaryotes. In vertebrates there are three different families: DNMT1, DNMT2 and DNMT3a and 3b. The DNMT1 has activity with the hemimethylated DNA. The DNMT3a and 3b are responsible for *de novo* methylation. While the DNMT2 subfamily has its catalytic sites highly conserved from prokaryotes to eukaryotes, having properties that allow performing functions of both *de novo* and maintenance methylation. Furthermore, the DNMT2 family can act methylating cytokines, tRNAs or DNA. In mammals, invertebrates and plants, DNMT2 is classified primarily *in vivo* as a tRNA MTase. However, DNA-MTase activity by these enzymes has been described, even at low levels. The question is the preferred activities of DNMT2 and mechanisms that modulate its activity, because in organisms that do not have the canonical DNA-MTases (DNMT1 and DNMT3), but have cytokines methylated in its genome, the DNMT2 acts as MTase on both substrates. *Drosophila* species are known as Dnmt2-only. The importance of DNMT2's role in ecological and evolutionary context in this species group is reflected by presence of a peculiar phenomenon: the sex-specific methylation described by our research group, presents in *willistoni* subgroup of *Drosophila*. In this study, DNMT2 of *D. melanogaster*, *D. willistoni* and *Mus musculus* were modeled by different methodologies. comparative analyzes with the crystallographic DNMT2 structures deposited in the PDB database were performed from these models, in order to establish the evolutionary and functional relationships between the different enzymes. Additionally, evolutionary and structural characterization studies of the 22 species, with the validated DNMT2 models of two species belonging to different evolutionary drosophilids groups, were conducted that have had their genomes sequenced and deposited in FlyBase data pack, adding the DNMT2 sequence of *Drosophila tropicalis* (*willistoni* subgroup) sequenced by our research group. The results of evolutionary and structural analyzes suggest differences between the DNMT2 properties of

subgroup *willistoni* species from the other drosophilids. These results indicate that even species that have close evolutionary relationships may have adaptive mechanisms that establish gradations of DNMT2 affinity for different substrates, without the need of drastic changes in enzyme architecture.

Capítulo I

Epigenética e o Fenômeno de Metilação

Introdução Geral

Justificativa

Objetivos

1 Introdução Geral

1.1.1 Epigenética

O termo epigenética foi cunhado por Conrad H. Waddington, em 1942, de origem da palavra grega *epigenesis*. Em parceria com Ernst Hadorn, Waddington tinha seus estudos focados em uma área que combinava a genética e o desenvolvimento biológico, procurando descrever e entender os processos genéticos envolvidos no desenvolvimento dos organismos. Através de suas observações, Waddington notou que fatores ambientais estressores poderiam causar a emergência de fenótipos não esperados em *Drosophila*, como um processo de assimilação genética (Waddington, 1942). Contudo, apenas no início da década de 90 foi possível a elucidação dos mecanismos moleculares envolvidos nas observações descritas por Waddington, devido ao interesse renovado na questão dos processos de assimilação genética.

Fundamentalmente o processo epigenético caracteriza-se por contribuir com diferentes fenótipos alternativos de um organismo a partir de seu genótipo. Esses fenótipos alternativos podem ser induzidos por determinados cenários ambientais, que por sua vez influenciam diferentes estados epigenéticos de um ou mais loci do genoma do organismo. Sabe-se ainda que a herança epigenética viabiliza a transmissão de informações que não são codificadas pelos genes de uma célula-mãe para a célula-filha, ou de geração para geração (Kiani *et al.*, 2013). De maneira resumida, as regulações epigenéticas podem ser descritas como as interações mútuas entre três eventos: a metilação de DNA, modificações de histonas com a remodelação do nucleossomo (Lo *et al.*, 2008) e a ação de RNA não-codificantes (Dozmorov *et al.*, 2013; Spitale *et al.*, 2011; Ørom and Shiekhattar, 2011).

Os mecanismos envolvidos nos fenômenos epigenéticos atuam na alteração da expressão gênica, mudando a acessibilidade à cromatina para a regulação transcricional por meio das modificações do DNA, mas sem haver alterações na sequência das bases do genoma. Esses processos são críticos no

desenvolvimento normal e na especialização de funções das diferentes células, dos diferentes tecidos de um organismo. Entretanto, as modificações epigenéticas desreguladas podem resultar em efeitos danosos ao organismo, resultando em diversas doenças como transtornos mentais, dependência de drogas (Tsankova *et al.*, 2007; Sweatt *et al.*, 2009; Kumar *et al.*, 2008) e câncer (Feinberg *et al.*, 1983; Hardy *et al.*, 2011; Park *et al.*, 2011).

1.1.2 O Fenômeno de metilação

A metilação do DNA genômico é um dos principais mecanismos de regulação epigenética nos organismos. O fenômeno de metilação do DNA consiste no processo de substituição de um átomo de hidrogênio (H) por um grupo metil (CH₃), sendo a reação catalisada por diversas metiltransferases (MTases). Os processos de *imprinting* genômico, a estabilidade genômica e o silenciamento de transposons e de retrovírus endógenos, por exemplo, são mediados por metilação (Bird, 2002; Yoder *et al.*, 1997).

Porém, os mecanismos de controle de expressão gênica envolvendo a metilação não atuam somente na atividade transcricional do gene, podendo atuar por vias indiretas, impedindo fatores de transcrição acessarem genes metilados ou interagindo com fatores *trans* do processo transcricional, como as histonas deacetilases e proteínas de ligação ao DNA metilado (MDBs) - essas últimas parecem competir por sítios alvos de fatores de transcrição. (Nan *et al.* 1998; Fuks *et al.*, 2000; Kim *et al.*, 2002). Quando da modificação promovida pelas histonas desacetilases, a cromatina adquire uma nova estrutura conformacional que desfavorece a iniciação transcricional. Neste contexto, os grupos metila interagem com as histonas formando nucleossomos que dificultam o acesso dos fatores transcricionais a determinadas regiões do DNA (Keshet *et al.*, 1986).

Outro aspecto interessante refere-se ao fenômeno de metilação. O mesmo não é homogêneo em posição e função ao longo do genoma. Em mamíferos são amplamente descritas metilações nas ilhas CpG nos sítios de

início de transcrição como sendo um mecanismo de silenciamento ou diminuição de expressão gênica (Okano *et al.*, 1999), contudo, em genes que não possuem sítios de início de transcrição ricos em CpG pode ocorrer metilação em sequências não-CpG, sendo o processo de silenciamento promovido primordialmente pelo nucleossomo (Farthing *et al.*, 2008; Han *et al.*, 2011).

Além disso, a localização das citosinas metiladas ao longo das diversas regiões que compõem um gene pode resultar em diferentes fenômenos, não estando somente envolvidas no processo de silenciamento gênico. Regiões a jusante do sítio promotor geralmente possuem depleção em dinucleotídeos CpG e possuem metilação em elementos repetitivos e no corpo do gene (Weber *et al.*, 2005). O que se observa, é que o processo de alongamento não é impedido pelas citosinas metiladas ao longo do gene, ao contrário do que ocorre no processo de início de transcrição, além delas poderem estar envolvidas no controle de *splicing* (Sharma *et al.*, 2011; Laurent *et al.*, 2010). Dessa forma, os estudos de epigenômica (mais especificamente o metiloma), favorecidos pelo crescente acúmulo de informações advindas dos mais diversos organismos que possuem seus genomas sequenciados, demonstram que os padrões de metilação se estendem para além das regiões promotoras dos genes, sendo mais heterogêneas em sua distribuição do que se supunha anteriormente.

1.1.3 As Metiltransferases

DNA Metiltransferases (DNA MTase) estão classificadas em três classes de acordo com o tipo de metilação que promovem. As duas primeiras classes de DNA MTase executam a metilação através da modificação dos nitrogênios exocíclicos dos nucleotídeos adenina e citosina, convertendo-os em N6-metiladenina (m6A-MTase) e N4-metilcitosina (m4C-MTase), respectivamente, sendo encontradas principalmente em procariotos.

A terceira classe refere-se às enzimas que catalisam a reação de metilação inserindo um grupamento metil (CH₃) no Carbono cinco do anel pirimidínico de citosinas, convertendo-as em 5-metilcitosina (**Figura 1**). As

enzimas pertencentes a essa última classe são denominadas DNA (citocina-5) metiltransferases (m5C-MTase) (Kumar *et al.*, 1994), distribuindo-se amplamente de procaríotos a eucariotos (Motorin *et al.*, 2010).

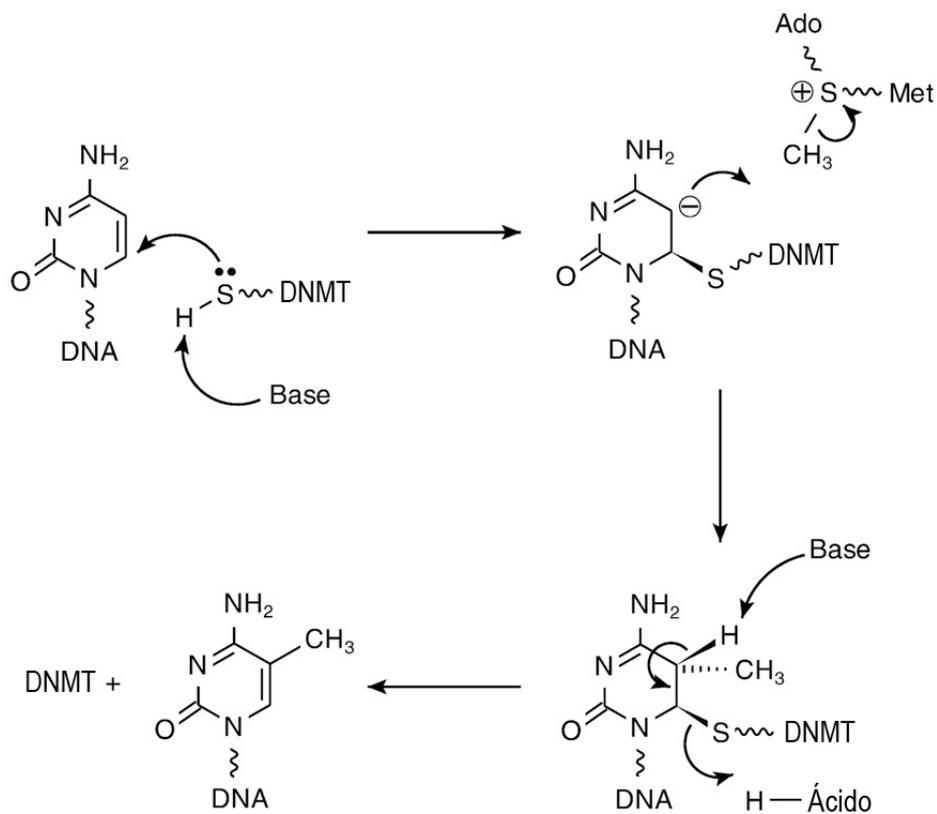


Figura 1. Representação esquemática do mecanismo de metilação de citosina por meio de DNMT. Num primeiro momento há um ataque nucleofílico na posição seis da citosina por parte da enzima (Cys81 em *MHal*, por exemplo). Esse ataque acaba por ativar a posição cinco para um ataque eletrofílico pelo cofator (AdoMet), quando o grupamento metil é transferido para a citosina.

Em vertebrados existem três diferentes famílias de m5C-MTase que são responsáveis por estabelecer e manter os padrões de metilação genômica ao longo do desenvolvimento dos organismos, sendo essas: a DNMT1, a DNMT2 e as DNMT3a e DNMT3b (**Figura 2**). É interessante compartilharem homologia de sequência no domínio catalítico (Kumar *et al.*, 1994; Jurkowski *et al.*, 2011, Mund *et al.*, 2004).

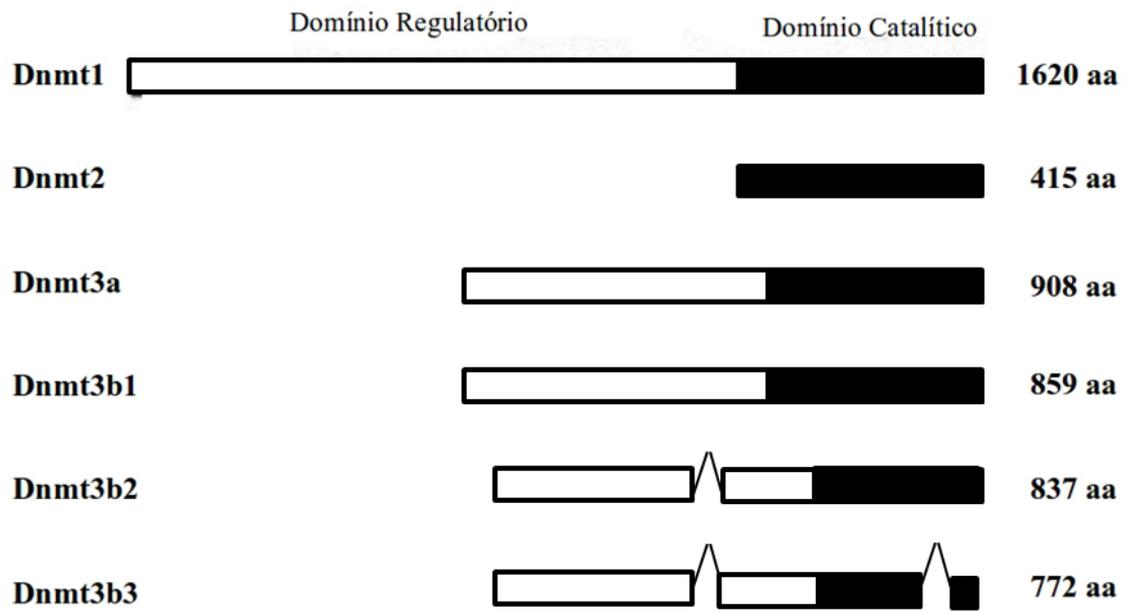


Figura 2. Ilustração esquemática de DNMTs conhecidas de vertebrados. A região em preto representa o domínio catalítico C-terminal conservado (imagem adaptada de Mund *et al.*, 2004).

1.1.4 A DNMT1: mantenedora das marcações epigenéticas

A DNMT1 é a maior das MTases, com massa molecular de 184 kDa (Yoder *et al.*, 1996), possui duas isoformas DNMT1a e DNMT1b, na qual a DNMT1b resulta do processamento de *splicing* alternativo do gene de *DNMT1* (Bonfils *et al.*, 2000). A DNMT1 possui maior atividade junto ao DNA hemimetilado oriundo das características de replicação semiconservativa do DNA, onde a fita molde que carrega os padrões de metilação guia a metilação das novas fitas sintetizadas (Yoder *et al.*, 1997). O silenciamento da *DNMT1* resulta em fenótipos letais por não haver a manutenção dos padrões de metilação no genoma replicado. Em estudos com camundongos, verificou-se que indivíduos homozigóticos *DNMT1*^{-/-} não sobrevivem à gestação. Disfunções de MTases também estão envolvidas em processos de tumorigênese, onde padrões de metilação anormais interferem nas expressões apropriadas de genes supressores tumorais e de genes regulatórios de crescimento. Essa associação é amplamente descrita, onde elevados níveis de transcritos de *DNMT1* e atividade MTase em células cancerígenas *in vitro* e tumores *in vivo* foram observados (Kautiainen *et al.*,

1986; el-Deiry *et al.*, 1991; Saito *et al.*, 2003).

Com aproximadamente 1600 aminoácidos, a DNMT1 possui seus primeiros 1100 resíduos localizados na região N-terminal (domínio regulatório), sendo os demais resíduos pertencentes à região C-terminal composta pelo domínio catalítico (Pradhan *et al.*, 2003). Ambos os domínios encontram-se ligados por uma região de repetição Gly-Cys.

A grande região regulatória da DNMT1 é capaz de interagir com diversos fatores envolvidos em diferentes processos moleculares de remodelação de cromatina. A DNMT1 é abundantemente encontrada em forquilhas de replicação de DNA, localização condizente com sua preferência por DNA hemimetilado. Estudos visando o entendimento dos mecanismos de controle e atividade enzimática da DNMT1 demonstraram interação entre essa e as proteínas PCNA (*Proliferative Cell Nuclear Antigen*) (Chuang *et al.*, 1997), histonas deacetilases (Fuks *et al.*, 2000; Rountree *et al.*, 2000), produtos do gene de retinoblastoma (Rb) (Robertson *et al.*, 2000; Pradhan *et al.*, 2002), DNMT3a e DNMT3b (Kim *et al.*, 2002).

Interessantemente, o trabalho cooperativo entre DNMT1 e DNMT3a (e/ou DNMT3b) faz-se importante no reconhecimento de regiões a serem metiladas, no processamento da metilação *de novo* e na manutenção da metilação em classes de sequências pertinentes aos sítios promotores de determinados elementos transponíveis e genes (Kim *et al.*, 2002).

Em mamíferos, no estágio de pré-implantação do embrião, o genoma encontra-se praticamente livre de metilação, entretanto, Mertineit *et al.* (1998) demonstraram que nesse estágio o gene *DNMT1* é transcrito e, posteriormente, traduzido no citoplasma, porém a DNMT1 não é direcionada para o núcleo, mantendo-se no citoplasma. Registra-se nesse período inicial de desenvolvimento uma intensa atividade catalítica das DNMT3a e DNMT3b, atuando como estabelecadores de padrões de metilação. Após esse período, os níveis e atividade da DNMT3a e DNMT3b diminuem, passando, então, a atuar primordialmente a DNMT1 (Okano *et al.* 1999). Dessa maneira, as DNMT3a e

DNMT3b parecem recrutar a DNMT1 para agilizar o estabelecimento da metilação *de novo* e posterior manutenção dessas marcas epigenéticas.

1.1.5 DNMT3a e DNMT3b: metilação *de novo*

As DNMT3a e DNMT3b são as principais moléculas envolvidas na metilação *de novo* de vertebrados, que é o fenômeno de estabilização dos padrões de metilação em estágios de desenvolvimento iniciais (Okano *et al.*, 1998). A inativação de ambas DNMT3 acarreta em fenótipos letais. A metilação *de novo* é essencial para que ocorra a correta compartimentalização do genoma durante o processo de diferenciação tecidual, pois dessa maneira se estabelecem os padrões de metilação peculiares nos genes tecido-específicos (Okano *et al.*, 1999). As DNMT3a e a DNMT3b diferenciam-se em suas afinidades espaço-temporais ao longo do estágio do desenvolvimento embrionário. A DNMT3b possui seu padrão de ação mais intenso nos estágios iniciais de desenvolvimento, mais especificamente metilando regiões centroméricas, enquanto a DNMT3a atua nos estágios posteriores, metilando genes ou sequências fundamentais para o desenvolvimento pós-implantação e depois do nascimento (Okano *et al.*, 1999).

Análises filogenéticas sugerem que os genes ancestrais de *DNMT3a* e *DNMT3b* surgiram depois da emergência dos urocordados e antes da radiação evolutiva dos gnatostomados, através de processo de duplicação gênica (Campos *et al.*, 2012). Cabe dizer que, DNMT3a e DNMT3b também diferem em sua afinidade por sequências de DNA (-3b possui maior afinidade por extensas regiões repetitivas) e processividade de metilação, sendo que a DNMT3b possui uma maior atividade de metilação de DNA do que a DNMT3a (Liu *et al.*, 2003). Após o fenômeno de duplicação o gene da DNMT3b sofreu diversas substituições não-sinônimas. Em mamíferos a simples substituição de um aminoácido, uma isoleucina por uma asparagina na posição 662 da DNMT3b, é considerada um evento preponderante para conceber o aumento de sua atividade catalítica (Shen *et al.*, 2010).

1.1.6 A DNMT2: sua ancestralidade comum e sua ampla distribuição

A subfamília DNMT2 tem características interessantes, pois as MTases pertencentes a esse grupo possuem seus sítios catalíticos altamente conservados, desde procaríotos até eucariotos (Jurkowski & Jeltsch, 2011) (**Figura 3**). Por exemplo, Shulz *et al.* (2012) verificaram que a família de MTases DNMT2 apresenta grande homologia, mesmo entre a enzima humana e a DNA MTase de procaríotos.

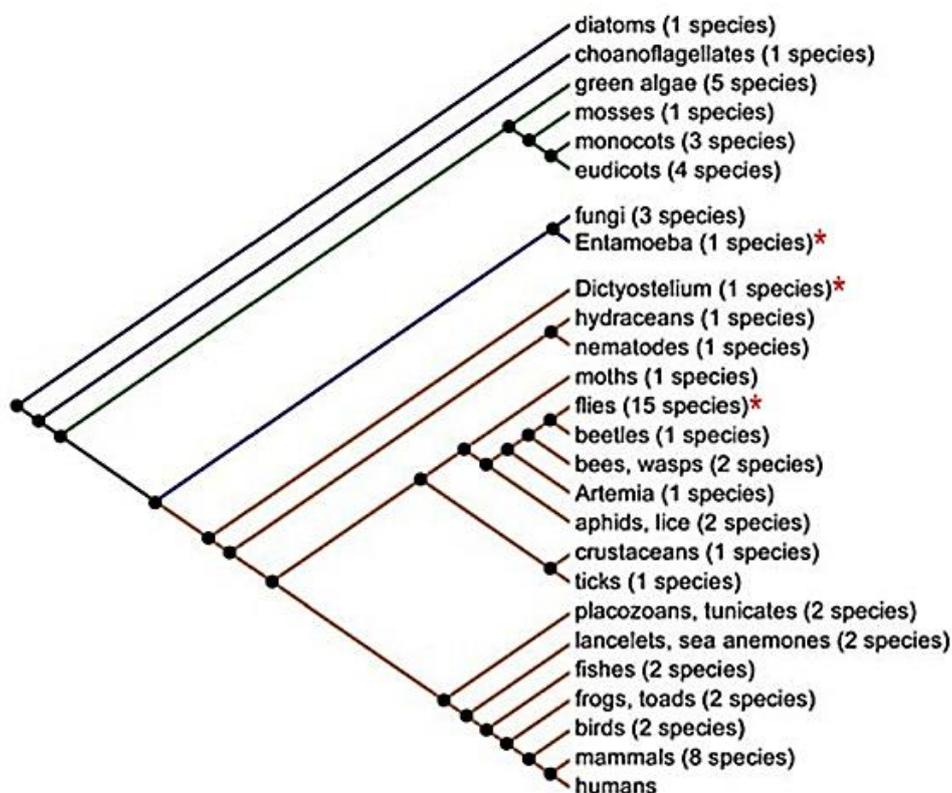


Figura 3. Conservação evolutiva de DNMT2 em diversos reinos. Destacados com asterisco vermelho os grupos onde a DNMT2 é a única metiltransferase até o momento descrita (imagem adaptada de Schaefer & Lyko, 2010).

A DNMT2 carrega propriedades que permitem executar funções tanto de metilação *de novo*, como de manutenção de metilação. Além da homologia e conservação de seus sítios catalíticos, a DNMT2 possui peculiaridades que acabam por atrair a atenção de diversos grupos de pesquisa, como, por exemplo, sua plasticidade ao substrato em que atua. Não é por acaso que ela é considerada uma subfamília enigmática (Dong *et al.*, 2001; Durdevic *et*

al., 2013; Ponger *et al.*, 2005; Jeltsch *et al.*, 2006; Schaefer *et al.*, 2006; Motorin *et al.*, 2010; Liu *et al.*, 2003).

Reconhece-se que a DNMT2 possui diferentes afinidades por DNA e RNA. Em mamíferos, invertebrados e plantas a DNMT2 é classificada *in vivo* como uma tRNA^{Asp} MTase (Goll *et al.*, 2006; Tuorto *et al.*, 2012). Entretanto, trabalhos realizados por Hermann (2003), Kunert (2003), Fisher (2004) e Kuhlmann (2005) relatam a atividade de DNA MTase da DNMT2, mesmo que em baixos níveis. O que se discute são as atividades preferenciais da DNMT2 e os mecanismos ecológicos e evolutivos que modulam sua atividade, pois se sabe que em organismos que não possuem as MTases canônicas (DNMT1 e DNMT3), mas apresentam metilação em seu genoma, a DNMT2 atua como MTase em ambos os substratos: DNA e tRNA (Krauss, 2011).

Interessantemente, a plasticidade da DNMT2 não se limita ao substrato em que atua, ela estende-se às dimensões espacial, temporal e tecidual. Devido a sua conservação e caracterização inicial como uma DNA MTase, a DNMT2 foi classificada como uma enzima nuclear. Entretanto, estudos posteriores com DNMT2 de humanos verificaram a sua presença em regiões citoplasmáticas (Goll *et al.*, 2006). Essa característica contrasta com a alta especificidade local das enzimas DNMT1 e DNMT3 à região do núcleo celular (Bachman *et al.*, 2001).

Contudo, a mobilidade da DNMT2 não é amplamente compartilhada ao longo da história evolutiva. Em *Dyctiostelium discoideum* e *Entamoeba histolytica*, a DNMT2 parece residir unicamente na matriz nuclear (Kuhlmann, 2005; Banerjee *et al.*, 2005). No que se refere à sua atividade tecido específica, a DNMT2 apresenta níveis de expressão variáveis, sendo que em humanos os tecidos cardíaco, pulmonar e testículos possuem elevadas taxas de expressão (Okano *et al.*, 1998, Yoder, 1998). Em *Drosophila melanogaster* e *Danio rerio* a expressão de DNMT2 apresenta-se em altos níveis nos estágios iniciais de desenvolvimento (Hung *et al.*, 2000; Lyko, 2000; Kunert, 2003), denotando sua expressão temporal.

Tais evidências demonstram que a DNMT2 é fundamentalmente diferente das demais DNA e RNA MTases, possuindo múltiplas atividades moleculares (Schaefer *et al.*, 2008). Partindo do princípio de que a DNMT2 possui essa característica polivalente, pode-se propor a hipótese de que a mesma interaja com diversas proteínas relacionadas à sinalização e transporte intracelular, além de estar envolvida em diferentes rotas metabólicas. Interações entre DNMT2 e enzima enolase - característica da rota glicolítica – em *Entamoeba histolytica* já foram descritas, onde a enolase estaria envolvida na modulação da atividade da DNMT2 como tRNA^{Asp} MTase, formando o complexo enolase-*E. histolytica* e DNMT2 homólogo (Ehmeth) em cenários de escassez de glicose (Tovy *et al.*, 2010).

Geralmente presente no núcleo, quando em situações de estresse a DNMT2 é mobilizada para regiões de grânulos de estresse citoplasmático e corpos de processamento de RNA, que estão envolvidos no silenciamento de retrotransposons, sugerindo que tal atividade esteja ligada a algum tipo de mecanismo de defesa celular primitivo contra infecções virais (Thiagarajan *et al.*, 2011). Logo, essa mobilização da DNMT2 provavelmente é mediada pela ação de moléculas de sinalização e migração da enzima do ambiente nuclear para o citoplasmático. Além do silenciamento de retrotransposons, a DNMT2 é essencial para a atividade da Dicer-2 durante a resposta de tratamentos *heat shock* (Durdevic *et al.*, 2013), estando envolvida no controle de expressão e regulação de diversas proteínas *Heat shock* (Fisher *et al.*, 2006; Schaefer *et al.*, 2010; Thiagarajan *et al.*, 2011), na longevidade das moscas (Lin *et al.*, 2005) e também na integridade estrutural da cromatina (Phalke *et al.*, 2009).

Portanto, apesar da DNMT2 possuir forte conservação de seus sítios catalíticos, sua principal funcionalidade e seus mecanismos de modulação ainda são uma incógnita para os pesquisadores. Estudos promovendo a depleção de DNMT2 em diversos organismos falharam em elucidar sua função primordial. Dessa forma, novos estudos estão procurando encontrar as respostas através de uma nova abordagem, identificando as capacidades funcionais da DNMT2 com estudos da sua relação com as proteínas com as quais interage.

1.1.7 O Fenômeno de Metilação em Drosofilídeos

Quanto aos fenômenos de metilação em invertebrados, até final da década de 90 o genoma de *Drosophila melanogaster* (organismo modelo amplamente empregado em estudos genéticos e epigenéticos) era considerado livre de metilação, como descritos em diversos trabalhos que analisavam os estágios de pupa e adultos de *Drosophila* (Patel & Gopinathan, 1987; Tweedie, *et al.*, 1999). Contudo, estudos realizados por Gowher *et al.* (2000), Lyko *et al.* (2000) apresentaram indícios de que haveria baixos níveis de metilação em embriões e adultos de *Drosophila* (**Figura 4**). Gowher (2000), verificou que aproximadamente 1 em 1000-2000 resíduos de citosina são metilados em moscas adultas, sendo que a metilação do DNA foi detectada em todos os estágios de desenvolvimento de *D. melanogaster*. Com análises das sequências genômicas de *Drosophila*, empregadas por Lyko (2000), confirmou-se a existência de metilação nesses organismos, mas com a peculiaridade de serem em nucleotídeos CpA, CpT e CpC em embriões de *D. melanogaster*, e não em CpG como é comumente encontrado em mamíferos.



Figura 4. Níveis de metilação de citosinas ao longo dos estágios de desenvolvimento em *Drosophila melanogaster* (imagem modificada de Lyko *et al.*, 2000)

O advento de novas técnicas de análise molecular e genômica proporcionaram maiores esclarecimentos a respeito da importância e funções exercidas pela enzima DNMT2 em invertebrados, especialmente em *Drosophila*.

Em nosso grupo de pesquisa, com foco no estudo de espécies de *Drosophila* Neotropicais, como o subgrupo *willistoni*, evidenciou-se pela primeira vez, o fenômeno de metilação sexo-específica do DNA em *Drosophila willistoni* (Garcia *et al.*, 2007).

O peculiar fenômeno detectado demonstra a importância de ampliarem-se os estudos em espécies Neotropicais para aferir maior entendimento evolutivo referente a esse grupo de espécies de *Drosophila*. Em especial, o subgrupo *willistoni* chama atenção por sua versatilidade ecológica e por sua capacidade de explorar vários nichos diferenciados (D'Ávila, 2011). Além disso, o grande polimorfismo cromossômico e as inversões distribuídas em todos os braços cromossômicos politênicos (Da Cunha *et al.*, 1950) acaba por conferir a esse grupo de espécies uma grande plasticidade cariotípica, sendo de grande interesse para estudos evolutivos, moleculares, ecológicos e taxonômicos.

Garcia *et al.* (2007), de nosso grupo de trabalho, em estudo comparativo das sequências de aminoácidos de DNMT2 de *D. willistoni* e *D. melanogaster* demonstraram significativa conservação dos motivos pertencentes aos sítios catalíticos, responsáveis pela metilação da citosina alvo. No entanto, a região de reconhecimento da sequência alvo a ser metilada (TRD) apresentou-se significativamente variável entre as duas espécies. Estudos posteriores, também realizados por nosso grupo de pesquisa, demonstraram que entre as MTases de diversas espécies do gênero *Drosophila*, existe significativa conservação dos dez motivos constituintes da enzima DNMT2 (D'Ávila *et al.*, 2010), incluindo os motivos consensos de ligação com a S-adenosilmetionina (AdoMet, estrutura doadora do grupamento metil) e o sítio ativo do dipeptídeo ProCys, responsável pelo ataque nucleofílico da citosina que receberá o grupamento metil.

Ainda hoje, alguns pesquisadores consideram incipiente o fenômeno de metilação em drosofilídeos, baseados no baixo percentual de citosinas metiladas no genoma dos indivíduos desse grupo e pela postura consensual de que a DNMT2 é muito mais uma tRNA^{asp} metiltransferase do que propriamente uma DNA metiltransferase (Hermann *et al.*, 2003; Liu *et al.*, 2003; Tang *et al.*, 2003; Jeltsch *et al.*, 2006). Porém, a descoberta do fenômeno de metilação sexo-

específico reestabelece a importância de manter um olhar crítico sobre tais cenários, justamente devido ao fato de serem peculiares a um dado grupo de organismos descritos como *DNMT2-only*.

1.1.8 Sobre a Evolução das Proteínas: Estrutura e Função

A evolução molecular é ocorre basicamente por mutações, as quais podem – ou não – afetar a aptidão de um organismo e assim ser alvo da seleção natural ou deriva genética. A análise de sequências primárias e terciárias de proteínas trazem interessantes sugestões a respeito da evolução da função de proteínas (Bhabha *et al.*, 2013). Dessa maneira, as pressões evolutivas podem influenciar a dinâmica estrutural da proteína em um processo adaptativo a diferentes ambientes, facilitando dessa maneira o surgimento de novas funcionalidades.

As assinaturas de processos evolutivos se expandem além da análise de substituições sinônimas e não sinônimas, mas se encontram também nas “pistas” conformacionais de novas estruturas que possam representar modificações funcionais das proteínas. Similaridades nas estruturas 3D de proteínas podem existir na ausência de identidade de sequências (Kinch *et al.*, 2002; Lupas *et al.*, 2001). Porém, tais congruências de empacotamento nem sempre resgatam a homologia existente entre duas proteínas similares tridimensionalmente. Isso se deve, porque ao longo da evolução, pressões evolutivas similares podem engendrar o surgimento de estruturas proteicas parecidas, tanto em forma, como em função, mas sem possuírem homologia – convergência evolutiva molecular. Contudo o inverso também é válido.

Claramente, a organização espacial proteica tende a ser mais conservada do que sua sequência durante a evolução. Entretanto, o crescente acúmulo de informações de estruturas tridimensionais de proteínas tem mostrado que proteínas homólogas podem ser significativamente distintas em termos estruturais, apesar da conservação significativa de suas sequências primárias e similaridades funcionais (Grishin, 2001). A existência dessas relações evolutivas

distintas trazem novos desafios para o campo de estudos da evolução estrutural de proteínas e modelagem computacional.

Além das análises de estruturas estáticas das proteínas e suas geometrias, é muito importante observar a dinâmica estrutural das mesmas (Liberles *et al.*, 2012). Para isso há esforço conjunto de diversas áreas como a biofísica, bioquímica, biologia evolutiva e bioinformática.

1.1.9 Bioinformática Estrutural: um campo interdisciplinar para resolução de estruturas e sistemas de proteínas alvo

A bioinformática é um campo interdisciplinar relativamente novo e que possui vasto potencial a ser explorado, propiciado pelos avanços tecnológicos na área da computação e aumento da capacidade de processamento de dados de máquinas e estruturação de servidores. A interação entre as ciências biológicas, química, física, matemática e computação forma um ambiente instigante, que leva a uma maior compreensão das relações estruturais e funcionais de proteínas e enzimas. De acordo com Sánchez (1998), o entendimento e a possibilidade do estudo das funções, estruturas e mecanismos pelas quais as proteínas executam suas tarefas é um passo importante fornecido pela biotecnologia.

Em um primeiro momento a compreensão das estruturas tridimensionais de moléculas possibilitou avanços significativos nas áreas farmacológicas e de diagnose (Santos-Filho & Alencastro, 2003). Com o crescente número de estruturas de proteínas oriundas de técnicas como a Cristalografia de Raios-X e Ressonância Magnética Nuclear (RMN) depositadas em bancos de dados, emergiu a possibilidade de estruturar essas informações em projetos proteômicos amplos. Esses projetos, por conseguinte, sistematizaram as informações depositadas ao ponto de poder se estabelecer um arcabouço de conhecimento das mais diversas interações moleculares e funções biológicas de inúmeras proteínas. Dessa maneira, é possível afirmar que a função de uma proteína pode ser concebida a partir de sua estrutura tridimensional, o que leva à

importância de se conhecer a estrutura 3D de sequências proteicas (Sánchez, 1998).

Os métodos experimentais de resolução de estruturas tridimensionais de proteínas (Cristalografia de Raios-X e RMN) - apesar de reconhecida importância - possuem diversas limitações, como por exemplo, a necessidade de grandes quantidades de amostras e a possível baixa qualidade dos cristais para o trabalho experimental. Outro revés pode ser encontrado em determinadas classes de proteínas como as proteínas de membrana, das quais dificilmente se obtém uma cristalização que possa ser utilizada satisfatoriamente por RMN.

Diante dessas dificuldades, é compreensível a discrepância entre o número de estruturas primárias e estruturas tridimensionais de proteínas disponíveis nos bancos de dados. Atualmente encontram-se depositadas 96.700 estruturas cristalográficas proteicas no PDB (até 02 de novembro de 2014) contra 546.800 sequências de proteínas anotadas no UniProt (2014). Por essas e outras razões, a utilização de métodos computacionais se tornou tão importante para a ampliação do conhecimento e para a resolução de estruturas proteicas ainda não conhecidas (Schwede *et al.*, 2003; Hillisch *et al.*, 2004)

1.1.10 Modelagem comparativa por homologia: MODELLER e SWISS-MODEL

A modelagem comparativa por homologia é uma das abordagens com maior nível de sucesso para predição de estruturas 3D de proteínas (Blundell, 2003). Esse tipo de abordagem parte de alguns pressupostos destacados por Branden & Tooze (1991), que observaram que as estruturas proteicas possuem um maior grau de conservação do que as suas sequências de aminoácidos. Além disso, (i) a homologia existente entre sequências primárias resulta em semelhanças estruturais e funcionais; (ii) as regiões mais internas de proteínas homólogas apresentam-se mais conservadas do que suas superfícies; (iii) grande parte das diferenças estruturais em proteínas homólogas ocorrem em regiões

menos ordenadas, como regiões de alças. Outro aspecto que se pode destacar é a existência de um número limitado de famílias tridimensionais conhecidas entre as proteínas, o que possibilita uma predição da estrutura da proteína alvo quando houver alguma outra estrutura proteica resolvida por meios experimentais.

Assim, para a construção de um modelo é necessário que exista uma estrutura 3D resolvida por técnicas experimentais, além de uma identidade mínima entre as estruturas primárias da sequência alvo e do molde. De modo geral, um valor de identidade acima de 25% entre proteína alvo e seu molde é o recomendado para se efetuar a construção por homologia de um modelo preditivo confiável (Cavasotto & Phatak, 2009; D'Alfonso et al., 2001; Marti-Renom et al., 2000).

A Modelagem Molecular por homologia apresenta quatro etapas principais: a procura por sequências de proteínas homólogas, o alinhamento das sequências, a construção e otimização dos modelos e por fim a avaliação e validação das estruturas geradas (Cavasotto & Phatak, 2009).

O MODELLER é um programa amplamente utilizado para modelagem por homologia ou comparativa de proteínas (Eswar, 2006; Marti-Renom, 2000), sendo capaz de calcular automaticamente modelos contendo todos os átomos que não sejam hidrogênio. O programa implementa o método de modelagem comparativa da estrutura de proteínas satisfazendo suas restrições espaciais (Sali, 1993; Fiser 2000). Nesse método é realizada a geração de um conjunto de diretrizes de restrições espaciais que são transferidas à sequência alvo. Dentre as restrições, consideram-se as ligações angulares (entre três átomos) e os ângulos diedrais (entre quatro átomos). Além das restrições, um campo de força controla as propriedades estereoquímicas entre os átomos (Rössle *et al.*, 2004). Todas as restrições químicas e espaciais aplicadas ao modelo são combinadas em uma função, que é minimizada durante o processo de construção do modelo (Fiser, 2003). Além disso, o programa MODELLER pode executar muitas tarefas adicionais, incluindo modelagem *de novo* de alças em estruturas de proteínas e otimização de vários modelos de estrutura de proteínas em relação a uma função, objetivo definido de forma flexível.

Já o software SWISS-MODEL é uma ferramenta que está disponível no servidor da *Swiss Institute of Bioinformatics*. O software também realiza modelagem de proteína por homologia, mas através do método de união dos corpos rígidos. Nessa metodologia a construção da proteína-problema ocorre por meio das transferências de coordenadas atômicas dos moldes utilizados. Após o alinhamento entre as sequências molde e alvo, os resíduos são sobrepostos e trocados, mas as coordenadas espaciais são mantidas (Goldsmith-Fischman and Honig, 2003). A busca da estrutura molde é efetuada através do banco de dados PDB. O programa, então, encaminha de forma automática todo o processo de modelagem. Agregam-se ao servidor, ferramentas de análise e validação da estrutura gerada, como WHATCHECK, PROCHECK e CASP (Arnold, 2006).

1.1.11 Modelagem molecular por métodos mistos: I-TASSER e Robetta

Através da plataforma on-line do servidor I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>), é possível a submissão de sequências alvo para a modelagem e estruturação de proteínas. Os modelos 3D são construídos utilizando uma metodologia intermediária entre o método por homologia e o *ab initio*, denominada de *threading*.

Esta ferramenta utiliza a técnica de reconhecimento de padrões de enovelamento através de alinhamentos de múltiplos fragmentos de diversos *templates* localizados no banco de dados da BioLip (<http://zhanglab.ccmb.med.umich.edu/BioLiP/>) (Zhang, 2008; Roy, 2010; Roy, 2012). Os fragmentos identificados são então agrupados e as regiões que não possuem moldes são resolvidas por *ab initio*. A montagem dos fragmentos é feita através de simulações de Monte Carlo. Finalmente é realizado o refinamento onde são recuperados cinco confôrmeros para a realização de uma nova rodada de agrupamento por troca de réplicas, esse processo visa eliminar possíveis impedimentos estéricos. Ao final são adicionadas as cadeias laterais e os confôrmeros são submetidos a um último passo de refinamento para otimização dos rotâmeros (Roy *et al.*, 2007; Zhang, 2008).

O programa Robetta permite a modelagem de proteínas através de abordagens *ab initio* e por comparação. Assim, domínios proteicos sem homologia com estruturas cristalográficas depositadas em PDB, podem ser modeladas com o protocolo *de novo* do algoritmo Rosetta. O protocolo é chamado de "reconstrói e refina", que simplificada é composto por uma etapa que emprega um modelo minimalista, chamada modelagem de baixa resolução, seguida de uma etapa que utiliza um modelo *full-atom*, chamada modelagem de alta resolução. Na etapa de modelagem de baixa resolução, utiliza-se um campo de força reduzido, também chamado de baixa resolução. Essas regiões são construídas a partir de bibliotecas de fragmentos. A modelagem comparativa é construída a partir de moldes de estruturas cristalizadas (PDBs) detectados e alinhados com algoritmos presentes no HHSEARCH/HHpred, Raptor X e Sparks-X. O processo é completamente automatizado.

1.1.12 Validação dos Modelos

Todos os modelos construídos por homologia conterão algum tipo de erro. A verificação do modelo desenvolvido, a estimativa de verossimilhança e magnitude desses possíveis erros são passos importantes no processo de modelagem por homologia (Xiang, 2006). Entre as diversas formas de validação de modelos, a principal é a função de escore (*scoring function*), que é capaz de discriminar um bom modelo de um modelo errado.

Basicamente as funções de escore são divididas em duas grandes categorias:

- Os escores estatísticos, que se baseiam em propriedades experimentalmente observadas dos aminoácidos, analisando os pontos de contatos entre resíduos, os contatos internos dos resíduos e os contatos interatômicos e/ou os contatos com solventes (Sippl, 1995).
- As funções efetivas de energia física, que diretamente

avaliam a energia livre conformacional de uma proteína. Essa função distingue por análises estatísticas a conformação nativa de uma proteína a partir de um empacotamento incorreto. É uma análise custosa computacionalmente e baseia-se em interações físicas bem definidas (Lazaridis & Karplus, 1999).

Assim, a validação fornecerá os níveis de qualidade e confiabilidade da estrutura modelada e será imprescindível para otimização do modelo (Cavasotto & Phatak, 2009). Existem diversos programas que podem ser utilizados para a validação dos modelos gerados e que podem ser categorizados da seguinte forma (Xiang, 2006):

- Primeira categoria: programas de análise de propriedades estereoquímicas, assim como verificação de simetria, de geometria (quiralidade, distância das ligações, ângulo de ligações, ângulo de torções, etc.) e qualidade do empacotamento estrutural (exemplo, PROCHECK e WHAT IF);
- Segunda categoria: programas que verificam a adequação da sequência à estrutura, assinalando uma pontuação para cada resíduo que esteja de acordo com ambiente específico (exemplo, VERIFY3D).

Após o processo de validação, caso o modelo não seja considerado satisfatório, faz-se necessário checar o alinhamento utilizado e refazer o processo de modelagem utilizando novas metodologias.

1.2 Justificativa

Este trabalho possui o intuito de contribuir para um maior conhecimento acerca da estrutura tridimensional da proteína DNMT2 de drosofilídeos e suas relações evolutivas, além de estudar aspectos de expressão

e localização espaço-temporal da proteína em diferentes etapas de desenvolvimento de *Drosophila melanogaster* e *Drosophila willistoni*.

Particularmente, o fenômeno de metilação em drosofilídeos caracteriza-se por apresentar baixos níveis de citosinas metiladas, com prevalência em nucleotídeos CpT e CpA. Nosso grupo de pesquisa demonstrou a existência de padrões de metilação sexo-específica em espécies Neotropicais do grupo *willistoni* de *Drosophila*, não apresentando recorrência desse fenômeno em outras espécies de drosofilídeos fora desse grupo (D'Ávila *et al.*, 2010).

O maior entendimento em relação às características estruturais e físico-químicas, além da localização espaço-temporal da proteína DNMT2, pode contribuir para a compreensão dos fenômenos peculiares a esse grupo de espécies Neotropicais, favorecendo a identificação de possíveis proteínas interagentes com DNMT2 que possibilitem o entendimento de sua funcionalidade, importância evolutiva e ecológica.

1.3 Objetivos

3.1 Objetivo geral

O objetivo do presente trabalho é realizar a modelagem molecular da enzima DNMT2 de *Drosophila melanogaster* e *Drosophila willistoni* utilizando diferentes metodologias. Além disso, visa detectar a presença da DNMT2 em diferentes fases de desenvolvimento desses organismos associando com a importância evolutiva e funcional de sua expressão espaço-temporal.

3.2 Objetivos específicos

- Sugerir a estrutura tridimensional da proteína DNMT2 de *Drosophila melanogaster* e *D. willistoni* utilizando diferentes ferramentas de modelagem molecular;

- Avaliar o melhor modelo obtido a partir de cada uma das metodologias através de diferentes métodos de validação de modelos (PROCHECK, WHAT IF, Q-MEAN6, VERIFY3D e ModFOLD);
- Realizar dinâmica molecular com os modelos de DNMT2 desenvolvidos e estruturas cristalográficas depositadas no PDB, em um período de 50 ns com o software GROMACS, de forma a verificar a qualidade e comportamento das estruturas em condições fisiológicas aproximadas;
- Relacionar aspectos evolutivos e funcionais da família DNMT2 através de estudos comparativos com demais estruturas resolvidas de DNMT2 de outros organismos;
- Analisar a presença da enzima DNMT2 em diferentes estágios de desenvolvimento de *D. melanogaster* e *D. willistoni* através de *immunoblot* com anticorpos anti-DNMT2.

CAPÍTULO II

Modelagem Molecular & Análises Evolutivas

Metodologia

Resultados e Discussão

2.1. Metodologia

2.1.1 Predição e resolução das estruturas tridimensionais de DNMT2 humana, *Mus musculus*, *Drosophila melanogaster* e *Drosophila willistoni*

O modelo cristalográfico da DNMT2 humana (PDB 1G55), depositado no Protein Data Bank (PDB) possui duas regiões não resolvidas que compreendem os resíduos 79-96 e 189-247. Os resíduos 79-96 compreendem a alça do domínio catalítico, onde se encontram os dipeptídeos ProCys que estão envolvidos no processo de transferência do grupamento metil ao carbono-5 da citosina-alvo. Os resíduos 189-247 pertencem a uma região não conservada, presente apenas nas DNMT2 de mamíferos. Yoder *et al.* (2001) verificaram que esse segmento representa um domínio independente da proteína, não interagindo fortemente com o restante da estrutura. A sequência fasta completa da DNMT2 humana (NP_004403.1) e de *Mus musculus* (AAC40130.1) foram recuperadas a partir da busca no banco de dados Genbank (<http://www.ncbi.nlm.nih.gov/genbank/>) (Figura 5).

```
>Dnmt2_humana (NP_004403.1)
MEPLRVLELYSGVGGMHHALRESCIPAQVVAIDVNTVANEVYKYNFPHTQLLAKTIEGITLEEFDRLSF
DMILMSPPCQPFTRIGRQGDMTDSRTNSFLHILDILPRLQKLPKYILLENVKGFVSSSTRDLLIQTIENC
GFQYQEFILSPTSLGIPNSRLRYFLIAKLQSEPLPFQAPGQVLMFEFPKIESVHPQKYAMDVENKIQEKNV
EPNISFDGSIQCSGKDAILFKLETAEEIHRKNQDSDLSVKMLKDFLEDDTDVNQYLLPPKSLRLRYALLL
DIVQPTCRRSVCFTKGYGSYIEGTGSVLQTAEDVQVENIYKSLTNLSQEEQITKLLILKLRIFTPEKIAN
LLGFPPPEFGFPEKITVKQRYRLLGNLNVHVVAKLKILYE

>Mus musculus (AAC40130.1)
MEPLRVLELYSGIGGMHHALRESHIPAHVVAIDVNTVANEVYKHNFPHTHLLSKTIEGISLEDFDKLSF
NMILMSPPCQPFTRIGLQGDMDPRTTSFLYILDILPRLQKLPKYILLENVKGFVSSSTRGLLIQTIIEAC
GFQYQEFLLSPSSLGIPNSRLRYFLIAKLQSEPFQAPGQILMEFVKIVTVEPQKYAVVEESQPRVQRT
GPRICAESSSTQSSGKDTILFKLETVEERDRKHQDSDLSVQMLKDFLEDGDTDEYLLPPKLLRLRYALLL
DIVKPTSRSSMCFTKGYGSYIEGTGSVLQAEDAQIENIYKSLPDLPPPEEKIAKLSMLKLRIFTPEKIAN
LQGFPPPEFGFPEKTTVKQRYRLLGNLNVHVVAKLLTVLCEGFGNASESCHKMPLILDSNSKILS
```

Figura 5. Sequências proteicas de DNMT2 humana (391 aminoácidos) e *M. musculus* (415 aminoácidos) traduzida a partir da junção dos éxons. Destacado em cinza os resíduos pertencentes às regiões não resolvidas na estrutura cristalográfica da DNMT2 humana (PDB 1G55).

Para a DNMT2 de *D. melanogaster* foi utilizada a sua isoforma C, cuja sequência está depositada no Genbank sob o número de acesso NP_001036355, possuindo 345 aminoácidos. Já para a sequência de DNMT2 de *D. willistoni* utilizou-se sua isoforma B, conforme descrita no trabalho de Garcia *et al.* (2007), contendo 341 aminoácidos. Em ambas as espécies, a proteína pode apresentar variações quanto ao número de aminoácidos, como resultado de processamento alternativo dos transcritos. Além disso, efetuou-se a busca das demais estruturas cristalográficas de DNMT2 depositadas para estudos evolutivos da estrutura terciária dessa família de 5mC MTase. Antes do processo de modelagem molecular computacional as sequências-alvo de *D. melanogaster* e *D. willistoni*, *M. musculus* e DNMT2 humana foram submetidas à análise do método preditivo de estrutura secundária PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) (Jones, 1999).

```

>Drosophila_melanogaster (NP_001036355)
MVFRLVLELFSGIGGMHYAFNYAQLDGGQIVAALDVNTVANAVYAHNYGSNLVKTRNIQSLSVKEVTKLQAN
MLLMSPPCQPHTRQGLQRDTEKRSALDTHLCGLIPECQELEYILMENVKGFESSQARNQFIESLERSGF
HWREFILTPQTQFNVPNTRYRYYCIARKGADFPFAGGKIWEEMPGAI AQNQLSQAIEIVEENVSPDFLVP
DDVLTKRVLVMDIIHPAQRSRSMCF TKGYTHYTEGTGSAYTPLSEDESHRIFELVKEIDTSNQDASKSEKI
LQQRLDLLHQVRLRYFTP REVARLMSFPENFEFPPETTNRQKYRLLGNSINVKVVGELIKLLTIK

>Drosophila_willistoni (Garcia et al. 2007)
MKEFRVLELFSGVGGMHYAFNCTKLQGEVVAAMDVNTVANKVYAHNYKNPKYLKTRNIQSLSEKEVNLNA
NMLLMSPPCQPHTRQGLEQRDTEKRSALNQLCSLLPKCETIKYILMENVKGFECQARQQFIEALQQA
KFYWREFILTPQTQFQVPNTRYRYYCLARKDKDFDFPSGKIWEQMPGQAPHPTNEMDTISSLLEPSISTSE
FLVPDDVLRVLM DIVNSDQNRSMCF TKGYTHYTEGTGSTFTPLVRDESDRLF EIVKEIDNDPNRTNWE
RLDVLRLQKLRYFTP REVARLMKFPEYDFDFPPDITNRQKYRLLGNSINVKVVGELIKLLTT

```

Figura 6. Sequências proteicas de DNMT2 de *D. melanogaster* (345 aminoácidos) e *D. willistoni* (341 aminoácidos) traduzida a partir da junção dos éxons.

De maneira esquemática, as etapas desenvolvidas no presente trabalho seguem as seguintes etapas apresentadas no fluxograma da **figura 7**.

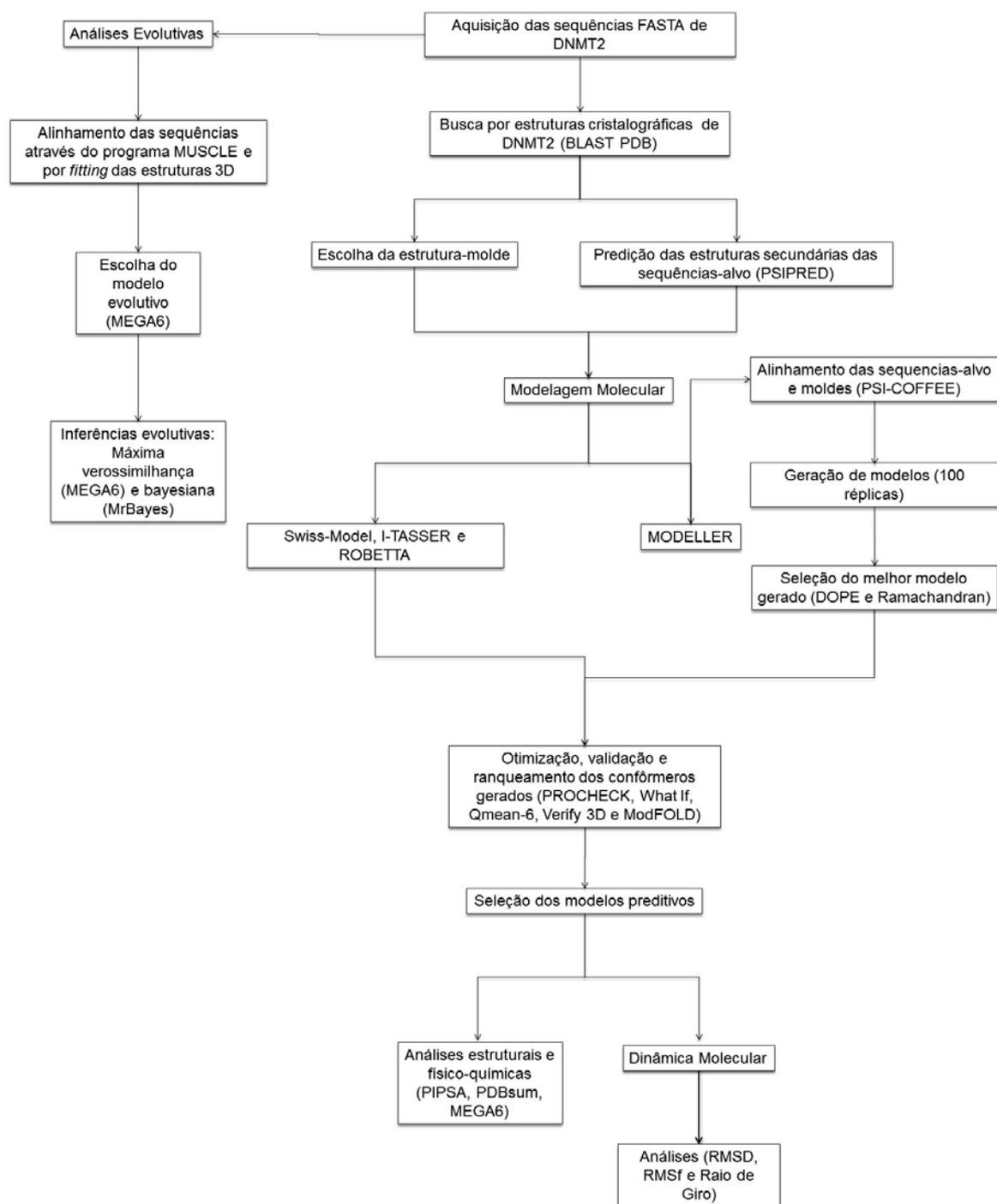


Figura 7. Fluxograma apresentando as etapas desenvolvidas no presente trabalho.

2.1.1.1 Modelos DNMT2 I-TASSER

A partir das sequências-alvo de *D. melanogaster*, *D. willistoni*, *M. musculus* e *humana* foram gerados cinco modelos, utilizando o software I-TASSER (Zhang, 2008). Após o processo de modelagem conforme descrito na introdução, os modelos foram ordenados de acordo com os valores C-score e TM-Score.

O C-score é uma medida que estima a qualidade dos modelos preditos. Ele é calculado a partir da significância dos enovelamentos oriundos do alinhamento com as estruturas-moldes e dos parâmetros de convergência das simulações de montagem das estruturas. O C-score varia de -5 a 2, de modo que quanto maior o valor de C-score maior o índice de confiança do modelo gerado. A acurácia do modelo é obtida através do TM-score, no qual valores entre 0,0 e 0,17 significam que a similaridade estrutural é ao acaso, enquanto valores entre 0,5 e 1,0 são indicativos de uma topologia correta do modelo em relação à proteína nativa (Zhang & Skolnick, 2004).

2.1.1.2 Modelos DNMT2 MODELLER

Os modelos gerados a partir das sequências completas das DNMT2 humana e de *M. musculus* foram construídos através de modelagem com múltiplos moldes, pelo fato do modelo cristalográfico da DNMT2 humana (PDB 1G55) possuir regiões não resolvidas. Foram utilizadas como moldes as estruturas da própria DNMT2 humana e a estrutura da DNMT2 de *Spodoptera frugiperda* (PDB 4H0N). Esta segunda estrutura foi obtida através de busca por BLASTp, com pesquisa nos bancos de dados PDB.

As sequências da MTase de humanos e *S. frugiperda* apresentaram 46% de identidade. Já a identidade entre *M. musculus* e DNMT2 humana é de 80.8%, e de 47% em relação a *S. frugiperda*. As identidades entre as sequências de DNMT2 de *D. melanogaster* / humana e de *D. melanogaster* / *S. frugiperda* são de 42% e 42.2%, respectivamente. *D. willistoni* compartilha identidade de 41.1% e

41.6% com as MTases de humanos e *S. frugiperda*, respectivamente. Entre *D. melanogaster* e *D. willistoni* há identidade de 69.2% dos resíduos. O alinhamento entre as sequências foi gerado através do programa PSI-COFFEE, que tem significativa acurácia no alinhamento de proteínas distantemente relacionadas através da extensão por homologia (Kemena, 2009; Di Tommaso, 2011).

Visando obter o melhor modelo tridimensional para *D. melanogaster* e *D. willistoni*, foram realizadas três modelagens independentes com diferentes alinhamentos:

- i) sequência alvo-1G55 (*Homo sapiens*);
- ii) sequência alvo-4H0N (*Spodoptera frugiperda*);
- iii) sequência alvo-1G55 e 4H0N,

Ao todo foram gerados 100 Modelos de cada espécie pelo programa Modeller v9.11, e de cada processo de modelagem o confôrmero que apresentou o melhor valor DOPE (além de análise do mapa de Ramachandran) foi selecionado para as validações.

2.1.1.3 Modelos DNMT2 SWISS-MODEL

A modelagem por homologia das proteínas de DNMT2 também foi realizada pelo algoritmo do SWISS-MODEL disponível no site do *Swiss Institute of Bioinformatics* (<http://swissMDEl.expasy.org/>). O programa efetua de forma automática todo o processo de modelagem, bastando alimentar o algoritmo com a sequência-alvo.

Em seguida o servidor buscou por estruturas-moldes existentes no banco de dados PDB e as disponibilizou para escolha do usuário. Para as sequências de *D. melanogaster* e *D. willistoni* a estrutura molde utilizada foi o de *S. frugiperda* (PDB 4H0N). Para a modelagem da DNMT2 humana e de *M. musculus* foi utilizada como molde apenas a estrutura PDB 1G55 (DNMT2 humana), já que o servidor não permite modelagem por múltiplos *templates*. Em

seguida o SWISS-MODEL efetuou a predição de estrutura secundária e alinhamento das sequências alvo e molde, realizando a posterior modelagem por homologia. Ao final o programa retornou os modelos elaborados com os dados de validação por QMEAN4.

2.1.1.4 Modelos DNMT2 ROBETTA

Através do servidor Robetta (<http://RBetta.bakerlab.org/>) as quatro sequências alvo alimentaram o programa de modelagem mista – uma rodada de modelagem para cada. O algoritmo Robetta é completamente automatizado e efetuou a busca por moldes e estabeleceu o alinhamento entre as sequências, conforme mencionado na introdução. Cinco modelos foram recuperados ao final para DNMT2 de humanos, *M. musculus*, *D. melanogaster* e *D. willistoni*.

2.1.1.5 Validação dos modelos

Os modelos foram validados através de diferentes ferramentas como o Verify3D (Bowie *et al.*, 1991; Lüthy *et al.*, 1992), What If (Vriend e Sander, 1993), Qmean6 (Benkert *et al.*, 2011), Procheck (Laskowski *et al.*, 1996) e ModFOLD (McGuffin, 2008).

2.1.1.6 Simulação de Dinâmica Molecular

Os melhores modelos gerados de cada espécie (através das metodologias descritas anteriormente), e também as estruturas cristalográficas de DNMT2 disponíveis no banco de dados PDB, foram submetidas às simulações de Dinâmica Molecular. As simulações foram realizadas com o pacote GROMACS v4.5.1 (Van Der Spoel *et al.*, 2005), em uma plataforma Linux (Ubuntu 12.04), utilizando protocolo já estabelecido em nosso laboratório.

O protocolo consiste numa série de passos, que podem ser descritos

em cinco etapas: (1) preparação do sistema, (2) minimização de energia, (3) solvatação da proteína, (4) termalização e (5) produção, descritas a seguir:

- 1) Preparação do Sistema: As proteínas estudadas foram simuladas em uma caixa cúbica com uma distância de 15 Å entre o soluto e a caixa, utilizando-se o campo de força GROMOS 53a6 e o modelo de água SPC. O sistema foi neutralizado com íons Sódio (Na^+) e Cloro (Cl^-) até a concentração final de 0,15 mol/L.
- 2) Minimização de Energia (EM): Composta por três etapas. Primeiramente o algoritmo *steepest-decent* foi aplicado para minimizar a energia do solvente (moléculas de água), enquanto a proteína era mantida rígida através da aplicação de um potencial harmônico ($100 \text{ kJ}^{-1}\text{mol}^{-1}\text{nm}^{-1}$). Após esta etapa foi executada uma nova aplicação do *steepest-decent*, agora sem restrições. Então, o sistema foi submetido a uma etapa de minimização de energia utilizando o algoritmo de gradiente conjugado (CG).
- 3) Solvatação do Complexo: Após as etapas de EM o sistema foi submetido a um procedimento que permite a formação das camadas de solvatação ao redor da molécula submetida à simulação de Dinâmica Molecular. O sistema foi mantido aquecido (100 K) durante 50 ps induzindo agitação nas moléculas de água, enquanto a posição da molécula de IDUA foi mantida estável através da aplicação de um potencial harmônico com intensidade de $5.000 \text{ kJ}^{-1}\text{mol}^{-1}\text{nm}^{-1}$. A temperatura foi então reduzida a 20 K, o que foi seguido pela redução gradual da força de restrição sobre os átomos da proteína utilizada na simulação, até o valor mínimo de $0,2 \text{ kJ}^{-1}\text{mol}^{-1}\text{nm}^{-1}$.
- 4) Termalização: Após remoção completa das restrições, o

sistema passa por um aquecimento gradual até atingir a temperatura de 300 K. Ao todo a termalização consiste em sete passos, sendo que a diferença de temperatura entre as etapas foi sempre inferior a 50K.

- 5) Produção: A temperatura foi mantida estável em 300 K durante um período de 47,5 ns e o sistema foi mantido completamente flexível. Isso permite que a proteína se movimente no solvente de forma similar ao que ocorreria *in vivo*, bem como a ocorrência de alterações conformacionais, formação e rompimento de ligações não covalentes, etc. Para fins de análise, foram aproveitadas apenas as medidas realizadas durante o período de produção.

2.1.1.7 Análise dos dados da Dinâmica Molecular

Depois de efetuadas as simulações com os modelos desenvolvidos e com as estruturas cristalográficas de DNMT2 disponíveis, utilizou-se uma série de programas disponíveis no pacote GROMACS que permitem realizar diversas análises sobre os dados gerados na etapa de produção. Dentre essas, destacam-se o desvio médio das estruturas ao longo da simulação (*g_rms*), o raio de giro da molécula (*g_gyrate*), a flutuação média de cada aminoácido ao longo do tempo (*g_RMSF*), entre outras.

2.1.1.8 Análise do potencial eletrostático de superfície

As estruturas cristalográficas e modelos desenvolvidos foram analisadas levando-se em conta o potencial eletrostático de superfície de cada molécula. Para o cálculo da equação de Poisson-Boltzmann, utilizou-se o software *Adaptive Poisson-Boltzmann Solver* (APBS) (Baker *et al.*, 2001), que é capaz de descrever as interações eletrostáticas entre molécula e soluto. O APBS foi rodado como *plugin* nos programas Chimera 1.9 (Pettersen *et al.*, 2004) e VMD

1.9.1 (Humphrey *et al.*, 1996).

Análises comparativas das propriedades eletrostáticas das proteínas utilizadas no presente trabalho foram realizadas através do algoritmo PIPSA (*Protein Interaction Property Similarity Analysis*) (Blomberg *et al.*, 1999). O programa permite a classificação das proteínas de acordo com as propriedades dos campos de interações moleculares, sendo que o potencial eletrostático molecular é o campo de interação molecular mais informativo para esses casos.

2.1.1.9 Análises das relações evolutivas estruturais de DNMT2

As sequências nucleotídicas e de aminoácidos de todas as estruturas empregadas no trabalho foram utilizadas para estabelecer as inferências evolutivas, tomando-se o cuidado de averiguar possíveis sítios mutados nos modelos cristalográficos e substituindo os resíduos mutados pelos selvagens, quando detectados. Para a análise, todas as sequências proteicas foram alinhadas, num primeiro momento, com o auxílio do programa MUSCLE (Edgar, 2004). Além disso, efetuou-se a busca pelas sequências homólogas de DNMT2 das 22 espécies que possuem seus genomas depositados no banco de dados Flybase (<http://flybase.org/>), através do algoritmo BLAST, tendo como entrada a sequência de DNMT2 de *D. melanogaster*.

Através do programa MEGA 6 (Tamura *et al.*, 2013) foi escolhido o melhor modelo evolutivo que se adapta ao grupo de sequências utilizadas, o Critério de Informação de Akaike (AIC) (Sakamoto *et al.*, 1986) foi escolhido como método estatístico. Nesse critério, o valor absoluto do logaritmo da máxima verossimilhança de cada modelo é somado ao número de parâmetros variáveis no modelo (há uma penalidade maior para o maior número de parâmetros existentes), assim o valor obtido é, então, comparado ao dos outros modelos. Aquele modelo que obtiver o menor valor de AIC é considerado o que melhor representa os dados.

As relações filogenéticas foram obtidas pelo método de máxima-

verossimilhança, com *bootstrap* de 10.000 réplicas. Também foram implementadas análises Bayesianas com o programa MrBayes (Ronquist & Huelsenbeck, 2003), com validação de no mínimo 1.000.000 de gerações, onde o parâmetro mínimo para aferição de convergência utilizado foi o desvio padrão médio das frequências. Para as análises de divergência de aminoácidos, as sequências foram agrupadas de acordo com sua ordem taxonômica, utilizando a análise de distância-p implementada pelo programa MEGA 6.

As sequências de aminoácidos também foram alinhadas de acordo com as estruturas terciárias das moléculas. As nove (9) estruturas foram sobrepostas utilizando o *software* Chimera 1.9. A partir da sobreposição das estruturas foi obtido o alinhamento global das sequências primárias. Em seguida foram feitos ajustes manuais no alinhamento das sequências levando-se em conta que regiões de alça acabam por apresentar grandes variações no espaço, o que acaba por induzir a abertura de *gap* em regiões de sequências que possuem alto grau de similaridade. Em regiões em que não houvesse similaridade ou homologia estrutural, adotou-se o distanciamento de 2 Å (angstroms) como penalidade para abertura de *gap*. Após, as relações filogenéticas foram obtidas conforme a metodologia descrita anteriormente.

2.2 Resultados e Discussão

2.2.1 Busca de estruturas cristalográficas

A família das DNMT2 por se diferenciar das demais m5C-MTase em diversos aspectos, acaba por ser um objeto de estudo interessante no âmbito funcional, fisiológico e evolutivo. Como descrito anteriormente, ainda há espaço para discussão quanto ao seu papel principal, já que diversos grupos acabaram por perder o gene *DNMT2*, ou em outros casos, a DNMT2 passou a atuar em diferentes processos extranucleares, como por exemplo, na metilação de citosinas em tRNA, ao longo da evolução. De maneira geral, a DNMT2 é classificada como uma tRNA MTase *in vivo*, porém a generalização de sua funcionalidade vai de encontro a diversos estudos que relatam atividade de metilação em DNA. Além disso, dentro de grupos de espécies muito próximas evolutivamente, a DNMT2 parece desempenhar papéis diferentes, ou com funcionalidades distintas.

Esse é o cenário que encontramos em drosofilídeos, onde recentemente nosso grupo de trabalho identificou padrão de metilação sexo-específica em espécies do grupo *willistoni*, não encontrando nenhuma outra correspondência desse fenômeno em qualquer outra espécie de drosofilídeos fora desse grupo (D'Ávila *et al.*, 2010). Diante do exposto, fez-se interessante uma abordagem analítica estrutural da DNMT2 de *D. melanogaster*, que não possui padrão de metilação sexo-específico, e *D. willistoni* que apresenta tal peculiaridade. Assim, o primeiro passo do presente trabalho foi a modelagem da estrutura 3D da DNMT2 de ambas as espécies, utilizando para isso diferentes metodologias nas buscas de um melhor modelo preditivo.

Utilizando como sequência de entrada a DNMT2 de *D. melanogaster*, efetuou-se a busca através do programa BLAST (*Basic Local Alignment Search Tool*) por estruturas homólogas no Protein Data Bank (PDB). Ao todo o sistema retornou 17 estruturas cristalográficas de DNMT2 (**Tabela 1**). As estruturas que apresentaram maior percentual de identidade foram: a 4H0N (MTase de *S.*

frugiperda) e a 1G55 (MTase de humanos), com identidades de 43% e 41%, respectivamente, sendo utilizadas como estruturas molde para a fase de modelagem computacional por homologia. As demais estruturas foram selecionadas para análises evolutivas posteriores.

Tabela 1. Modelos cristalográficos de DNMT2 depositados no banco de dados Protein Data Bank utilizados para análise e escolha da estrutura molde para o processo de modelagem por homologia e estudos evolutivos.

Família	Identidade <i>D. melanogaster</i> X estrutura (%)	PDB
DNMT2		
sfDNMT2 (<i>S. frugiperda</i>)	43.0	4H0N
DNMT2 humana	41.0	1G55
Ehmeth (<i>E. histolytica</i>)	30.0	3QV2
Hhal (<i>H. haemolyticus</i>)	24.0	1MHT
Mutante S87Q-Q237S	24.0	2UYH
Mutante T250G	24.0	1FJX
Mutante Q237w	24.0	1SVU
Mutante Glu119Gln	23.0	2ZCJ
Mutante Glu119Ala	23.0	2Z6U
Mutante R165N	23.0	2UZ4
Mutante R163N	23.0	2UYC
Mutante Arg165Ala	23.0	2Z6Q
Mutante C81A	23.0	2Z6A
Mutante F124A	23.0	2I9K
HaellI (<i>H. influenzae</i>)	25.0	1DCT
Mutante C71S		3UBT
M.Mpel (<i>M. penetrans</i>)	22.0	4DKJ

A análise de estrutura secundária é de grande importância quando se procura estabelecer um alinhamento de qualidade entre as sequências-alvo e molde durante os estágios prévios de modelagem computacional por homologia (**Figura 8**). A preocupação com um alinhamento adequado se faz pela presença de diversos *indels* entre as diferentes MTases. No caso de *D. melanogaster*,

existe uma inserção de seis (6) aminoácidos em sua sequência primária (resíduos Ile267-Leu281) e que não estão presentes nas sequências-molde selecionadas e nem em *D. willistoni*.

A região não resolvida no modelo cristalográfico da DNMT2 humana, compreendida entre os resíduos 189 – 247, se estrutura em uma sequência de três α -hélices, segundo os resultados do algoritmo PSIPRED (**Figura 9**). Efetuou-se a modelagem computacional da DNMT2 humana incluindo as regiões não resolvidas. Diante dos dados coletados e problemática abordada pelo presente trabalho, seria interessante ter mais uma estrutura 3D de Chordata, assim, a DNMT2 de *M. musculus* foi escolhida para ser modelada, visto que a mesma se faz presente em diversos trabalhos a respeito de metilação de citosinas (Kiani *et al.*, 2013; Tuorto *et al.*, 2012; Thiagarajan *et al.* 2011).

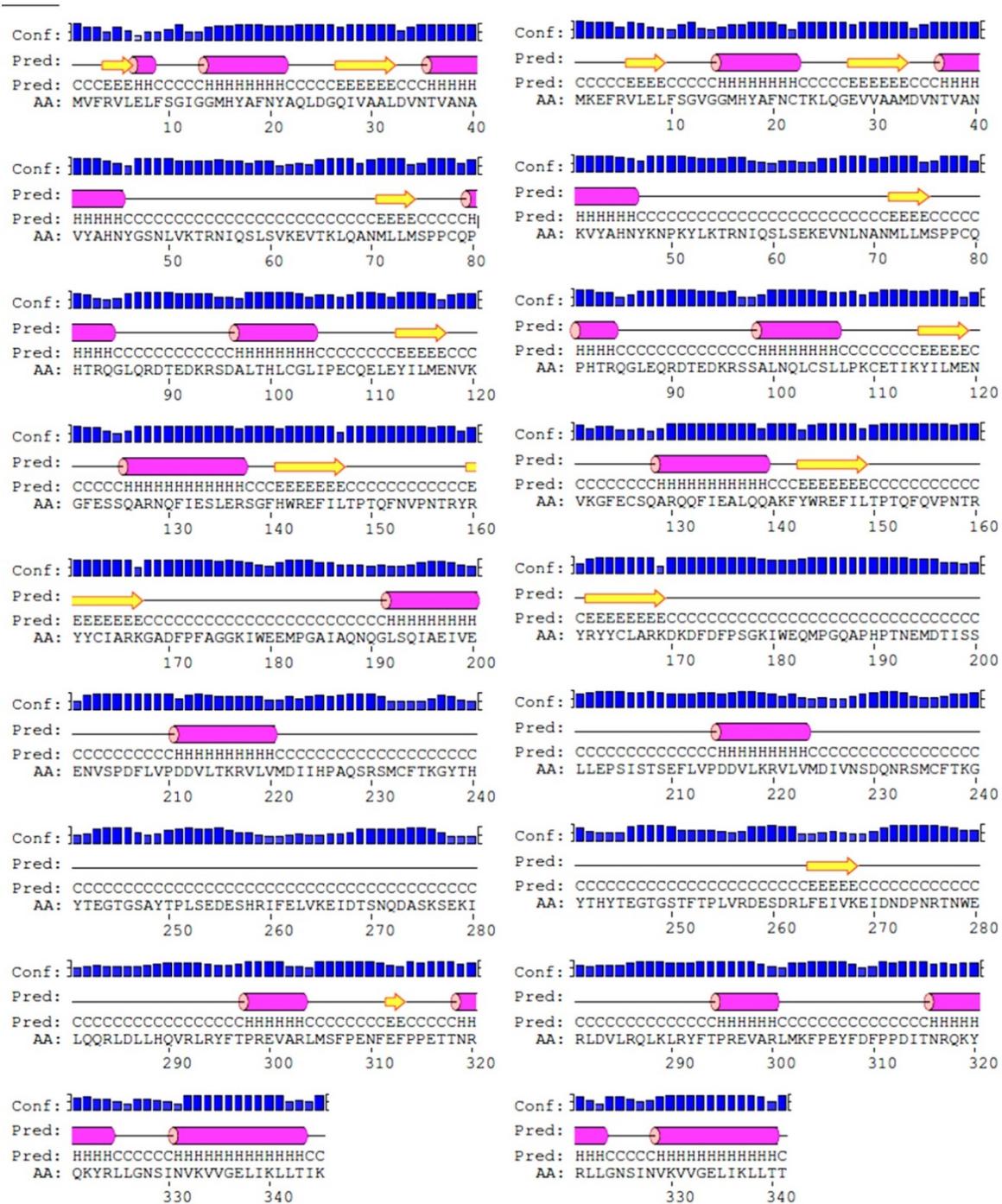


Figura 8. Estruturas secundárias previstas a partir das sequências de DNMT2 de *D. melanogaster* (esquerda) e *D. willistoni* (direita) provenientes da ferramenta PSIPRED. Representação: Folha β (setas amarelas), hélices α (cilindros rosa) e alças (linha preta). Barras na parte superior das estruturas previstas indicam o grau de confiança da previsão.

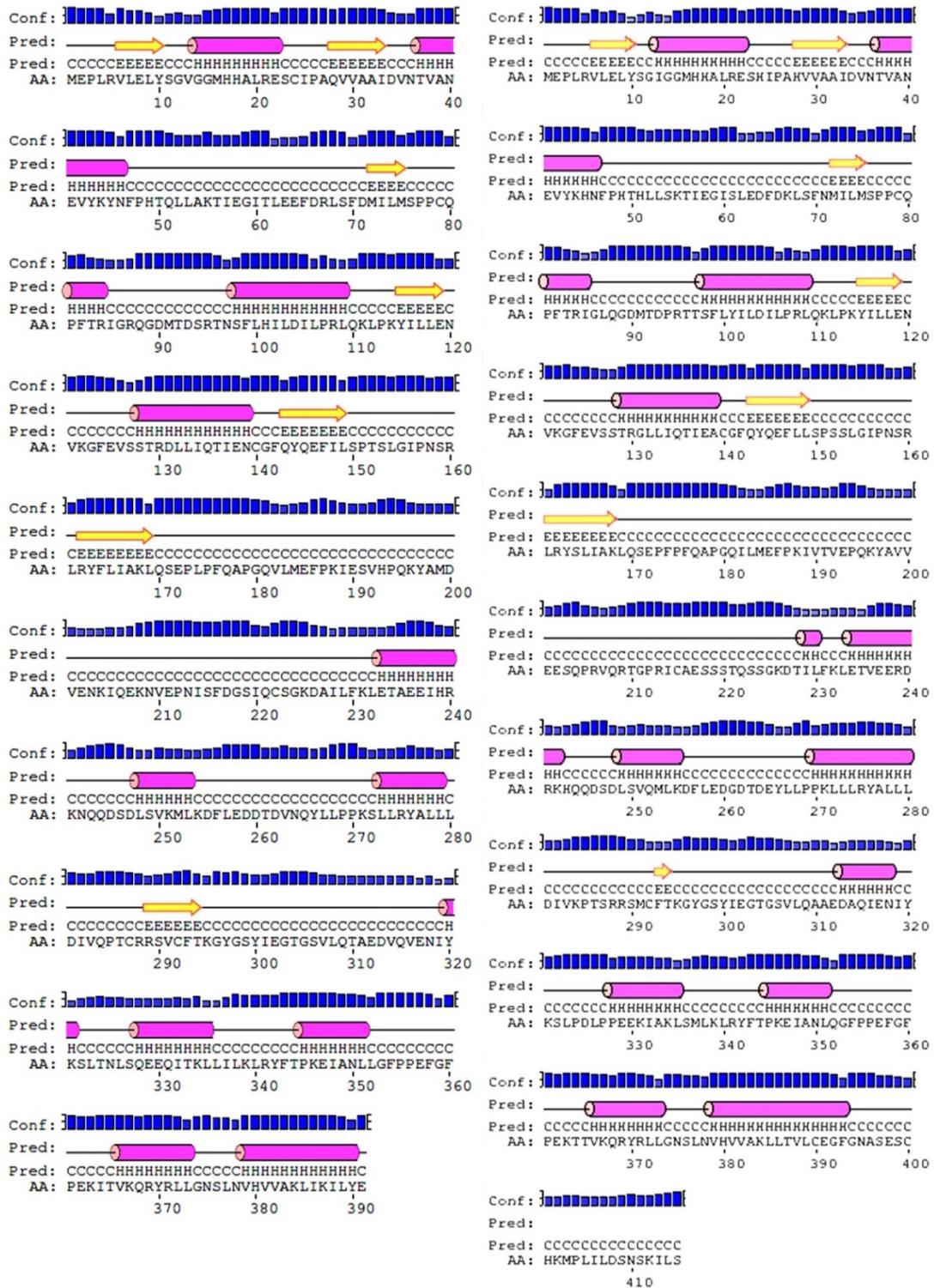


Figura 9. Representação da estrutura secundária predita pelo programa PSIPRED, a partir da sequência de DNMT2 de humanos (esquerda) e de *M. musculus* (direita). Folha β (setas amarela), hélices α (cilindros rosa) e alças (linha preta). Acima da estrutura predita encontra-se o grau de confiança da predição.

2.2.2 Validações e seleção dos modelos desenvolvidos

Ao todo foram gerados e avaliados 52 modelos, 12 modelos de DNMT2 de humanos e 12 de *M. musculus*, em que: 5 modelos resultantes do programa Robetta, 5 modelos do programa I-TASSER, um modelo do programa Swiss-Model e um modelo do programa Modeller, sendo que para a modelagem por homologia utilizando o programa Modeller 9.11, gerou-se 100 modelos para cada espécie, sendo escolhido aquele que apresentasse o melhor valor DOPE e distribuição adequada dos ângulos *phi* e *psi* no mapa de Ramachandran. Para *D. melanogaster* e *D. willistoni* foram gerados 14 modelos para cada espécie, com a diferença que através do programa Modeller foram gerados modelos a partir de três diferentes alinhamentos das sequências-alvo com as sequências molde de: i) com *S. frugiperda* (PDB: 4H0N); ii) com DNMT2 humana (PDB: 1G55); e iii) alinhamento múltiplo com as estruturas cristalográficas de *S. frugiperda* e DNMT2 humana.

Na **tabela 2** estão presentes os escores das validações realizadas com os modelos do presente trabalho. É importante salientar que os modelos das DNMT2 humana e de *M. musculus* tiveram suas validações realizadas excluindo-se a alça de ligação dos domínios I e II (resíduos 189-247) por apresentarem difícil resolução dessa região. Dong *et al.* (2001) só obtiveram sucesso na cristalização após a deleção dessa região. Além disso, a sequência desses resíduos foi expressa e purificada e análises subsequentes demonstraram ser esse um domínio independente, não interagindo fortemente com o restante da proteína.

Os modelos gerados foram ranqueados, de modo que aqueles que figuraram mais vezes entre os melhores colocados em cada um dos validadores empregados no presente trabalho, foram os escolhidos como sendo as estruturas preditivas de DNMT2 de cada espécie.

Tabela 2. Resultados das validações dos modelos obtidos das estruturas de DNMT2. A primeira linha indica as ferramentas utilizadas e a segunda os principais parâmetros analisados. MF – Mais favorável; NP – Não permitido; GP – Generosamente permitido; AP – Adicionalmente permitido. Em cinza: modelos selecionados como preditivos. Modelos que não foram aprovados no Verify 3D não foram validados (N.A.). “RB”, “IT”, “MD” e “SW” referem-se aos programas Robetta, I-TASSER, Modeller e Swiss-Model, respectivamente.

DNMT2	Verify 3D	What If	QMEAN6	PROCHECK				ModFOLD
				Ramachandran (%)				
Modelo	%	Z-score	Z-score	MF	NP	GP	AP	Global Score (P- Value)
<i>D. melanogaster</i>								
mRB1	> 80	-0,845	-1,51	87,7	0,3	0,3	11,7	0,7390 (4,158E-4)
mRB2	> 80	-0,837	-1,89	88,0	0,3	0,6	11,0	0,7346 (4,353E-4)
mRB3	> 80	-1,000	-1,60	87,0	0,3	1,0	11,7	0,7354 (4,317E-4)
mRB4	> 80	-0,811	-2,18	88,0	0,3	1,0	10,7	0,7043 (5,964E-4)
mRB5	> 80	-0,822	-2,16	88,3	1,0	0,6	10,1	0,7342 (4,514E-4)
mIT1	> 80	-1,189	-2,99	83,1	1,9	2,9	12,0	0,7411 (4,068E-4)
mIT2	< 80	-1,440	-3,85	78,9	3,2	4,2	13,6	0,5487 (3,006E-3)
mIT3	> 80	-1,124	-2,87	82,1	1,9	1,9	14,0	0,7097 (5,639E-4)
mIT4	< 80	-1,450	-3,10	80,8	1,9	3,2	14,0	0,6460 (1,093E-3)
mIT5	> 80	-1,137	-2,93	87,7	1,3	1,0	10,1	0,7788 (2,749E-4)
mMD1 (1G55) ^a	> 80	-1,100	-1,57	87,7	0,3	2,3	9,7	0,7286 (4,633E-4)
mMD2 (4H0N) ^a	> 80	-0,807	-2,30	91,6	0,6	0,0	7,8	0,7893 (2,465E-4)
mMD3 (multi) ^d	> 80	-0,859	-2,18	91,2	0,6	1,0	7,1	0,7885 (2,486E-4)
mSW (4H0N) ^a	> 80	-1,055	-2,34	86,6	0,3	1,3	11,8	0,7683 (3,066E-4)
<i>D. willistoni</i>								
wRB1	> 80	-0,842	-2,19	88,6	0,3	0,6	10,4	0,7792 (2,738E-4)
wRB2	> 80	-0,949	-2,11	87	0	0,6	12,3	0,7781 (2,769E-4)
wRB3	> 80	-0,841	-2,05	86,7	0,6	1	11,7	0,7783 (2,764E-4)
wRB4	> 80	-0,753	-1,85	87	0,3	0,3	12,3	0,7779 (2,775E-4)
wRB5	> 80	-0,846	-2,29	87,3	0,6	1,3	10,7	0,7781 (2,769E-4)
wIT1	< 80	-1,096	-2,67	87,9	1,3	1	9,8	0,7844 (2,594E-4)
wIT2	> 80	-1,438	-2,72	81,8	2,9	2,6	12,7	0,7668 (3,115E-4)
wIT3	> 80	-1,431	-2,64	81,8	2,6	2,6	13	0,7656 (3,154E-4)
wIT4	> 80	-1,424	-3,03	78,5	2,3	4,2	15	0,7420 (4,031E-4)
wIT5	> 80	-1,353	-2,65	78,5	2,3	2,6	16,6	0,7597 (3,353E-4)
wMD1 (1G55) ^a	> 80	-1,184	-2,07	86	1	2	11,1	0,7283 (4,647E-4)
wMD2 (4H0N) ^a	> 80	-0,885	-1,78	89,6	0,7	1	8,8	0,7913 (2,414E-4)
wMD3 (multi) ^d	> 80	-0,949	-2,8	91,5	0	0,7	7,8	0,7473 (3,815E-4)
wSW (4H0N) ^a	> 80	-1,267	-2,47	84,5	1,3	2,3	11,8	0,7710 (2,982E-4)

<i>M. musculus</i> ^c								
MmIT1	> 80	1,063	2,40	79.2	1.3	4.9	14.6	0.7250 (5.662E-5)
MmIT2	> 80	1,210	2,22	77.9	1.0	1.9	19.2	0.6987 (6.961E-5)
MmIT3	> 80	1,055	2,47	77.9	1.6	2.9	17.5	0.7259 (5.625E-5)
MmIT4	> 80	1,117	2,46	78.9	1.6	1.3	18.2	0.7013 (6.822E-5)
MmIT5	> 80	1,092	2,42	77.3	1.3	3.6	17.9	0.6994 (6.923E-5)
MmRB1	> 80	0,680	1,92	91.6	0.3	0.3	7.8	0.7006 (6.86E-5)
MmRB2	> 80	0,825	1,48	89.0	0.3	0.0	10.7	0.6920 (7.347E-5)
MmRB3	< 65	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
MmRB4	> 80	0,817	1,37	89.9	0.7	0.6	8.8	0.7013 (6.819E-5)
MmRB5	> 80	0,662	1,31	91.2	0.6	0.3	7.8	0.7027 (6.745E-5)
MmMD (multi) ^d	> 80	1,210	1,91	91.6	0.6	0.3	7.4	0.7802 (2.709E-4)
MnSW	> 80	0,640	1,01	93.4	0.3	0.3	5.9	0.7735 (2.904E-4)
Humana ^c								
HsIT1	> 80	1,072	1,51	73.7	2.7	3.8	19.8	0.7758 (3.865E-5)
HsIT2	> 80	1,019	1,78	75.4	1.4	2.4	20.8	0.7795 (3.763E-5)
HsIT3	> 80	0,946	1,84	77.5	0.7	4.1	17.7	0.7856 (3.601E-5)
HsIT4	> 80	0,984	1,89	79.7	1.4	2.0	17.1	0.7127 (6.233E-5)
HsIT5	< 65	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
HsRB1	> 80	0,469	0,88	90.1	0.0	0.7	9.2	0.7733 (3.939E-5)
HsRB2	> 80	0,646	0,70	89.8	0.3	0.3	9.6	0.7259 (5.625E-5)
HsRB3	< 65	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
HsRB4	> 80	0,517	1,13	89.4	0.7	1.4	8.5	0.7735 (3.931E-5)
HsRB5	> 80	0,558	0,68	88.1	0.0	1.7	10.2	0.7723 (3.966E-5)
HsMD (multi) ^b	> 80	0,705	1,25	93.2	0.3	0.7	5.8	0.7817 (3.705E-5)
HsSW	> 80	0,495	0,71	89.7	0.0	1.4	8.9	0.7674 (4.112E-5)

^a Anotação de estrutura cristalográfica utilizada como molde. ^b Modelagem por homologia efetuada com múltiplos moldes (4H0N e 1G55). ^c Modelos de DNMT2 humana e *M. musculus* validados sem a alça referente aos resíduos 189-247. Os validadores WHAT IF e PROCHECK possuem peso 1, enquanto QMEAN6 e ModFOLD por serem validadores robustos, possuem peso 2 para o ranqueamento dos modelos.

Assim, o modelo selecionado para *D. melanogaster* foi o mMD2 e para *D. willistoni*, o wRB1. Para *M. musculus*, o modelo selecionado foi o MmSW e para humanos o modelo HsRB1. Além das estruturas modeladas foram utilizadas as estruturas cristalográficas de *H. haemolyticus* (PDB 1MHT), *H. influenzae* (PDB 1DCT), *M. penetrans* (PDB 4DKJ), *E. histolytica* (PDB 3QV2) e *S. frugiperda* (PDB 4H0N) totalizando nove (9) estruturas analisadas ao longo do

trabalho (Figura 10).

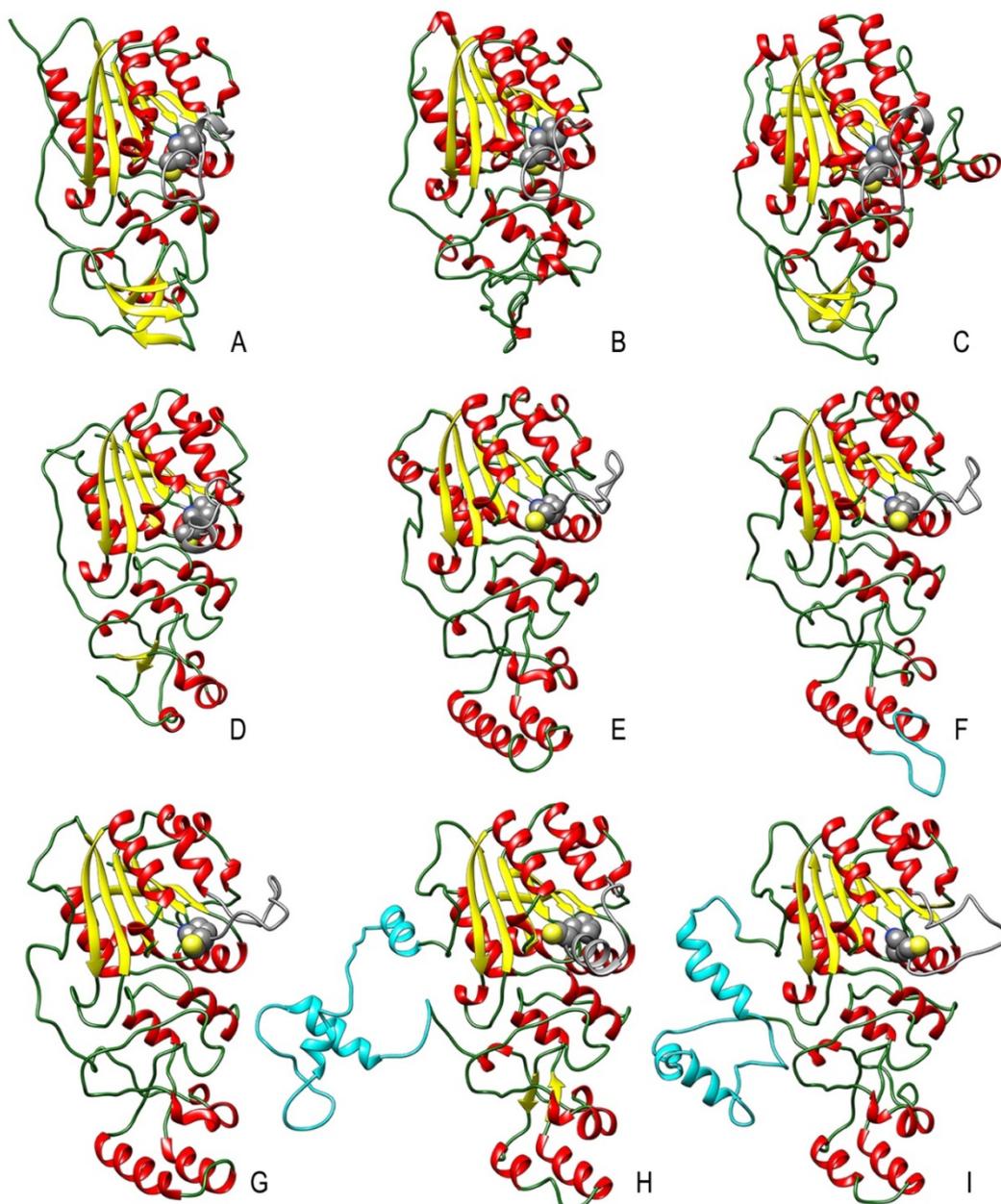


Figura 10. Representação em *ribbon* das estruturas 3D de DNMT2. Estruturas alfa-hélices (vermelho), folhas-beta (amarelo) e alças (verde). A alça do domínio catalítico está colorida em cinza e os átomos do dipeptídeo ProCys do sítio ativo representados como esferas (carbonos em cinza, enxofre em amarelo e nitrogênio em azul). (A) *H. haemolyticus* (PDB 1MHT); (B) *H. influenzae* (PDB 1DCT); (C) *M. penetrans* (PDB 4DKJ); (D) *E. histolytica* (PDB 3QV2); (E) *D. willistoni* (wRB1); (F) *D. melanogaster* (mMD2), em ciano a alça contendo a inserção dos resíduos 267-281; (G) *S. frugiperda* (PDB 4H0N); (H) *M. musculus* (MmSW); e (I) Humanos (HsRB1), em

ciano região não conservada referente aos resíduos 189 – 247 presentes apenas em DNMT2 de mamíferos.

2.2.3 Análises estruturais, físico-químicas e evolutivas das DNMT2

A família DNMT2 caracteriza-se por possuir dez motivos, sendo que cinco são altamente conservados (I, IV, VI, VIII e X) (Kumar; et al., 1994). Entre os motivos VIII e IX encontra-se o domínio TRD, que se apresenta substancialmente variável entre diferentes espécies. Assim, pode se dividir as enzimas da família DNMT2 em dois grandes domínios: o Domínio-I, que se caracteriza por possuir os motivos consensos do arcabouço de ligação da AdoMet e o sítio ativo do dipeptídeo ProCys, responsável pela aproximação da citosina-alvo a ser metilada à estrutura AdoMet; e o Domínio-II, que é responsável pela interação com o substrato-alvo (DNA ou tRNA), possuindo como característica diversidade não apenas em nível de sequência, mas também estrutural entre as diferentes enzimas da família DNMT2.

O domínio-I está estruturado em um clássico empacotamento de Rossmann. Esse empacotamento proteico caracteriza-se pela alternância do motivo folha- β / α -hélice/folha- β em sua estrutura secundária e é comumente encontrada em enzimas que interagem com nucleotídeos, no caso a AdoMet. Já o domínio-II se caracteriza por não possuir uma arquitetura rica em estruturas secundárias, sendo composta predominantemente por alças.

Ao todo, a homologia das sequências das DNMT2 estudadas não é alta, porém existem diversos motivos bem conservados, principalmente no domínio catalítico. Pode se destacar alguns motivos bastante relevantes, como o motivo III, contendo o dipeptídeo catalítico ProCys (**Figura 11**). A grande parte dos resíduos conservados localiza-se em regiões de interações com o substrato ou ligação com o cofator (AdoMet). Entre eles, o motivo ENV e a arginina conservada do motivo V, que são de extrema importância para a atividade catalítica da enzima, podendo haver comprometimento da atividade enzimática quando na ocorrência de mutações (Sankpal, 2002; Wyszynski *et al.*, 1993) (**Figura 12**).

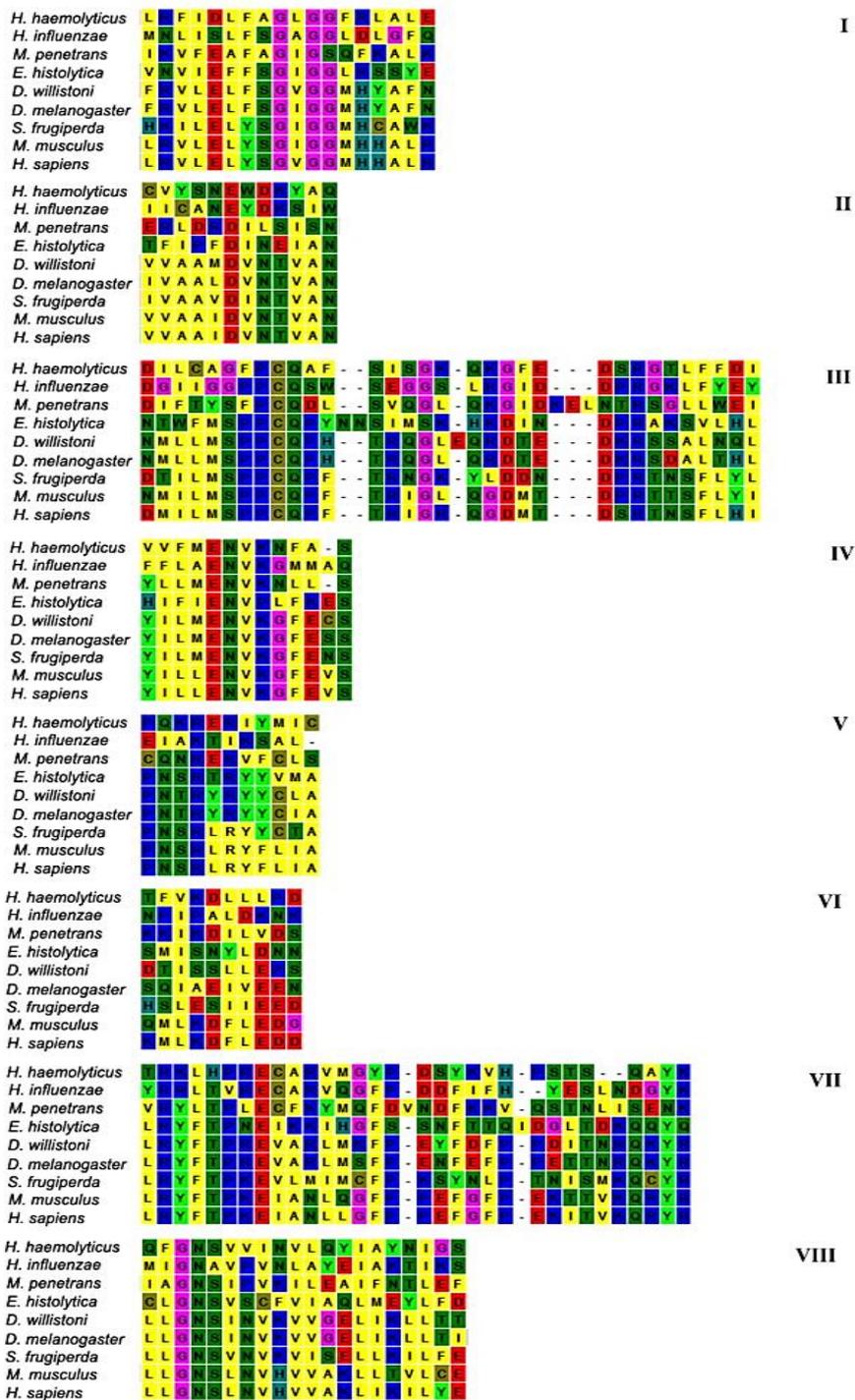


Figura 11. Sequência de oito motivos conservados presentes no domínio catalítico das estruturas analisadas.

No motivo I encontra-se a sequência F/YxGxG que está envolvida na ligação com o cofator AdoMet (*S. frugiperda*, *M. musculus* e humana possuem tirosina, ao invés de fenilalanina). Essa diferença provavelmente seja substancial

na relação proteína-cofator, pois a fenilalanina é um aminoácido não polar, enquanto a tirosina é uma aminoácido polar, o que deve ter implicação na forma de interação da AdoMet em seu arcabouço de ligação com a proteína. É interessante notar que a dinâmica existente entre cofatores e substratos com a proteína pode ser preponderante para sua atividade, podendo interferir na função da mesma. Evans & Bronowska (2010) sugerem que há uma rede de interações dinâmicas em DNMTs que podem regular a termodinâmica das interações proteína-cofator/proteína-substrato e, talvez, as sequências-alvo a serem metiladas.

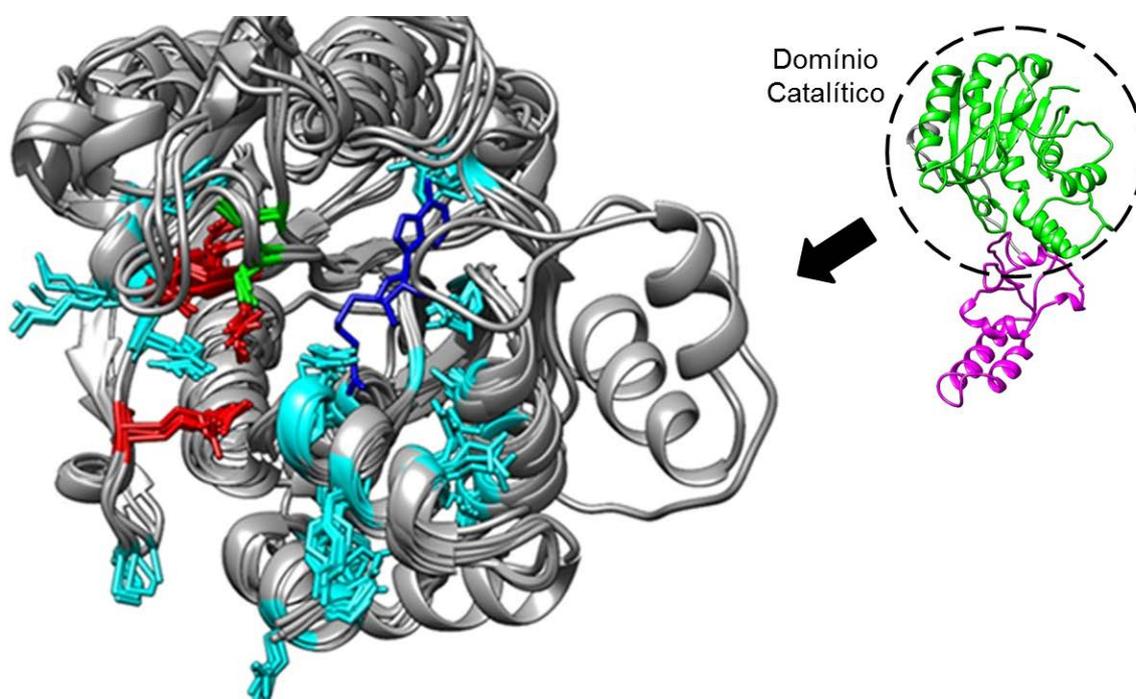


Figura 12. Sobreposição do domínio catalítico das nove DNMT2 analisadas. Cisteína catalítica colorida em verde, o motivo conservado ENV e a arginina do motivo V, que são cruciais para a atividade catalítica da enzima, estão coloridos em vermelho. Os demais resíduos conservados e/ou envolvidos com interações diretas tanto com o cofator, como o substrato, estão coloridos em ciano. O cofator SAM está colorido em azul. À direita representação da estrutura completa da DNMT2 de *D. willistoni*, onde o domínio catalítico está colorido em verde e o domínio contendo a TRD, em magenta.

Enquanto o domínio catalítico caracteriza-se por possuir uma relativa conservação estrutural entre as diversas DNMT2, as TRDs (região pertencente ao domínio menor) das DNMTs possuem diferenças marcantes em suas estruturas

terciárias, além da alta diversidade em sua composição de resíduos (**Figura 13**). Entre as próprias DNMT2 de procariotos há diferenças importantes, tanto na composição, como na arquitetura e no processo de reconhecimento da sequência-alvo. Essas diferenças se refletem nos sítios preferenciais de metilação dessas espécies. Em *H. haemolyticus* e *M. penetrans*, por exemplo, há a preferência por DNA hemimetilado, em dinucleotídeos CpG no contexto 5'-GCGC-3' (O'Gara *et al.*, 1996; Wojciechowski *et al.*, 2012). Já *H. influenzae* realiza a metilação de citosinas envolvidas em sítios 5'-GGCC-3' (Reinisch *et al.*, 1995). Além disso, os mecanismos para o reconhecimento, ação sobre a citosina-alvo e estabilização do DNA também são bem diversos em cada uma das enzimas. Entretanto, o que existe em comum nas DNMT2 de procariotos é o grande número de resíduos que interagem de alguma maneira com o substrato, em *Mhal*, por exemplo, há mais de 20 aminoácidos envolvidos na interação com o DNA (O'Gara *et al.*, 1996).

As diferenças estruturais das TRDs são mais acentuadas quando se compara as DNMT2 de procariotos com eucariotos. A TRD de *E. histolytica* assume uma arquitetura intermediária entre os dois reinos, mas possui uma sobreposição mais justaposta com as demais TRDs de eucariotos. A similaridade também está presente na conservação de alguns motivos importantes, que estão envolvidos no reconhecimento da sequência-alvo a ser metilada do substrato, como os motivos CFTxxYxxY/I (onde "x" pode ser qualquer outro resíduo), E/DGTGS e DIV/I/C. Diferentemente das enzimas procarióticas, as TRDs de DNMT2 eucarióticas possuem menos resíduos preditos na ação de reconhecimento e interação com o substrato. Estudos com a MTase de *E. histolytica* (EhMeth) demonstraram a dualidade de reconhecimento de substratos nesse grupo de DNMT2, onde as moléculas de DNA e a região composta pela haste contendo o anticódon do tRNA se ligam de maneira muito similar à enzima (Schulz *et al.*, 2012). Assim, as diferenças encontradas tanto na composição, como arquitetura das DNMT2 estudadas, estariam ligadas às diferenças de funções biológicas e especificidades observadas nas DNMT2 de diferentes espécies, onde até mesmo em espécies muito próximas evolutivamente (como *H.*

haemolyticus e *H. influenzae*) há alta divergência na composição, na estrutura e nos mecanismos de ação da TRD sobre o substrato.

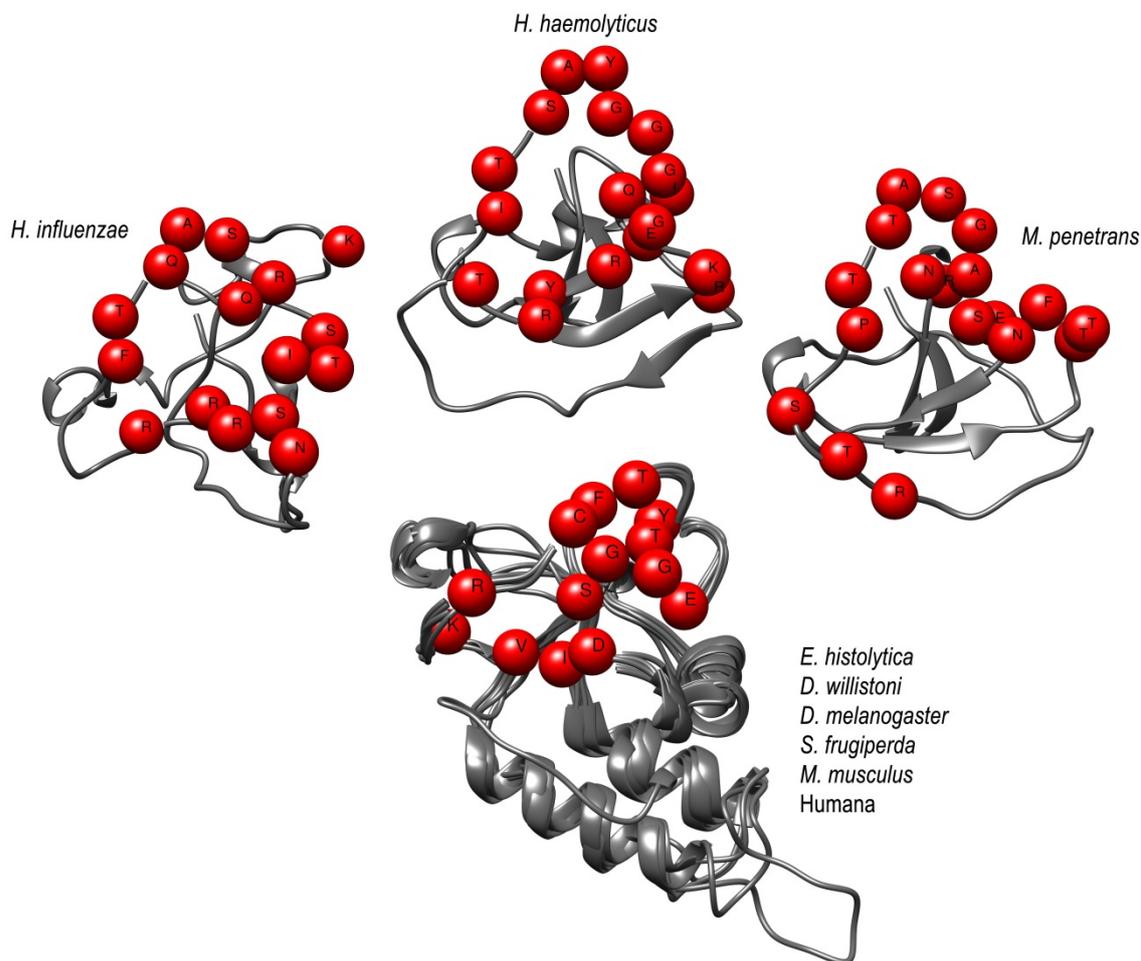


Figura 13. Representação das estruturas terciárias das TRDs das DNMT2 analisadas. Os resíduos descritos como aqueles que interagem direta ou indiretamente com o substrato (DNA/tRNA), tendo a função de reconhecimento da sequência alvo a ser metilada, estão representados como esferas vermelhas.

Analisando a composição geral das diferentes DNMTs analisadas, nota-se a prevalência de resíduos não polares (40,2%) e polares (33,6%) (**Figura 14**). A prevalência desses aminoácidos está relacionada principalmente com a estruturação tridimensional da proteína. Entretanto, quando analisados de maneira relativa cada uma das enzimas, emergem diferenças interessantes na composição dos resíduos. A MTase de *H. haemolyticus* (Hhal) possui 14,07% de

resíduos básicos, contra 11,93% de ácidos, já *H. influenzae* (Haelll) possui 16,05% de resíduos básicos e 10,49% de resíduos ácidos.

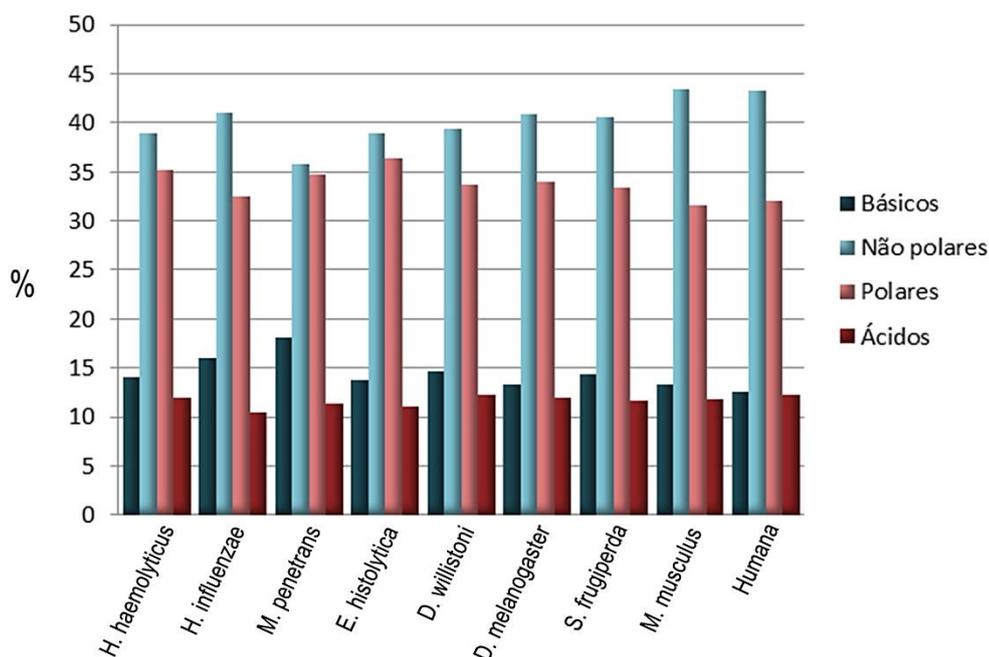


Figura 14. Distribuição das frequências de aminoácidos quanto às suas propriedades nos modelos de DNMT2. Todas as frequências estão expressas em percentual.

A DNMT2 de *M. penetrans* possui a composição de resíduos mais peculiar entre as DNMTs, possuindo um grande percentual de aminoácidos básicos (18,11%), enquanto resíduos ácidos fazem parte de 11,41% da proteína. A prevalência entre os resíduos básicos é de lisina (12,7%), além de um percentual significativo (10,9%) de asparagina, que geralmente está presente no início e na terminação de estruturas secundárias α -hélice (**Figura 15**). Para a DNMT2 de *E. histolytica*, existe uma diminuição de resíduos básicos (13,76%) quando comparado com o percentual desses resíduos às DNMT2 de procariotos.

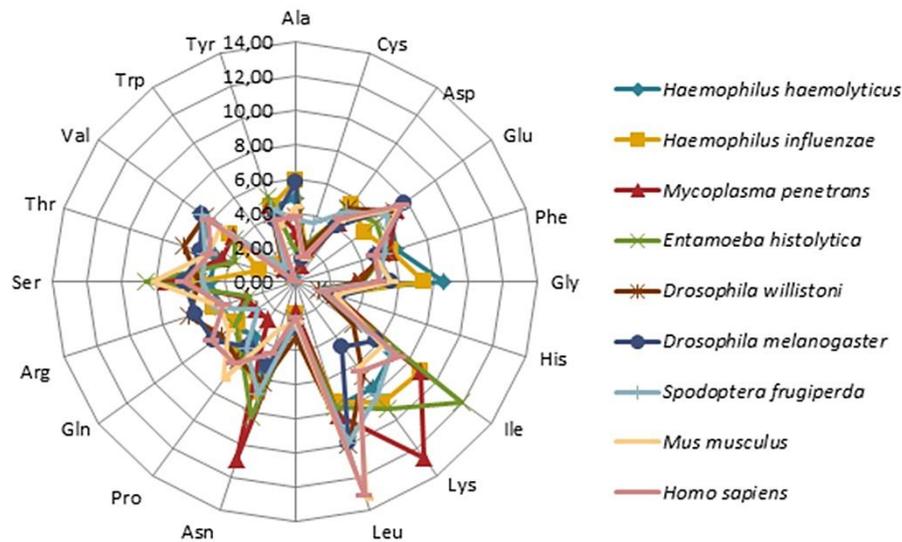


Figura 15. Distribuição das frequências de aminoácidos das DNMTs. Todas as frequências expressas em percentual.

Em *D. willistoni* tem-se um aumento no percentual de aminoácidos básicos (14,16%), percentual semelhante às DNMTs de procaríotos, e 12,32% de resíduos ácidos - esses percentuais também são similares à *sfDNMT2*, 14,41% e 11,71% respectivamente. Tal similaridade se torna interessante ao passo que estudos recentes com o genoma de *S. frugiperda* demonstraram padrões de distribuição de ilhas CpG e %CG parecidos com os de *Bombyx mori* e *Danaus plexippus*, o que indica padrões de organização genômica semelhantes. Essa uniformidade em espécies próximas de Lepidoptera pode estar associada à manutenção dos sistemas de controle de expressão gênica semelhantes, como a ação de DNA MTases, por exemplo (Kakumani *et al.*, 2014). Entretanto, a DNMT2 de *D. melanogaster* possui uma menor ocorrência de resíduos básicos (13,3%) e de resíduos ácidos (11,88%). Por sua vez, esses valores estão mais próximos às composições das DNMTs humana (12,53% resíduos básicos e 12,28% resíduos ácidos) e de *M. musculus* (13,25% resíduos básicos e 12,28% resíduos ácidos), que possuem clara função biológica preferencial na metilação de citosinas em tRNA^{asp}.

A composição de aminoácidos difere, também, entre o domínio catalítico e a TRD das DNMT2. De modo geral, no domínio catalítico há uma prevalência de aminoácidos não polares (41%) em relação aos polares (33,4%), enquanto que na TRD essa relação é de 38,6% e 34,1%, respectivamente. Contudo, enzimas classificadas como DNA MTase na literatura possuem frequência de resíduos não polares de 35,8% e polares de 38,6%, na TRD, enquanto que nas em tRNA MTase essa relação se inverte, havendo a prevalência de não polares com 39,4%, contra 32,5% de resíduos polares.

Através do desvio quadrático médio (RMSD) é possível mensurar quantitativamente as similaridades tridimensionais das estruturas proteicas, tomando como parâmetro as coordenadas dos C α após a elaboração da sobreposição dos confôrmeros. A **tabela 3** apresenta os valores de RMSD entre as DNMTs, levando-se em consideração o cálculo das proteínas inteiras e também apenas do domínio catalítico. Entre as estruturas inteiras, aquelas que apresentaram o menor RMSD foram as de *S. frugiperda* e *D. melanogaster* (mMD2) com um desvio de 0,396 Å. O baixo valor obtido é resultado do método de restrições químicas e espaciais empregado pelo programa Modeller para a elaboração do modelo, o que também representa o estabelecimento de um bom alinhamento entre as sequências alvo e molde. O desvio entre os modelos de *D. melanogaster* e *D. willistoni* (wRB1) foi de 0,718 Å para a proteína completa e 0,547 Å considerando apenas o domínio catalítico. O valor de RMSD entre as estruturas de *D. willistoni* e *S. frugiperda* também foi um dos mais baixos, 0,692 Å para proteína inteira e 0,468 Å entre os domínios catalíticos.

Tabela 3. RMSD referente comparação pareada dos alinhamentos estruturais dos modelos de Dnmt2. Diagonal inferior RMSD da proteína total. Diagonal superior, RMSD do domínio catalítico.

	1	2	3	4	5	6	7	8	9
1 <i>H. haemolyticus</i>		1,226	1,782	1,871	1,589	1,636	1,619	1,611	1,635
2 <i>H. influenzae</i>	2,093		1,958	2,177	1,885	1,938	1,929	1,902	1,924
3 <i>M. penetrans</i>	2,114	2,609		1,972	1,758	1,786	1,749	1,776	1,799
4 <i>E. histolytica</i>	2,221	2,632	2,383		1,335	1,239	1,215	1,093	1,122
5 <i>D. willistoni</i>	2,150	2,406	2,163	1,742		0,547	0,468	0,891	0,889
6 <i>D. melanogaster</i>	2,146	2,440	2,167	1,654	0,718		0,256	0,853	0,833
7 <i>S. frugiperda</i>	2,104	2,422	2,106	1,649	0,692	0,396		0,796	0,802
8 <i>M. musculus</i>	2,246	2,661	2,368	1,723	1,387	1,394	1,337		0,374
9 Humana	2,271	2,638	2,307	1,676	1,449	1,430	1,386	1,182	

Os maiores desvios encontrados foram entre DNMTs de *M. musculus* e humana em relação à estrutura de *H. influenzae*, com um RMSD de 2,638 Å entre DNMT2 humana e *H. influenzae* e 2,661 Å entre *M. musculus* e *H. influenzae*. Entretanto, essas diferenças subjazem principalmente no que se refere ao domínio composto pela TRD, quando se leva em consideração apenas o domínio catalítico, os maiores desvios encontrados são entre *H. influenzae* e *E. histolytica* (2,177 Å) e entre *H. influenzae* e *M. penetrans* (1,958 Å). Interessantemente, as estruturas cristalográficas de procariotos possuem maiores RMSD entre si do que as estruturas de eucariotos. Além disso, quando observados os desvios dos domínios catalíticos entre DNMT2 de drosofilídeos em relação às DNMT2 de mamíferos há diferenças por volta de 0,866 Å, em média.

Outra análise empregada nas estruturas de DNMT2 foi a de potencial eletrostático de superfície. As interações eletrostáticas de uma proteína se caracterizam como forças atrativas de longo alcance, capazes de interferir na velocidade de associação entre moléculas (como proteína-cofator e proteína-substrato). Por serem forças de longo alcance e fortes, o conhecimento de como o potencial eletrostático se distribui na superfície de uma proteína acaba sendo crítico para o entendimento do comportamento e função de uma molécula. Através do aplicativo APBS foi possível efetuar o cálculo da equação de Poisson-Boltzmann para cada estrutura (**Figura 16**).

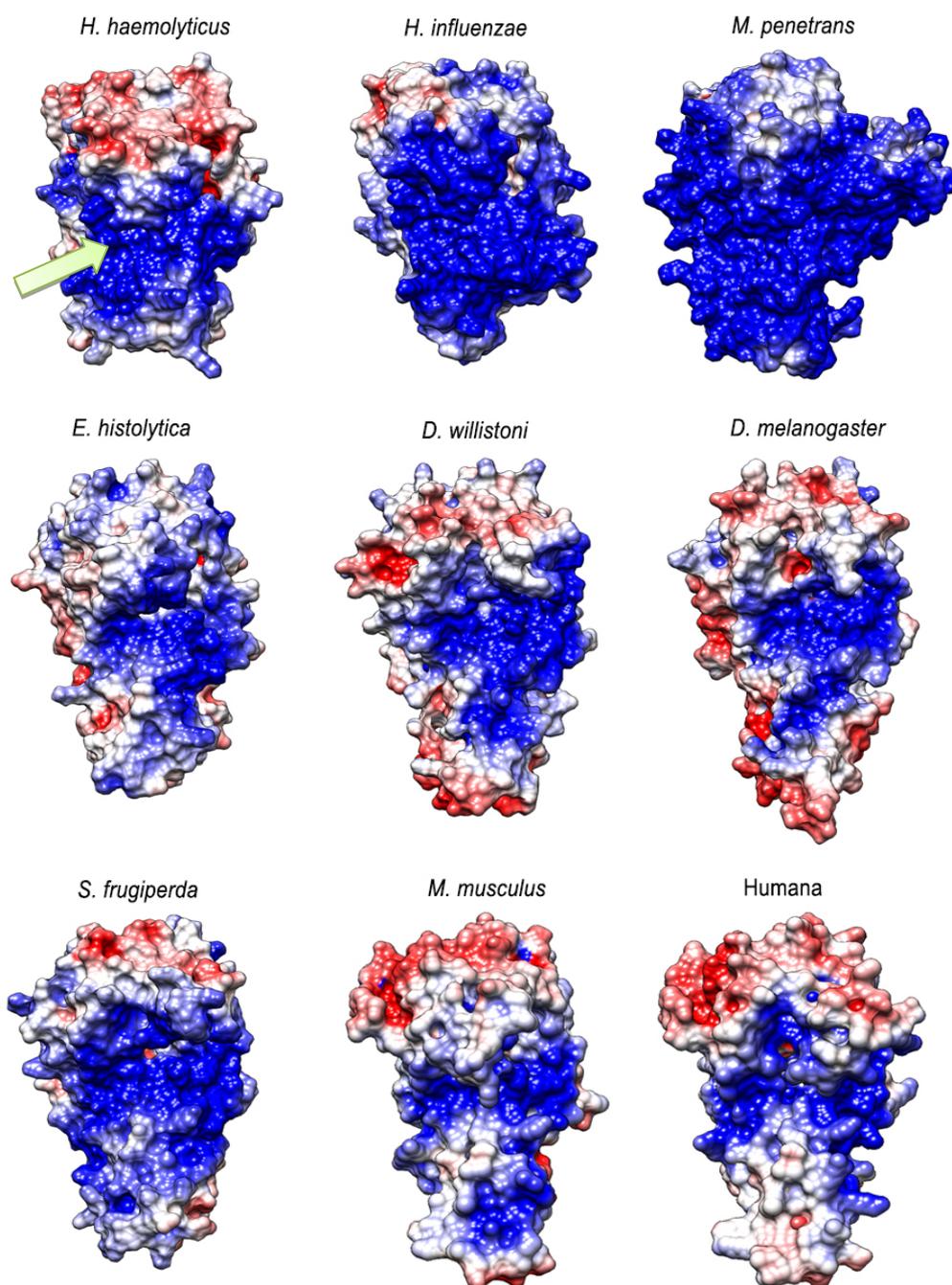


Figura 16. Representação do potencial eletrostático de superfície das estruturas de DNMT2. As superfícies estão coloridas de acordo com o potencial eletrostático, em azul para potencial positivo ($5 kT$), vermelho para negativo ($-5 kT$) e branco para neutro, onde k representa a constante de Boltzmann e T , a temperatura. As estruturas de DNMT2 humana e de *M. musculus* estão representadas sem a alça de ligação dos domínios I e II (resíduos 189 – 247) das MTase2 de mamíferos. A TRD (seta verde) localiza-se na região mais central da molécula, com característica predominantemente de potencial eletrostático positivo.

A composição dos resíduos de aminoácidos acaba se refletindo na distribuição do potencial eletrostático de superfície nas DNMT2 analisadas. O elevado percentual de resíduos básicos presentes nas enzimas de *M. penetrans*, por exemplo, acaba conferindo uma acentuada distribuição de cargas positivas na superfície da molécula, como se pode observar na **figura 16**. O perfil de distribuição de cargas positivas e negativas nas moléculas analisadas acaba por corroborar com a funcionalidade das mesmas em relação à afinidade por determinado substrato: DNA ou tRNA. As DNMT2 classificadas como DNA MTase possuem como característica uma ampla superfície eletropositiva, principalmente na TRD, região onde há a interação entre a proteína e o seu substrato. O mesmo padrão ocorre em *EhMeth*, onde a diminuição de resíduos básicos reflete a redução da área de potencial eletropositivo da enzima. DNMT2 de mamíferos, por serem prioritariamente tRNA MTases, apresentam a disposição de seus potenciais eletrostáticos positivos delimitados a uma pequena área da TRD.

Entre as DNMT2 de drosofilídeos é possível visualizar diferenças da densidade de cargas positivas nas TRDs de *D. willistoni* e *D. melanogaster*. *D. willistoni* visualmente apresenta uma maior área de potencial eletrostático positivo do que a DNMT2 de *D. melanogaster*. Como discutido anteriormente, espécies do grupo *willistoni* apresentam padrões de metilação sexo-específica (Garcia *et al.*, 2007; D'Ávila *et al.*, 2010), fenômeno peculiar a esse grupo e sem recorrência nas demais espécies de drosofilídeos. Logo, poderia se esperar algum tipo de diferença de afinidade junto ao DNA entre as DNMT2 das duas espécies, o que acaba se refletindo nas diferenças das áreas de distribuição de cargas positivas na TRD de *D. willistoni* e *D. melanogaster*.

A análise visual dos potenciais elétricos de superfície restringe-se ao âmbito qualitativo, porém através do programa PIPSA é possível estabelecer uma análise quantitativa das similaridades das propriedades de interações das proteínas. A **Figura 17** apresenta o epograma referente à análise de toda a superfície das nove (9) estruturas estudadas, nele é possível observar três agrupamentos principais. No primeiro agrupamento estão presentes as estruturas da DNMT2 humana e de *M. musculus*. No segundo agrupamento, têm-se as

estruturas de *H. haemolyticus*, *M. penetrans* e *H. influenzae*. No último agrupamento, encontram-se as DNMT2 de *E. histolytica*, *S. frugiperda*, *D. melanogaster* e *D. willistoni*.

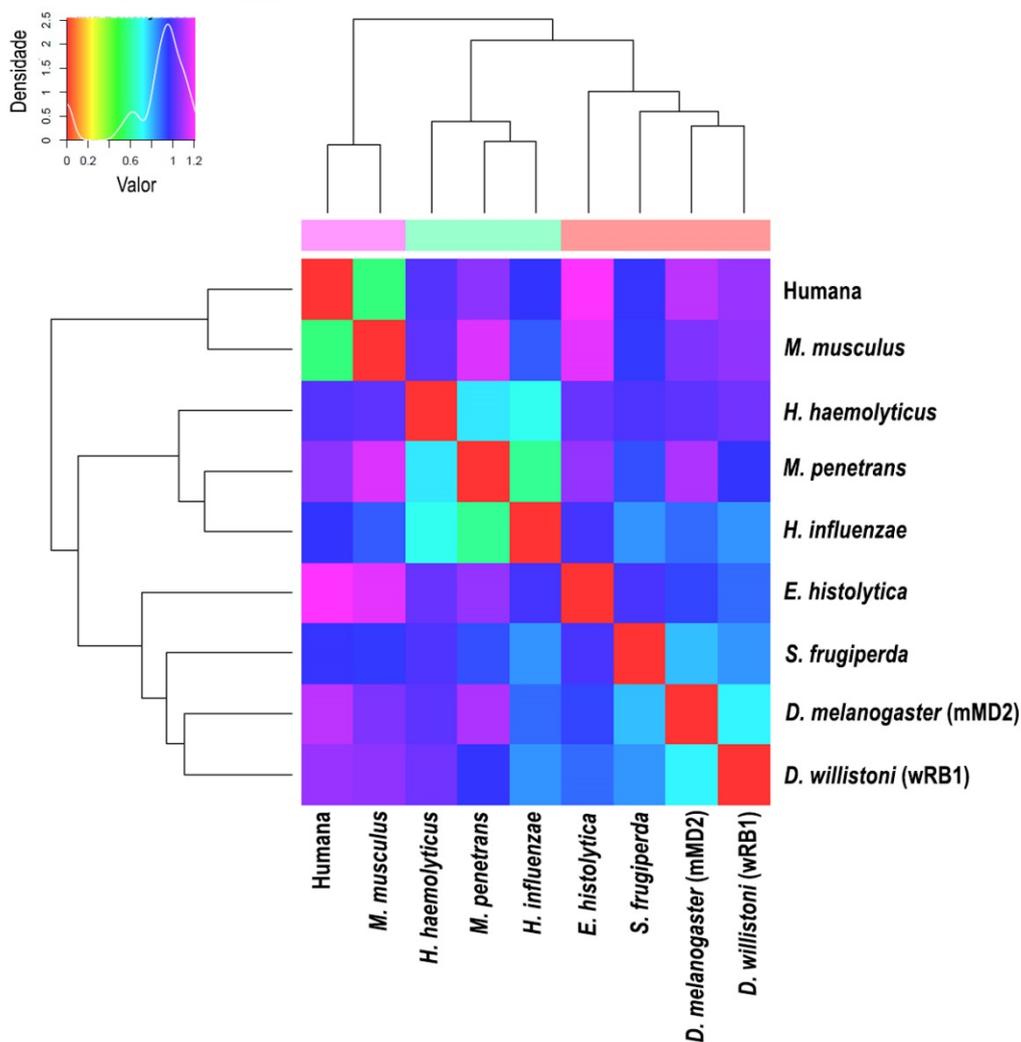


Figura 17. Epograma da análise por PIPSA (*Protein Interaction Property Similarity Analysis*). As estruturas estão agrupadas segundo a relação de similaridade dos potenciais eletrostáticos a partir da matriz de distância elaborada pelo programa. Acima à esquerda encontra-se a chave de cores, indicando a relação de similaridade, quanto menor o valor, maior similaridade entre as estruturas.

Drosophila willistoni, *D. melanogaster* e *S. frugiperda* demonstram potenciais eletrostáticos similares, já a MTase humana e de *M. musculus* apresentam-se distantes em valores de similaridade das demais DNMT2. *D. melanogaster* e *D. willistoni* parecem compartilhar potenciais eletrostáticos de superfície semelhantes, segundo a análise PIPSA, entretanto, é interessante

notar que *D. willistoni* possui valores de similaridade maiores com a DNA MTase de *H. influenzae*, do que *D. melanogaster* e *H. influenzae*.

Para inferir as relações funcionais em conjunto com os aspectos evolutivos, realizaram-se análises das sequências de resíduos de aminoácidos e nucleotídeos das DNMT2 pelos métodos de máxima-verossimilhança e Bayesiana. Partindo da hipótese de que as relações evolutivas são inferidas em análises filogenéticas por algoritmos que levam em conta diferentes modelos de evolução, que por sua vez, conferem pesos e tratamentos distintos aos processos de substituições sinônimas, não-sinônimas, deleções e inserções, utilizaram-se no presente estudo duas abordagens para os alinhamentos das sequências. A primeira abordagem baseia-se no protocolo padrão de análises filogenéticas, com o processo de alinhamento das sequências através de um programa (MUSCLE), avaliação do modelo evolutivo a ser empregado e construção das árvores contendo as relações evolutivas das espécies analisadas. A segunda abordagem empregada se diferencia em relação ao tratamento do alinhamento das sequências, nesse caso, levou-se em conta como estavam organizadas as estruturas terciárias das moléculas e essas informações eram levadas em conta na elaboração do alinhamento das sequências, o que se convencionou denominar esse tipo de alinhamento de 3D-1D.

Na **figura 18**, têm-se as inferências evolutivas das sequências de resíduos pelo método de máxima-verossimilhança. Pode se observar que o nó de divergência entre o clado composto pelas DNMT2 de *D. willistoni* e *D. melanogaster* e o clado que possui as sequências de *S. frugiperda*, *M. musculus* e humana, apresenta melhoria no *bootstrap* da árvore onde se utilizou o alinhamento híbrido (3D-1D). O ponto de divergência entre *H. haemolyticus* e *H. influenzae* também obteve um valor melhor de *bootstrap* (de 83 para 99). Além disso, o valor do nó de divergência entre *S. frugiperda* e as sequências de *M. musculus* e humana obteve grande ganho de robustez, resolvendo o agrupamento composto pelos artrópodes e vertebrados quando utilizado o alinhamento 3D-1D. É interessante notar como as filogenias possuem graus de similaridade com a organização presente no epograma das estruturas, com a

exceção de que no epograma, *S. frugiperda* encontra-se agrupada com *D. willistoni* e *D. melanogaster*.

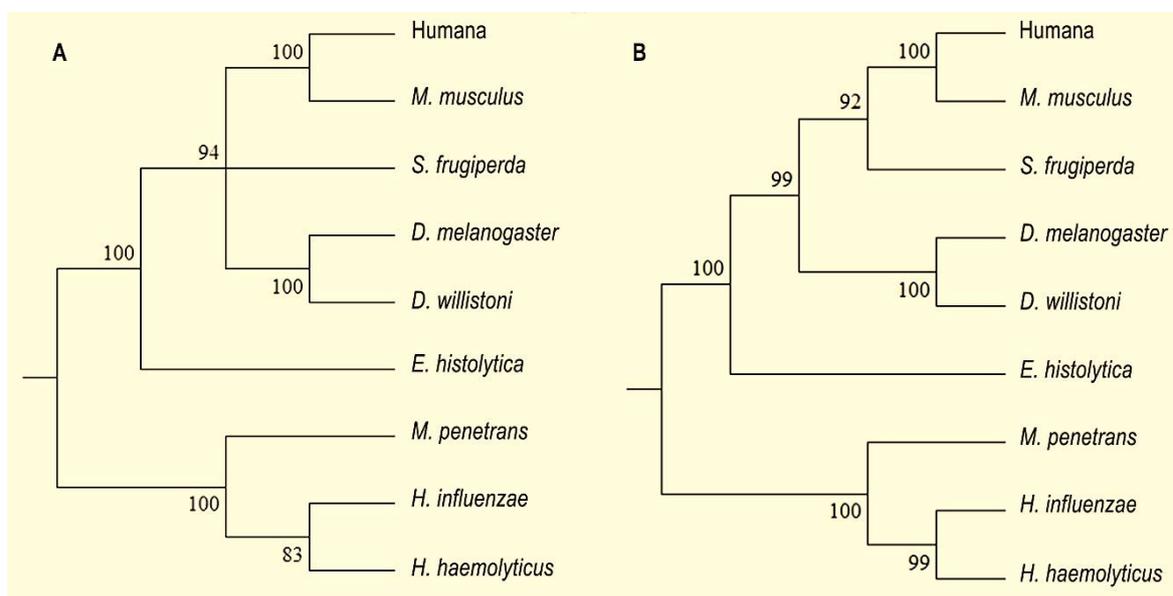


Figura 18. Análise filogenética pelo método de máxima verossimilhança das sequências de aminoácidos. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas, sendo tomada como representativa da história evolutiva dos taxos analisados. Os ramos que apresentaram *bootstrap* menor do que 75 foram condensados. O modelo evolutivo utilizado foi o de Le Gascuel (LG) com distribuição gama discreta com 4 categorias. **(A)** árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; **(B)** árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2.

Quando feitas as análises com as sequências nucleotídicas (**Figura 19**), obtiveram-se resultados semelhantes na melhoria da resolução das relações evolutivas das espécies, onde no alinhamento padrão não havia resolvido o clado composto por *H. haemolyticus*, *H. influenzae* e *M. penetrans*, o alinhamento 3D-1D promoveu o restabelecimento das relações evolutivas entre essas espécies, porém, não houve sucesso na definição da relação evolutiva de *S. frugiperda* com os clados próximos.

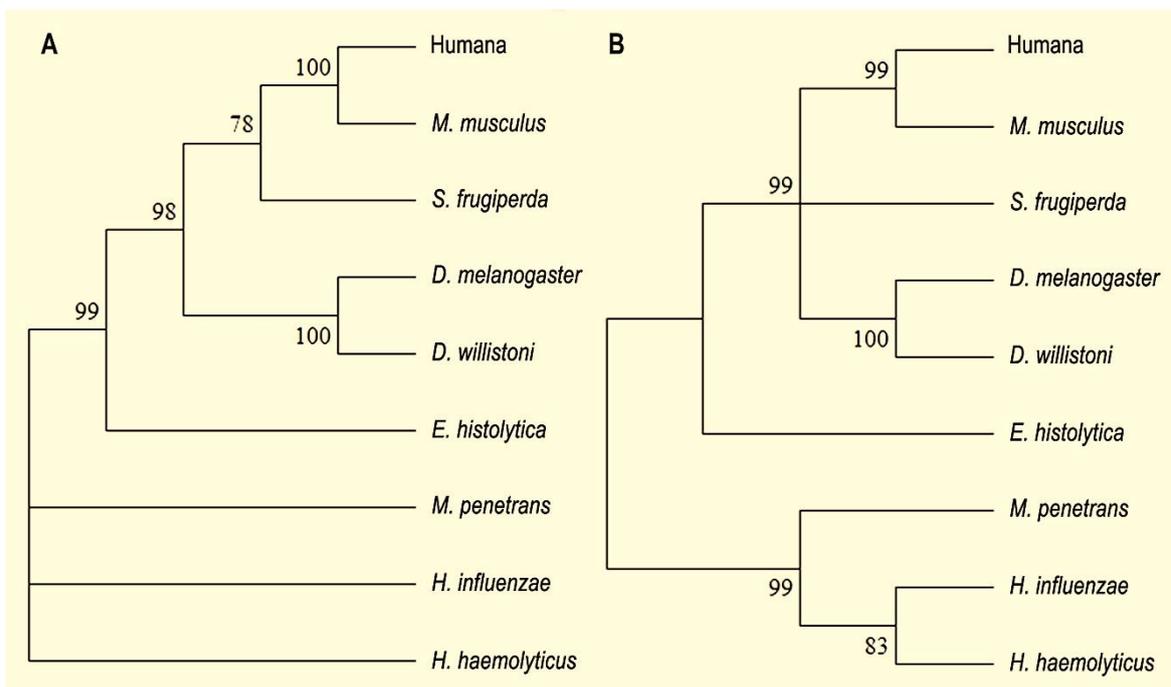


Figura 19. Análise filogenética pelo método de máxima verossimilhança das sequências nucleotídicas. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas. Os ramos que apresentaram *bootstrap* menor do que 75 foram condensados. O modelo evolutivo utilizado foi o GTR (*general time reversible*) com taxa de sítios invariáveis (I) e distribuição gama (G). **(A)** árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; **(B)** árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2.

A árvore proveniente da análise bayesiana do alinhamento das sequências primárias não resgatou uma relação evolutiva clara, agrupando as sequências de *S. frugiperda*, *E. histolytica*, *M. musculus* e humana (**Figura 20A**). Diferentemente, a filogenia obtida tendo como *input* o alinhamento 3D-1D das estruturas apresentou agrupamentos similares aos da análise por máxima-verossimilhança e com valores de probabilidade posterior bayesiana de ramos mais robustos (**Figura 20B**). Na **Figura 21A** tem-se a árvore com inferência evolutiva bayesiana com o alinhamento através do programa MUSCLE, onde não há a definição entre os ramos de *H. haemolyticus*, *H. influenzae*, *M. penetrans* e *E. histolytica*. Porém, quando levada em conta as estruturas terciárias no alinhamento das sequências nucleotídicas, as reconstruções das relações evolutivas se estabelecem com valores de suporte mais elevados. Interessantemente, as relações evolutivas, ao nível nucleotídico, acabam agrupando no mesmo clado *S. frugiperda*, *D. melanogaster* e *D. willistoni* (**Figura**

21B). Tal fato pode estar atrelado às taxas de substituições nucleotídicas mais próximas entre os invertebrados.

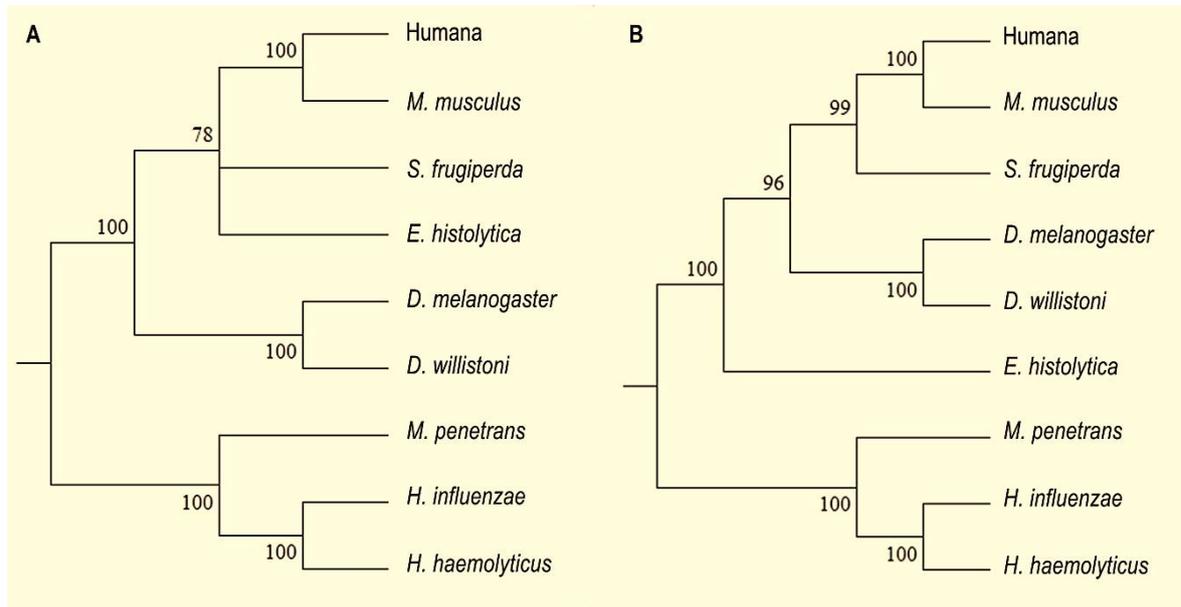


Figura 20. Inferência filogenética bayesiana das DNMT2 das sequências de aminoácidos. A árvore foi gerada utilizando o modelo evolutivo JTT com distribuição gama. Os ramos que apresentaram valores de probabilidade posterior bayesiana menor do que 75 foram condensados. (A) árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; (B) árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2.

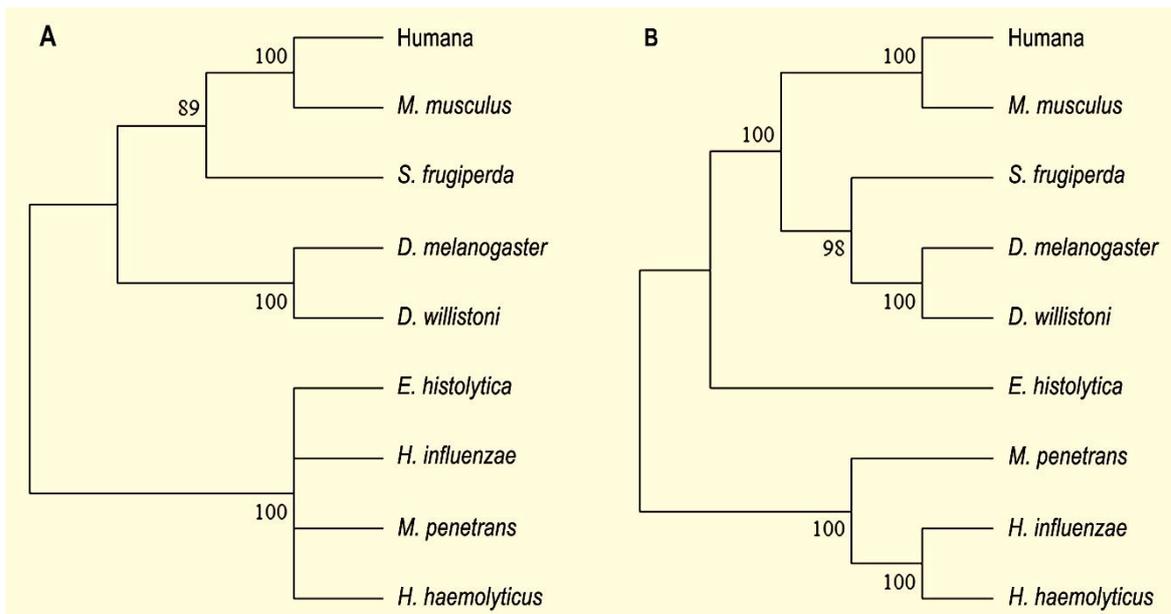


Figura 21. Inferência filogenética bayesiana das sequências nucleotídicas. Os ramos que apresentaram valores de probabilidade posterior bayesiana menor do que 75 foram condensados. O modelo evolutivo utilizado foi o GTR (*general time reversible*) com taxa de sítios invariáveis (I) e distribuição gama (G). **(A)** árvore consenso utilizando o algoritmo MUSCLE para o alinhamento das sequências; **(B)** árvore consenso utilizando informações de *fitting* das estruturas 3D das DNMT2.

As diferenças observadas nas análises filogenéticas remetem à questão de que as estruturas terciárias (domínios proteicos) se comportam como unidades evolutivas. Nesse caso, mudanças nas sequências (de nucleotídeos e de aminoácidos, conseqüentemente) podem ser toleradas enquanto não resultem em perturbações físico-químicas significantes nas propriedades que irão definir as estruturas secundárias, e essas por sua vez, na organização das estruturas terciárias de uma molécula. Assim, temos alta diversidade nas sequências nucleotídicas e de resíduos entre as DNMT2, mas quando se leva em conta as relações estruturais das moléculas, além das relações similaridade/identidade, há acréscimo no conjunto de informações que colaboram nas inferências filogenéticas, visto que as estruturas terciárias tendem a ser mais conservadas do que as sequências que as compõem.

2.2.4 Dinâmica molecular: Análises do comportamento das DNMT2 em meio fisiológico

Através dos princípios de mecânica clássica, a dinâmica molecular é capaz de fornecer dados a respeito do comportamento das moléculas em meio aquoso ou no vácuo, como por exemplo, a estabilidade da conformação proteica (Pikkemaat *et al.*, 2002) e alterações das estruturas secundárias ao longo do tempo (Campbell *et al.*, 2004), assim como as relações de interação intramoleculares e da própria molécula com o meio. No presente trabalho, o solvente é apresentado explicitamente e as interações eletrostáticas calculadas pelo método de Ewald (PME), onde os átomos tem a possibilidade de interagirem com as demais partículas do sistema.

A **figura 22** apresenta as análises do desvio médio quadrático (RMSD) das estruturas de DNMT2 analisadas, por um período de 50 ns de simulação. Como explicado anteriormente, o RMSD é uma medida em que as coordenadas dos C α podem ser aferidas a partir de um ponto (ou estrutura) de referência no espaço. No caso, o RMSD de uma dinâmica molecular mensura o quanto a molécula se desloca ao longo do tempo em relação às suas coordenadas iniciais. A maioria das estruturas não alcançam estabilização ao final dos 50 ns de simulação de dinâmica molecular. O RMSD de *H. haemolyticus*, *H. influenzae*, *M. penetrans*, *E. histolytica* e *S. frugiperda* apresentaram valores médios semelhantes, entre 0,30 nm e 0,37 nm. Já as moléculas de *D. willistoni*, *D. melanogaster*, *M. musculus* e humana tiveram valores médios entre 0,46 nm (*D. willistoni*) até 0,54 nm (*D. melanogaster*) (**tabela 4**). O maior desvio registrado foi de 0,70 nm em *M. musculus* aos 35 ns de simulação e o menor foi de 0,20 nm, em *H. influenzae* no tempo de 5 ns. A maior amplitude de RMSD registrada foi ao longo da simulação em *M. musculus* (0,32 nm) e a menor em *S. frugiperda* (0,16 nm).

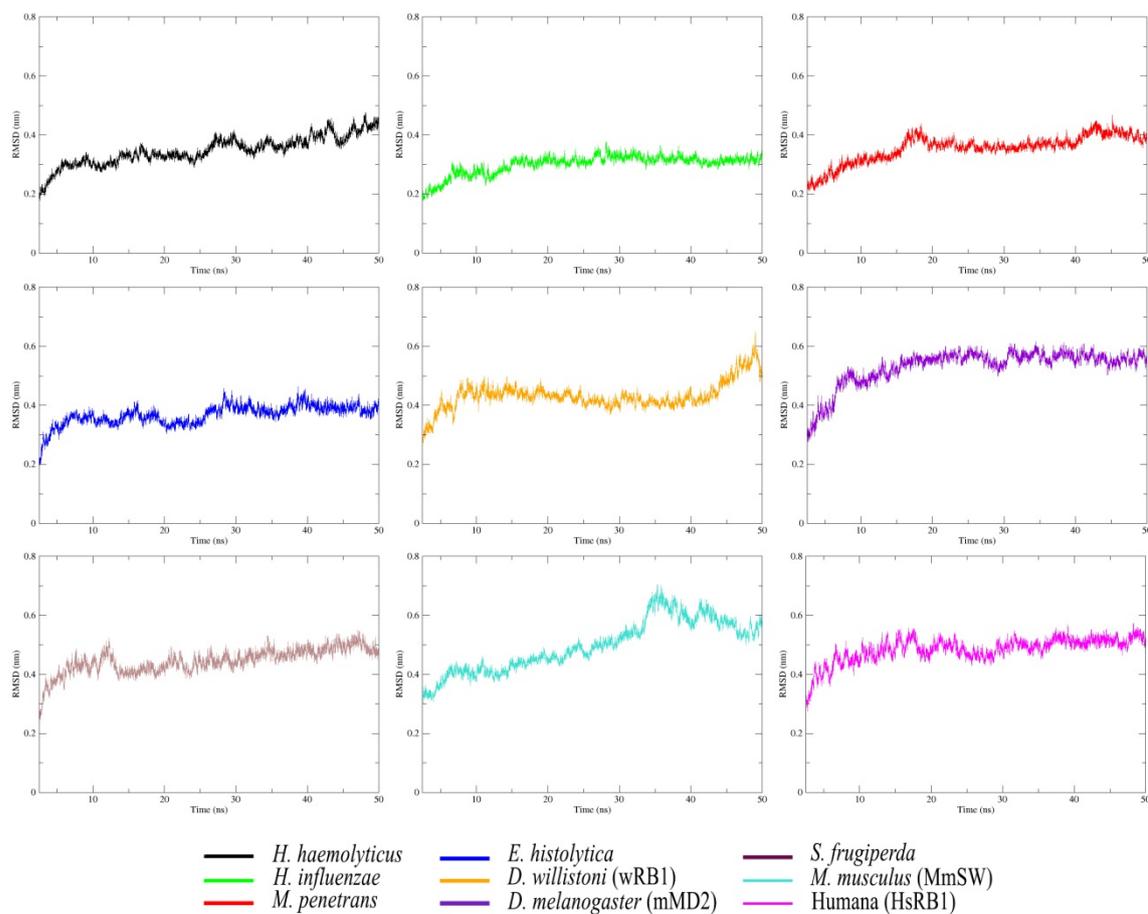


Figura 22. Desvio médio quadrático (RMSD) com todos os átomos das estruturas 3D de DNMT2 durante a simulação de dinâmica molecular, no intervalo de 2,5ns à 50 ns – que corresponde ao tempo de término da fase de minimização de energia e termalização das estruturas e início da fase de produção.

Cabe observar que os valores de RMSD das estruturas, nesse primeiro instante, levam em consideração as estruturas completas das DNMT2 com relação às suas posições no início da simulação, incluindo-se as regiões de alças (que tradicionalmente possuem uma maior flexibilidade). Em uma molécula, onde grande parte de sua composição estrutural é de alças (alças do domínio catalítico ProCys, de ligação entre os domínios catalítico e de reconhecimento de substrato e a própria TRD) espera-se uma maior mobilidade da molécula ao longo da simulação, impactando nos valores de RMSD. Além disso, os domínios que compõem as DNMT2 possuem processos de enovelamento diferenciados entre si. Enquanto o domínio catalítico é relativamente bem estruturado, com um arcabouço formado por um clássico empacotamento de Rossmann (alternância

de motivos folha- β / α -hélice/folha- β), a TRD é composta basicamente de alças, sem apresentarem uma arquitetura de maior complexidade (**Figura 23**).

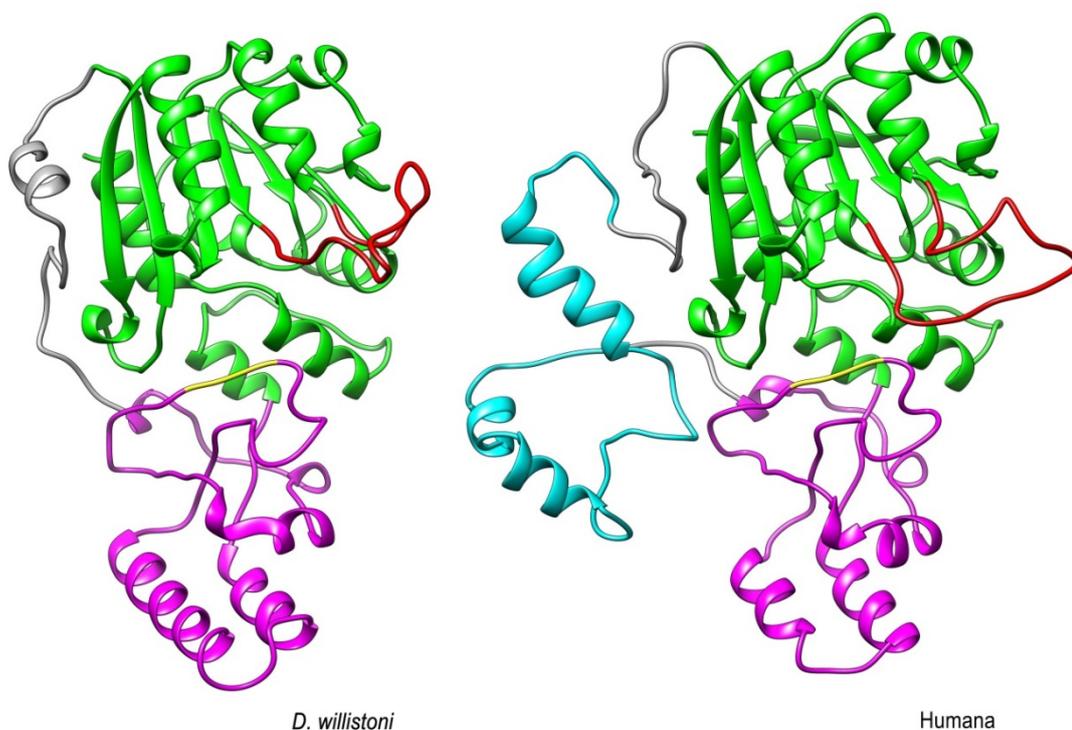


Figura 23. Representação *ribbon* das estruturas DNMT2 de *D. willistoni* e humana. Os diferentes domínios estão representados em diferentes cores: domínio catalítico (verde), alça contendo cisteína catalítica (vermelho), domínio II contendo a TRD (magenta), alça de ligação dos domínios (cinza), segmento não resolvido na estrutura cristalográfica 1G55, DNMT2 humana (ciano), motivo conservado CFT (amarelo).

Tabela 4. Valores de RMSD das estruturas entre os períodos de simulação de dinâmica molecular de 5 ns à 50 ns.

Espécie	RMSD (nm)		
	Média	Mínimo	Máximo
<i>H. haemolyticus</i>	0,35	0,25	0,48
<i>H. influenzae</i>	0,31	0,20	0,38
<i>M. penetrans</i>	0,36	0,23	0,47
<i>E. histolytica</i>	0,37	0,28	0,46
<i>D. willistoni</i>	0,46	0,35	0,57
<i>D. melanogaster</i>	0,54	0,34	0,62
<i>S. frugiperda</i>	0,36	0,28	0,44
<i>M. musculus</i>	0,51	0,38	0,70
Humana	0,49	0,36	0,57

Observando separadamente o comportamento dos dois domínios que compõem as DNMT2, verificam-se mais claramente as distinções entre esses, ao longo das simulações (**figura 24**). Além disso, a grande alça presente nas DNMT2 de vertebrados impactou substancialmente nos valores de RMSD das proteínas. Os domínios catalíticos das estruturas apresentam-se mais estáveis, variando entre 0,2 nm e 0,35 nm, enquanto os domínios TRD possuem valores de RMSD maiores, como era esperado pela arquitetura apresentada nessa região (principalmente alças). Porém, a diferença de RMSD entre o domínio catalítico e TRD de *H. haemolyticus*, *H. influenzae*, *M. penetrans*, *E. histolytica* e *S. frugiperda* é pequena, ao passo que as DNMT2 de *D. willistoni*, *D. melanogaster*, *M. musculus* e humana apresentam grandes diferenças de RMSD entre os dois domínios, durante a simulação de dinâmica molecular. Para Gerstein *et al.* (1994), a flexibilidade em determinadas regiões de uma proteína acaba por ser primordial para que essa desempenhe sua função biológica. Dessa forma, a rede de interações dinâmicas interna das moléculas analisadas acaba por influenciar o comportamento entre os diferentes domínios que as compõem. As diferenças observadas entre as TRDs podem estar associadas a uma maior ou menor flexibilidade no reconhecimento de substrato e sequência-alvo pelas enzimas. Assim, poderia se sugerir que, aquelas moléculas que possuem uma maior mobilidade na TRD comportariam a possibilidade de funções duais, metilando tRNA e/ou DNA, como já foi constatada em DNMT2 humana (Jeltsch *et al.*, 2006; Tuorto *et al.*, 2012).

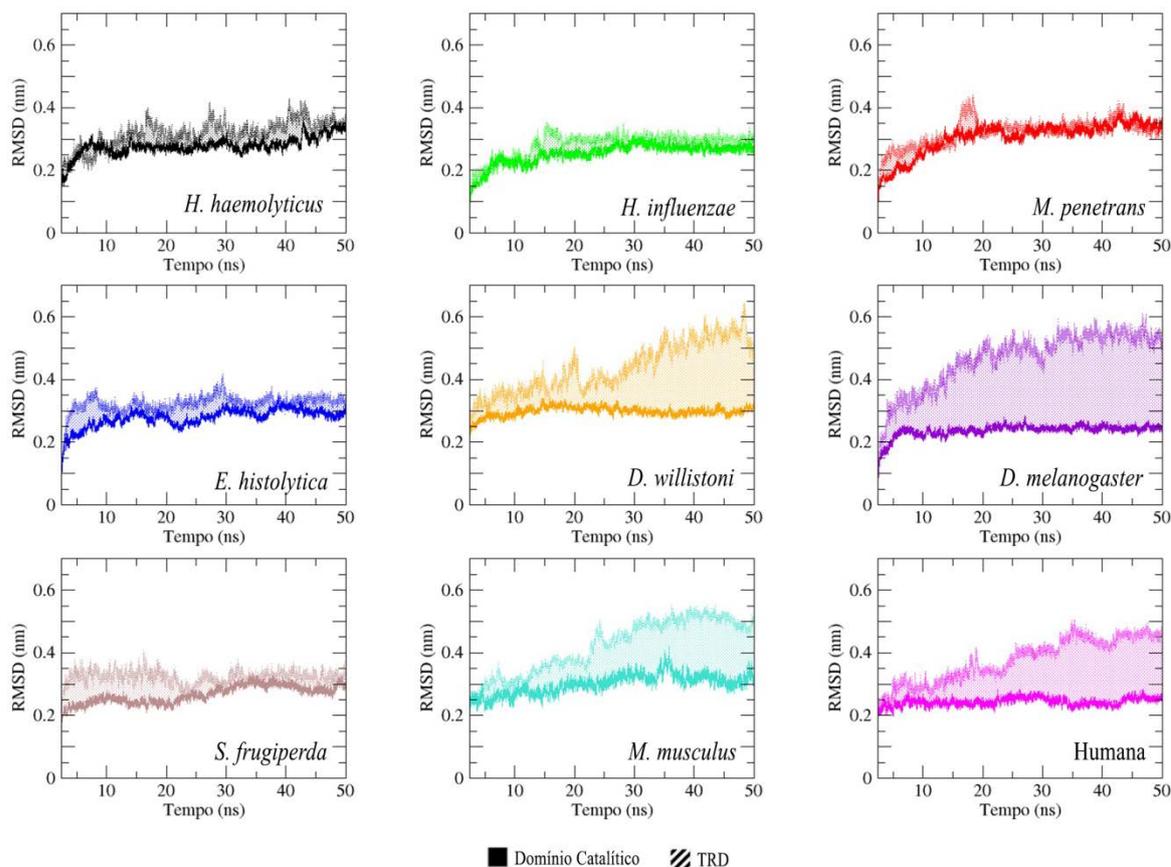


Figura 24. Comparação entre o RMSD dos domínios catalíticos e TRDs das DNMT2, no tempo de 2,5 ns à 50 ns. Linha contínua representando o RMSD do domínio catalítico e linha pontilhada, TRD. Região de preenchimento entre os gráficos destacando as diferenças entre os dois domínios. Na análise não foi levado em conta a alça de ligação dos domínios catalíticos e TRDs.

Dadas as marcantes diferenças de RMSD entre as enzimas, poderia se suspeitar da ocorrência de desestabilização de alguma estrutura secundária, principalmente por serem os modelos elaborados (wRB1, mMD2, MmSW e HsRB1) aqueles que apresentaram as maiores diferenças. Uma forma de avaliar as regiões que apresentaram maior ou menor estabilidade estrutural é a análise da flutuação da raiz quadrática média (RMSF), que fornece o deslocamento dos resíduos durante a simulação (**Figura 25**). Pode se observar a grande mobilidade dos resíduos envolvidos na alça que contém o dipeptídeo catalítico ProCys (entre os resíduos 75 à 100). Nessa região podem aparecer algumas estruturas secundárias momentâneas, mas não permanecem constantes ao longo da simulação. O resíduo Glu125 (dos quatro modelos) também apresentou destacada flutuação durante a dinâmica das estruturas. Esse resíduo se encontra

próximo ao motivo ENV, que é de grande importância no processo de estabilização da cisteína-alvo próximo ao arcabouço em que se encontra a estrutura doadora do grupamento metil, AdoMet. Essa mobilidade pode ter importância na acomodação das distâncias do motivo ENV durante as fases de *flipping* da cisteína-alvo da dupla-hélice do DNA e aproximação do nucleotídeo ao centro catalítico.

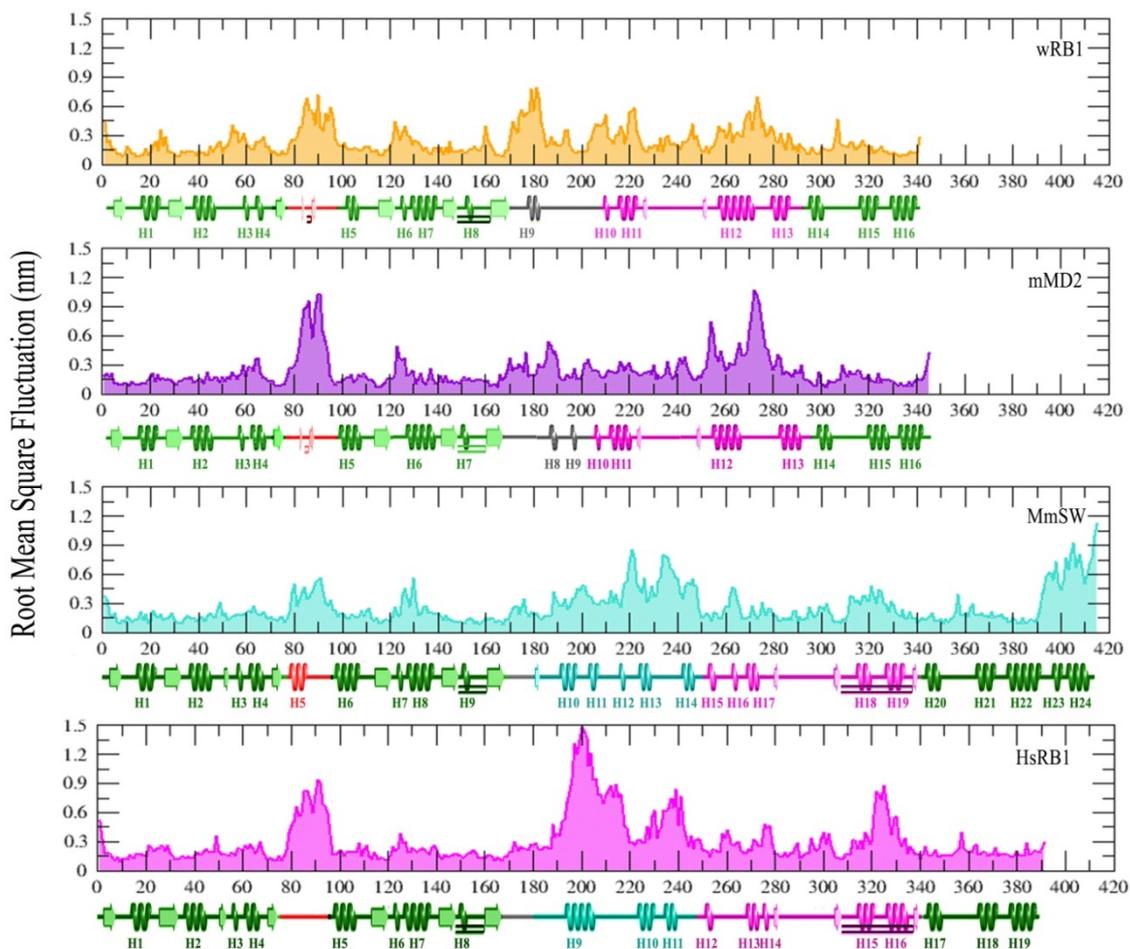


Figura 25. RMSF dos modelos wRB1 (*D. willistoni*), mMD2 (*D. melanogaster*), MmSW (*M. musculus*) e HsRB1 (humana) durante a simulação de dinâmica molecular. As estruturas secundárias das estruturas iniciais de cada modelo estão representadas abaixo de cada gráfico correspondente e coloridas de acordo com as representações da figura 24.

A partir do resíduo 170, inicia-se a alça de ligação dos dois domínios da DNMT2, que se caracteriza pela significativa flutuação dos resíduos que a compõem. Cabe ressaltar a destacada flutuação dos resíduos 200 até 250,

aproximadamente, nas estruturas de MmSW e HsRB1, que corresponde à grande inserção característica das DNMT2 de vertebrados (Dong *et al.* 2001). Interessantemente, a alça componente da TRD, onde o motivo conservado CFTxxYxxY está presente, se comporta de maneira mais estável (resíduos 236-244 em wRB1; 233-241 em mMD2; 292-300 em MmSW; 292-300 em HsRB1). Essa estabilidade também foi descrita por Evans & Bronowska (2010), quando na presença do cofator SAM, há um aumento na dinâmica de diversos resíduos do arcabouço de ligação do cofator, em contrapartida há a diminuição da dinâmica observada no motivo CFTxxYxxY.

Em seguida, existe uma região nos modelos wRB1, mMD2 e HsRB1, de arquitetura do tipo α -hélice/alça/ α -hélice (resíduos 250 à 280 em drosofilídeos e 308 à 336 na enzima humana), que apresentou alta flutuação. Entretanto, em drosofilídeos as análises das estruturas secundárias ao longo da dinâmica não apresentaram desestruturação das α -hélices (H12 e H13) constituintes desse motivo. Em *D. melanogaster*, nessa região existe uma inserção de 10 aminoácidos na alça entre as hélices, que não está presente nas demais DNMT2 analisadas.

A análise de raio de giro de uma proteína ao longo da simulação de dinâmica molecular fornece informações a respeito do raio médio do sistema molecular. Nesse tipo de medida, a molécula é tratada como uma esfera em que são tomadas medidas de expansão ou compactação da mesma. Assim, essa análise, aliada às análises anteriores pode indicar o tipo de mudança conformacional da proteína. Conforme se pode observar na **Figura 26**, o domínio catalítico e a TRD apresentam padrões de expansão e compactação semelhantes entre si ao longo das simulações e com uma leve tendência à compactação, o que parece indicar uma interatividade termodinâmica entre os dois domínios. Na DNMT2 de *M. musculus* existe uma maior variação de raio de giro no domínio catalítico, contudo, essa variação é decorrente à mobilidade existente na região C-terminal, que compreende as α -hélices H23 e H24.

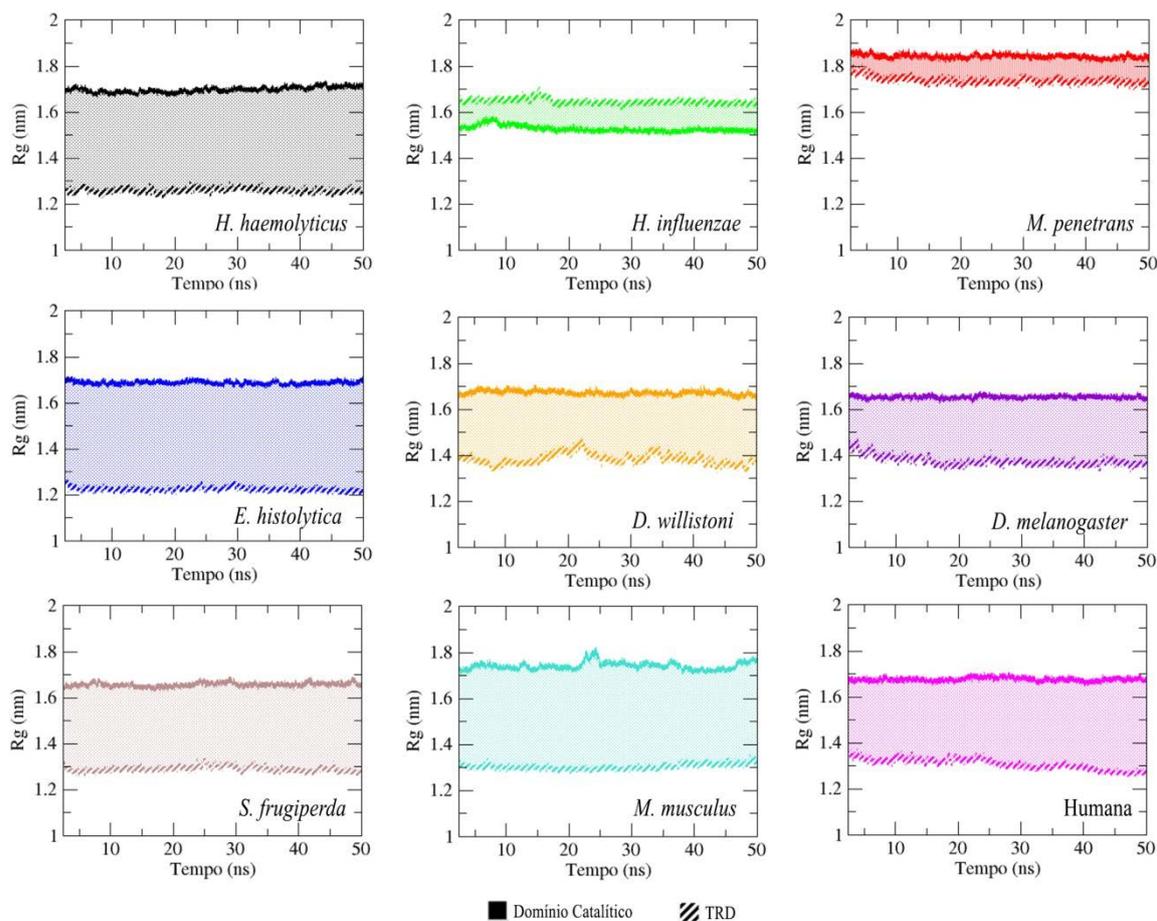


Figura 26. Análise de raio de giro das regiões catalíticas e TRDs das DNMT2 durante a simulação de dinâmica molecular, no intervalo de 2,5ns à 50 ns.

É válido ressaltar que os movimentos amostrados até então não são conclusivos, pois se está avaliando a ação das moléculas no decorrer de tempo de 50 ns, apenas. Mesmo que em alguns casos exista uma clara tendência de estabilidade conformacional, não existe uma amostragem temporal suficiente para abarcar, por exemplo, a análise de todos os possíveis movimentos de baixa frequência e/ou de grandes amplitudes, favorecendo a análise de flutuações locais. Além disso, as análises realizadas no presente trabalho estão focadas na estrutura apoproteína de DNMT2, sem a presença do cofator AdoMet, AdoHcy e substratos de ação (DNA e tRNA).

CAPÍTULO III

Modelagem Molecular por Homologia & Análises Evolutivas das DNMT2 de drosofilídeos

Metodologia

Resultados e Discussão

3.1 Metodologia

3.1.1 Busca de sequências homólogas no banco de dados Flybase e modelagem molecular por homologia das DNMT2 de drosofilídeos

Para averiguar se a relação de similaridade dos potenciais eletrostáticos entre *D. melanogaster* e *D. willistoni* não passaria de uma similaridade relativa devido ao distanciamento de relações evolutivas entre as nove (9) estruturas analisadas, passou a ser interessante avaliar essa relação dentro de um quadro de estruturas mais próximas evolutivamente. Assim, buscou-se no banco de dados Flybase (<http://flybase.org/>) DNMT2 homólogas de drosofilídeos. Ao todo foram recobradas 20 sequências de DNMT2 do gênero *Drosophila* (Tabela 5). Além disso, como o único representante do grupo *willistoni* neste trabalho é a *D. willistoni*, utilizou-se a sequência dos resíduos de DNMT2 de *D. tropicalis*, oriundos do trabalho de caracterização e conservação de DNMT2 de drosofilídeos de nosso grupo de pesquisa.

A modelagem por homologia (Modeller 9.11) das 20 sequências de drosofilídeos realizou-se conforme descrito no capítulo II (item 1.1.2), utilizando como moldes os modelos de *D. melanogaster* (mMD2) e de *D. willistoni* (wRB1) validados no presente trabalho.

Tabela 5. Sequências homólogas de Dnmt2 obtidas através da ferramenta BLASTn. A sequência de Dnmt2 de *D. melanogaster* foi utilizada como *query*.

Espécie	Anotação contendo o identificador do Flybase
<i>Drosophila simulans</i>	gn dsim 2L type=golden_path_region; loc=2L:1..22036055; ID=2L; dbxref=GB:CM000361
<i>Drosophila sechellia</i>	gn dsec scaffold_16 type=golden_path_region; loc=scaffold_16:1..1878335; ID=scaffold_16; dbxref=GB:CH480831
<i>Drosophila yakuba</i>	gn dyak 2L type=golden_path_region; loc=2L:1..22324452; ID=2L; dbxref=GB:CM000157
<i>Drosophila erecta</i>	gn dere scaffold_4929 type=golden_path_region; loc=scaffold_4929:1..26641161; ID=scaffold_4929; dbxref=GB:CH954177
<i>Drosophila ficusphila</i>	gi 459201587 gb KB457400.1 <i>Drosophila ficusphila</i> unplaced genomic scaffold scf7180000453936, whole genome shotgun sequence

<i>Drosophila eugracilis</i>	gi 459206169 gb KB464972.1 <i>Drosophila eugracilis</i> unplaced genomic scaffold scf7180000409122, whole genome shotgun sequence
<i>Drosophila biarmipes</i>	gi 459197679 gb KB462833.1 <i>Drosophila biarmipes</i> unplaced genomic scaffold scf7180000302422, whole genome shotgun sequence
<i>Drosophila takahashii</i>	gi 459204042 gb KB461661.1 <i>Drosophila takahashii</i> unplaced genomic scaffold scf7180000415863, whole genome shotgun sequence
<i>Drosophila elegans</i>	gi 459200570 gb KB458413.1 <i>Drosophila elegans</i> unplaced genomic scaffold scf7180000491028, whole genome shotgun sequence
<i>Drosophila rhopaloa</i>	gi 452191607 gb KB448653.1 <i>Drosophila rhopaloa</i> unplaced genomic scaffold scf7180000761302, whole genome shotgun sequence
<i>Drosophila kikkawai</i>	gi 459202993 gb KB459629.1 <i>Drosophila kikkawai</i> unplaced genomic scaffold scf7180000302408, whole genome shotgun sequence
<i>Drosophila ananassae</i>	gnl dana scaffold_12943 type=golden_path_region; loc=scaffold_12943:1..5039921; ID=scaffold_12943; dbxref=GB:CH902624
<i>Drosophila bipectinata</i>	gi 459198809 gb KB464388.1 <i>Drosophila bipectinata</i> unplaced genomic scaffold scf7180000396728, whole genome shotgun sequence
<i>Drosophila pseudoobscura</i>	gnl dpse 4_group2 type=golden_path_region; loc=4_group2:1..1235759; ID=4_group2; dbxref=GB:CH379059
<i>Drosophila persimilis</i>	gnl dper scaffold_8 type=golden_path_region; loc=scaffold_8:1..3966273; ID=scaffold_8; dbxref=GB:CH479187
<i>Drosophila miranda</i>	gi 480995217 gb CM001520.2 <i>Drosophila miranda</i> strain MSH22 chromosome 4, whole genome shotgun sequence
<i>Drosophila mojavensis</i>	gnl dmoj scaffold_6500 type=golden_path_region; loc=scaffold_6500:1..32352404; ID=scaffold_6500; dbxref=GB:CH933807
<i>Drosophila virilis</i>	gnl dvir scaffold_12963 type=golden_path_region; loc=scaffold_12963:1..20206255; ID=scaffold_12963; dbxref=GB:CH940649
<i>Drosophila albomicans</i>	gi 405988621 gb JH859534.1 <i>Drosophila albomicans</i> strain KM55-5 unplaced genomic scaffold Dalb_scaffold_67169, whole genome shotgun sequence
<i>Drosophila grimshawi</i>	gnl dgri scaffold_15252 type=golden_path_region; loc=scaffold_15252:1..17193109; ID=scaffold_15252; dbxref=GB:CH916368

3.2 Resultados e Discussão

3.2.1 As DNMT2 de drosofilídeos são conservadas em sequência e estrutura, mas divergem em suas propriedades

Os 23 modelos de *Drosophila* foram carregados no servidor do programa PIPSA para análise comparativa dos potenciais de superfície (**Figura**

27). Na análise de similaridades dos potenciais eletrostáticos de superfície com as espécies do gênero *Drosophila*, ficam evidentes as diferenças existentes entre os potenciais entre as DNMT2 de espécies do grupo *willistoni* em relação aos demais drosofilídeos. Grande parte das espécies do grupo *melanogaster* parece partilhar valores de potenciais muito próximos, porém *D. bipectinata*, *D. ananassae* e *D. yakuba* agrupam-se distintamente às demais espécies filogeneticamente próximas. *D. eugracilis* (grupo *melanogaster*) acaba relacionando-se em termos de potenciais de superfície com as espécies do grupo *obscura*.

É importante lembrar que a análise oriunda do PIPSA não tem como objetivo a averiguação das relações filogenéticas entre as espécies, mas as semelhanças existentes na distribuição do campo eletrostático na superfície das moléculas estudadas. Como apresentado anteriormente, existe a possibilidade de que espécies muito próximas em termos evolutivos, como *H. haemolyticus* e *H. influenzae*, mesmo compartilhando percentuais moderados de conservação de resíduos, venham a ter mecanismos e preferências por sítios de metilação diferentes. Garcia *et al.* (2007) demonstrou em seu trabalho a elevada conservação de resíduos nos domínios catalíticos de DNMT2 de *D. willistoni* e *D. melanogaster*, mas grande variabilidade na TRD dessas espécies, o que pode indicar diferenças na afinidade e funções das DNMT2.

A alta taxa de polimorfismos intraespecíficos em espécies do grupo *willistoni* é uma rica fonte para a emergência de novidades evolutivas em diversos níveis, nesses organismos. O fenômeno de metilação sexo-específico é uma das peculiaridades pertencentes a esse grupo, o que enseja modificações nas ferramentas envolvidas no processo de metilação de citosinas. Como apresentado na análise de potencial de superfície, as diferenças encontradas podem ser importantes para determinar uma modulação ou afinidade diferencial das DNMT2 de espécies do grupo *willistoni*, comparativamente às demais espécies de drosofilídeos.

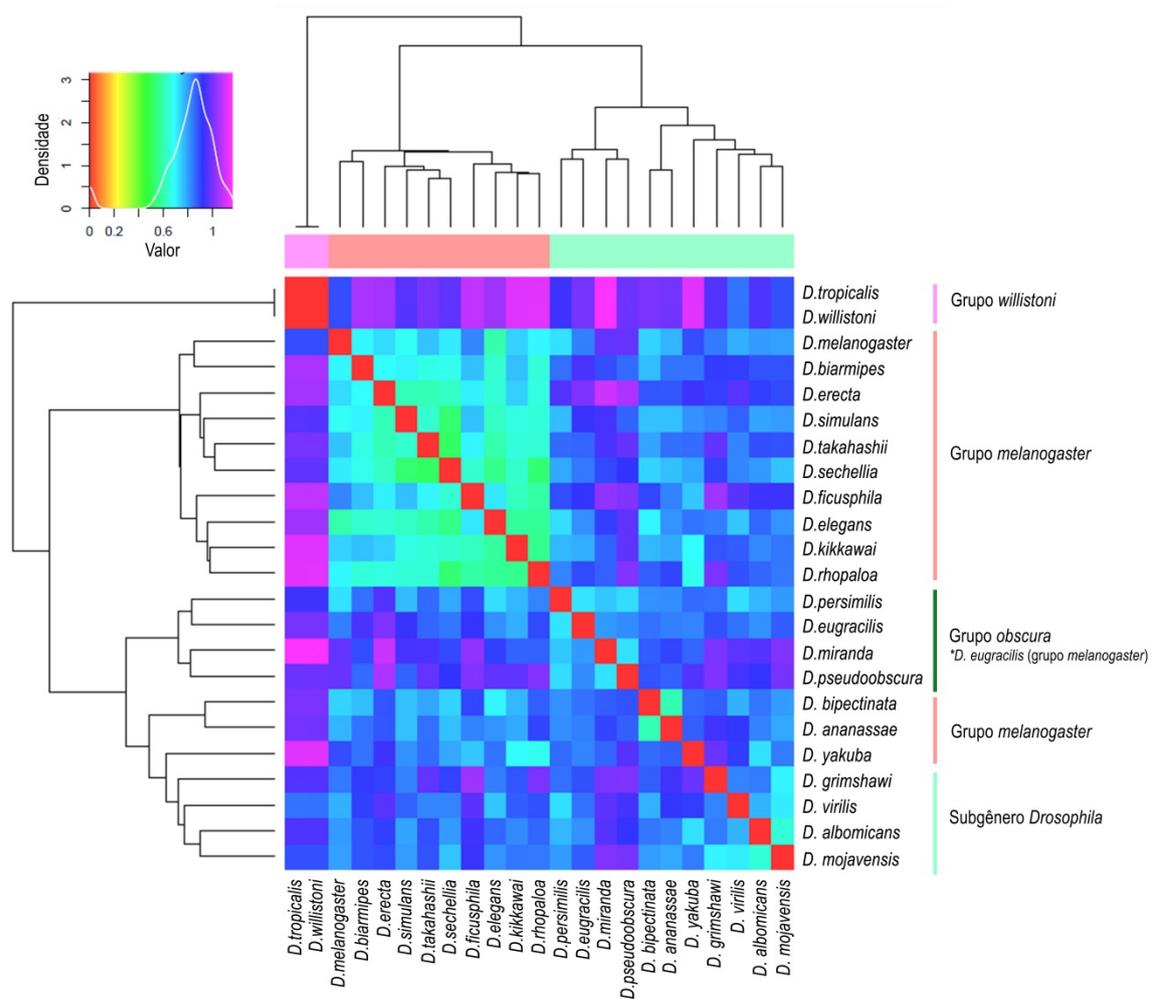


Figura 27. Epograma da análise por PIPSA (*Protein Interaction Property Similarity Analysis*) das 23 DNMT2 de drosofilídeos. As estruturas estão agrupadas segundo a relação de similaridade dos potenciais eletrostáticos a partir da matriz de distância elaborada pelo programa. Acima à esquerda encontra-se a chave de cores, indicando a relação de similaridade, quanto menor o valor, maior similaridade entre as estruturas.

Na **figura 28** podem ser visualizados os padrões de distribuição de potencial eletrostático e superfície de diferentes espécies de drosofilídeos. Nota-se que dentro de grupos mais relacionados evolutivamente ainda assim há diferenças em como se distribuem as cargas na superfície de suas moléculas.

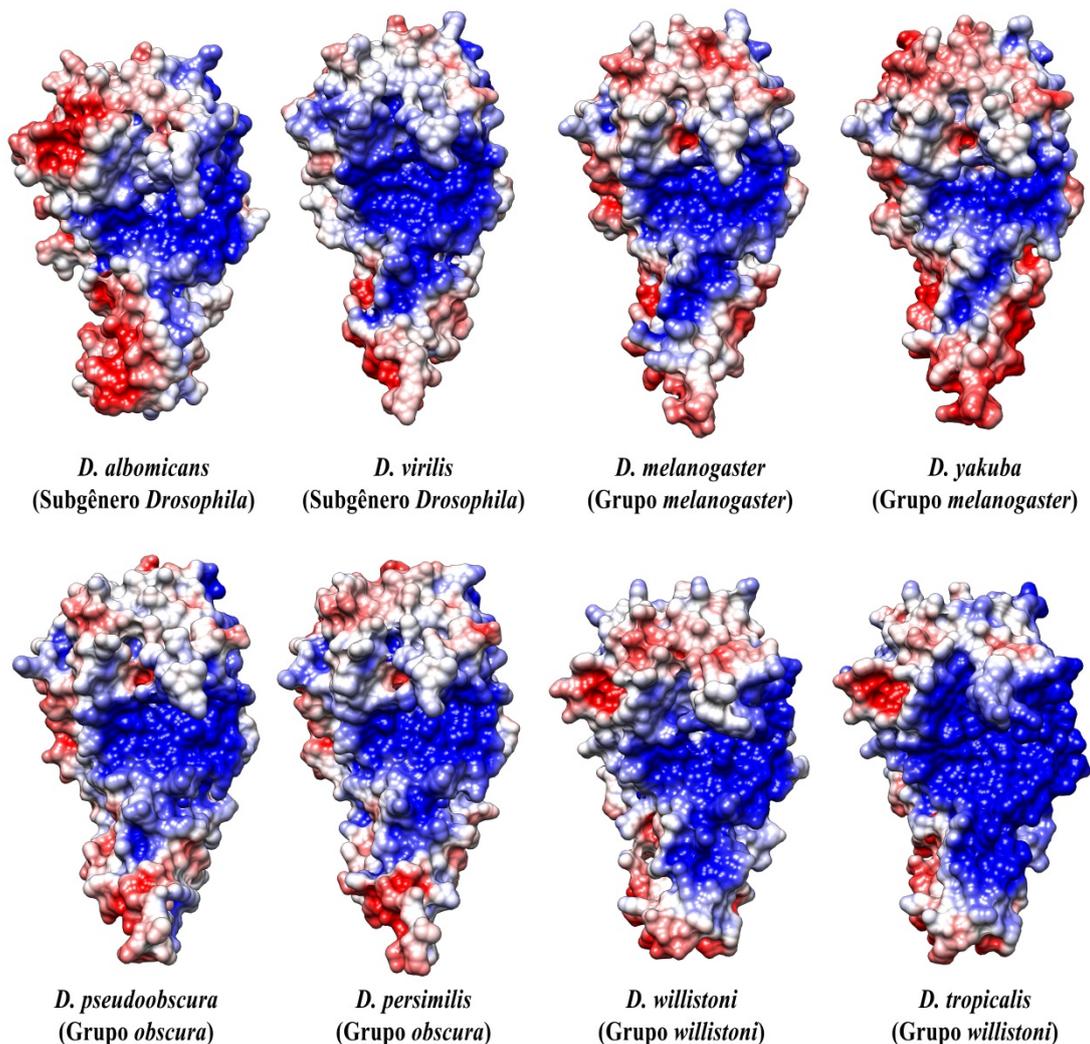


Figura 28. Potencial eletrostático de superfície das estruturas de DNMT2 de representantes de diferentes grupos de drosofilídeos. As superfícies estão coloridas de acordo com o potencial eletrostático, em azul para potencial positivo ($5 kT$), vermelho para negativo ($-5 kT$) e branco para neutro, onde k representa a constante de Boltzmann e T , a temperatura.

Em relação às sequências de DNMT2, drosofilídeos apresentam grande conservação de seus motivos catalíticos, mas há grau de variabilidade na TRD. A análise de divergência evolutiva de aminoácidos demonstra maior divergência entre o grupo *willistoni* e grupo *melanogaster* (27,2%) e a menor divergência entre o grupo *obscura* e grupo *melanogaster* (17,9%) (**Tabela 6**). A menor divergência de resíduos entre espécies é de *D. persimilis* e de *D. pseudoobscura* (0,6%), ambas pertencentes ao grupo *obscura*. Já a maior divergência entre espécies se encontra entre *D. biarmipes* e *D. tropicalis* (30,1%)

(**Tabela 7**). Entre os nucleotídeos, a análise entre grupos apresenta o grupo *obscura* e grupo *willistoni* com a maior divergência evolutiva (31,1%), enquanto o grupo *melanogaster* e grupo *obscura* tiveram a menor diferença observada (24,1%). A menor divergência observada entre espécies foi de *D. persimilis* com *D. pseudoobscura* (0,3%), e a maior estimativa de divergência foi entre *D. tropicalis* e *D. kikkawai* (33,7%).

Tabela 6. Estimativa da divergência evolutiva dos aminoácidos entre grupos (diagonal inferior) e de nucleotídeos (diagonal superior). Todas as posições que continham *gap* foram eliminadas da análise.

	1	2	3	4
1 Subgênero <i>Drosophila</i>		0,282	0,268	0,305
2 Grupo <i>melanogaster</i>	0,209		0,240	0,310
3 Grupo <i>obscura</i>	0,213	0,179		0,311
4 Grupo <i>willistoni</i>	0,267	0,272	0,250	

Na **tabela 8** apresenta o desvio de códon das 23 espécies de drosofilídeos em relação à *D. melanogaster*. De acordo com a tabela, há um maior desvio de uso de códon para os resíduos de fenilalanina (PHE), isoleucina (ILE), histidina (HIS), asparagina (ASN) e glicina (GLY). Espécies do subgênero *Drosophila* e do grupo *willistoni* apresentam maior grau de afastamento da preferência de uso de códons em relação à *D. melanogaster*. Destaca-se, também, as diferenças encontradas no uso de códons por *D. bipectinata*, *D. eugracilis*, *D. ficusphila* e *D. rhopaloa* que pertencem ao grupo *melanogaster*.

As análises filogenéticas por máxima-verossimilhança (**Figura 29**) e bayesiana (**Figura 30**) das sequências de aminoácido apresentaram a mesma topologia, entretanto, os valores de suporte da árvore advinda da análise bayesiana apresentaram-se mais robustos. Quando feitas as inferências filogenéticas das sequências de nucleotídeos (**Figura 31** e **Figura 32**), são mantidas as mesmas relações evolutivas entre as espécies, com a divisão dos clados compostos pelas espécies do grupo *melanogaster*, grupo *obscura*, grupo *willistoni* e as espécies do subgênero *Drosophila* separadas das espécies do subgênero *Sophophora*. *D. willistoni* e *D. tropicalis* assumem posição mais

externa na filogenia, como o esperado para essas espécies basais do subgênero *Sophophora* (Robe *et al.* 2010).

Tabela 7. Estimativa de divergência evolutiva das seqüências de aminoácidos (diagonal inferior) e das seqüências de nucleotídeos (diagonal superior). Todas as posições que continham *gaps* foram eliminadas da análise. Maiores divergências destacadas em vermelho e menores, em azul.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
1 <i>D.albomicans</i>																								
2 <i>D.ananassae</i>	0,306																							
3 <i>D.biarmipes</i>	0,239	0,263																						
4 <i>D.bipectinata</i>	0,230	0,224	0,262																					
5 <i>D.elegans</i>	0,239	0,081	0,224	0,243																				
6 <i>D.erecta</i>	0,202	0,183	0,143	0,183	0,178																			
7 <i>D.eugracilis</i>	0,186	0,202	0,140	0,199	0,130	0,195																		
8 <i>D.ficusphila</i>	0,202	0,199	0,146	0,205	0,121	0,130	0,186																	
9 <i>D.grimshawi</i>	0,214	0,183	0,146	0,193	0,127	0,140	0,112	0,294																
10 <i>D.kikkawai</i>	0,180	0,252	0,224	0,255	0,199	0,202	0,205	0,230	0,282															
11 <i>D.melanogaster</i>	0,183	0,186	0,171	0,189	0,152	0,140	0,152	0,140	0,208	0,226														
12 <i>D.miranda</i>	0,193	0,205	0,161	0,205	0,130	0,068	0,140	0,146	0,208	0,143	0,242													
13 <i>D.mojavensis</i>	0,224	0,208	0,208	0,177	0,174	0,174	0,174	0,208	0,183	0,177	0,275	0,013												
14 <i>D.persimilis</i>	0,186	0,236	0,227	0,245	0,205	0,224	0,220	0,230	0,155	0,214	0,220	0,233	0,270	0,012										
15 <i>D.pseudoboscuro</i>	0,217	0,202	0,205	0,199	0,171	0,171	0,171	0,205	0,177	0,177	0,019	0,227	0,270	0,257	0,273	0,271	0,277	0,191	0,313	0,314	0,280			
16 <i>D.rhopala</i>	0,214	0,202	0,199	0,199	0,168	0,168	0,171	0,168	0,199	0,174	0,174	0,019	0,220	0,006	0,223	0,235	0,232	0,230	0,255	0,299	0,320	0,222		
17 <i>D.sechella</i>	0,199	0,189	0,149	0,196	0,068	0,124	0,118	0,118	0,196	0,137	0,124	0,174	0,196	0,171	0,165	0,177	0,178	0,143	0,265	0,294	0,316	0,163		
18 <i>D.simulans</i>	0,193	0,205	0,149	0,202	0,134	0,056	0,137	0,143	0,205	0,140	0,034	0,177	0,220	0,174	0,171	0,127	0,014	0,178	0,276	0,288	0,309	0,097		
19 <i>D.takahashii</i>	0,186	0,199	0,146	0,196	0,130	0,050	0,130	0,137	0,202	0,137	0,028	0,174	0,214	0,171	0,168	0,121	0,006	0,176	0,276	0,292	0,310	0,094		
20 <i>D.virilis</i>	0,189	0,186	0,118	0,193	0,106	0,109	0,118	0,093	0,202	0,124	0,099	0,165	0,211	0,161	0,158	0,093	0,096	0,093	0,278	0,304	0,320	0,160		
21 <i>D.willistoni</i>	0,146	0,233	0,205	0,236	0,193	0,202	0,199	0,208	0,124	0,193	0,193	0,208	0,130	0,202	0,196	0,177	0,189	0,183	0,177	0,291	0,299	0,280		
22 <i>D.tropicalis</i>	0,270	0,280	0,292	0,280	0,264	0,258	0,276	0,252	0,242	0,258	0,264	0,245	0,258	0,245	0,239	0,255	0,270	0,264	0,255	0,264	0,050	0,302		
23 <i>D.yakuba</i>	0,280	0,289	0,301	0,280	0,273	0,267	0,289	0,264	0,261	0,264	0,280	0,258	0,283	0,258	0,255	0,270	0,286	0,280	0,267	0,283	0,053	0,326		
	0,196	0,180	0,155	0,183	0,130	0,059	0,127	0,127	0,211	0,140	0,081	0,180	0,220	0,177	0,174	0,121	0,068	0,062	0,109	0,205	0,264	0,273		

Tabela 8. Desvio de uso de codons da sequência codificante de *Dnmt2* de *D. melanogaster* em comparação às 22 espécies de drosofilídeos analisadas. O asterisco indica preferência de codon idêntica entre as espécies. As menores relações encontradas no uso de códons estão destacadas em negrito.

	PHE	LEU	ILE	VAL	SER	PRO	THR	ALA	TYR	HIS	GLN	ASN	LYS	ASP	GLU	CYS	ARG	SER	GLY	Proporção
<i>D.melanogaster</i> (grupo <i>melanogaster</i>)	TTT	CTG	ATT	GTC	TCG	CCA	ACG	GCC	TAC	CAT	CAG	AAT	AAG	GAC	GAA	TGC	CGC	AGT	GGC	1.00
<i>D.ananassae</i> (grupo <i>melanogaster</i>)	*	CTT	* GTT	* CCG	*		*	*	*	*	*	*	AAA	GAT	*	*	AGA	AGC	GGT	0.58
<i>D.biarmipes</i> (grupo <i>melanogaster</i>)	*	*	ATC	GTG	* CCC	ACC	GCG	*	*	*	*	*	*	*	GAG	*	*	AGC	*	0.63
<i>D.bipectinata</i> (grupo <i>melanogaster</i>)	*	CTT	ATA	* CCT	*		GCA	*	CAC	CAA	*	AAA	GAT	*	*	CGG	AGC	GGT/A	0.42	
<i>D.elegans</i> (grupo <i>melanogaster</i>)	*	*	* GTT	TCC	CCC	ACC/AG	*	TAT	*	*	*	*	*	*	GAG	*	*	*	*	0.68
<i>D.erecta</i> (grupo <i>melanogaster</i>)	TTC	*	* GTG	TCC	CCG	*	*	TAT	*	*	*	*	*	*	GAG	TGT	*	*	*	0.63
<i>D.eugracilis</i> (grupo <i>melanogaster</i>)	*	TTG	ATC	GTT	TCC	CCG	ACT	*	TAT	*	GAG	*	AAA	GAT	*	TGT	*	*	GGA	0.37
<i>D.ficusphila</i> (grupo <i>melanogaster</i>)	*	TTG - CTA/G	* GTT	TCC	CCC	ACA	*	TAT	*	CAA	*	*	GAT	*	*	CGA	*	GGA	0.47	
<i>D.kikkawai</i> (grupo <i>melanogaster</i>)	TTC	*	ATA	GTG	* CCC	*	*	*	*	*	AAC	*	*	GAG	*	*	AGC	*	0.63	
<i>D.rhopaloea</i> (grupo <i>melanogaster</i>)	*	*	* GTT	* CCC	ACC	GCT/A	TAT	*	CAA	*	AAA	GAT	GAG	*	CGG	*	*	*	0.47	
<i>D.sechellia</i> (grupo <i>melanogaster</i>)	TTC	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	0.95
<i>D.simulans</i> (grupo <i>melanogaster</i>)	TTC	*	* CAC	*	*	*	*	*	*	*	*	*	*	GAG	*	*	*	*	*	0.84
<i>D.takahashii</i> (grupo <i>melanogaster</i>)	*	*	* GTT	TCC	CCC	*	*	*	*	CAA	*	*	*	*	*	*	AGC	*	*	0.74
<i>D.yakuba</i> (grupo <i>melanogaster</i>)	*	*	* GTG	TCC	*	*	*	TAT	*	*	*	*	*	GAG	TGT	*	*	*	*	0.74
<i>D.miranda</i> (grupo <i>obscura</i>)	*	TTG	* GTT	TCC	CCC	ACC/A	*	TAT	*	CAA	*	*	*	*	*	CGG	*	*	*	0.58
<i>D.persimilis</i> (grupo <i>obscura</i>)	*	TTG	* GTT	TCT	CCC	ACA	*	TAT	*	*	*	*	*	*	*	CGG	*	*	*	0.63
<i>D.pseudoobscura</i> (grupo <i>obscura</i>)	*	TTG	* GTT	TCT	CCC	ACA	*	TAT	*	*	*	*	*	*	*	CGG	*	*	*	0.63
<i>D.tropicalis</i> (grupo <i>willistoni</i>)	*	TTA	* GTT	TCA	* ACC	*	TAT	*	CAA	*	AAA	GAT	*	TGT	*	*	GGT/C/AG	0.47		
<i>D.willistoni</i> (grupo <i>willistoni</i>)	*	TTA	* GTT	TCA	* ACC	GCA	TAT	CAC	CAA	*	AAA	GAT	*	TGT	*	*	GGA	0.37		
<i>D.albomicans</i> (Subgênero <i>Drosophila</i>)	*	TTA/G - CTA/G	ATA	* TCC	* ACA	GCA	TAT	*	CAA	*	*	*	*	TGT	CGA	*	*	0.47		
<i>D.grimshawi</i> (Subgênero <i>Drosophila</i>)	*	TTG	* GTG	TCA	CCG	ACA	GCA	TAT	*	*	*	AAA	GAT	*	TGT	CGT	AGC	*	0.37	
<i>D.mojavensis</i> (Subgênero <i>Drosophila</i>)	*	CTT	* GTG	TCT	* ACA	GCT	TAT	CAC	CAA	*	AAA	GAT	*	CGA	*	*	0.42			
<i>D.virililis</i> (Subgênero <i>Drosophila</i>)	*	*	* GTG	TCT	CCC	ACA	GCG	TAT	CAC	*	*	AAA	GAT	*	*	AGC	*	0.47		

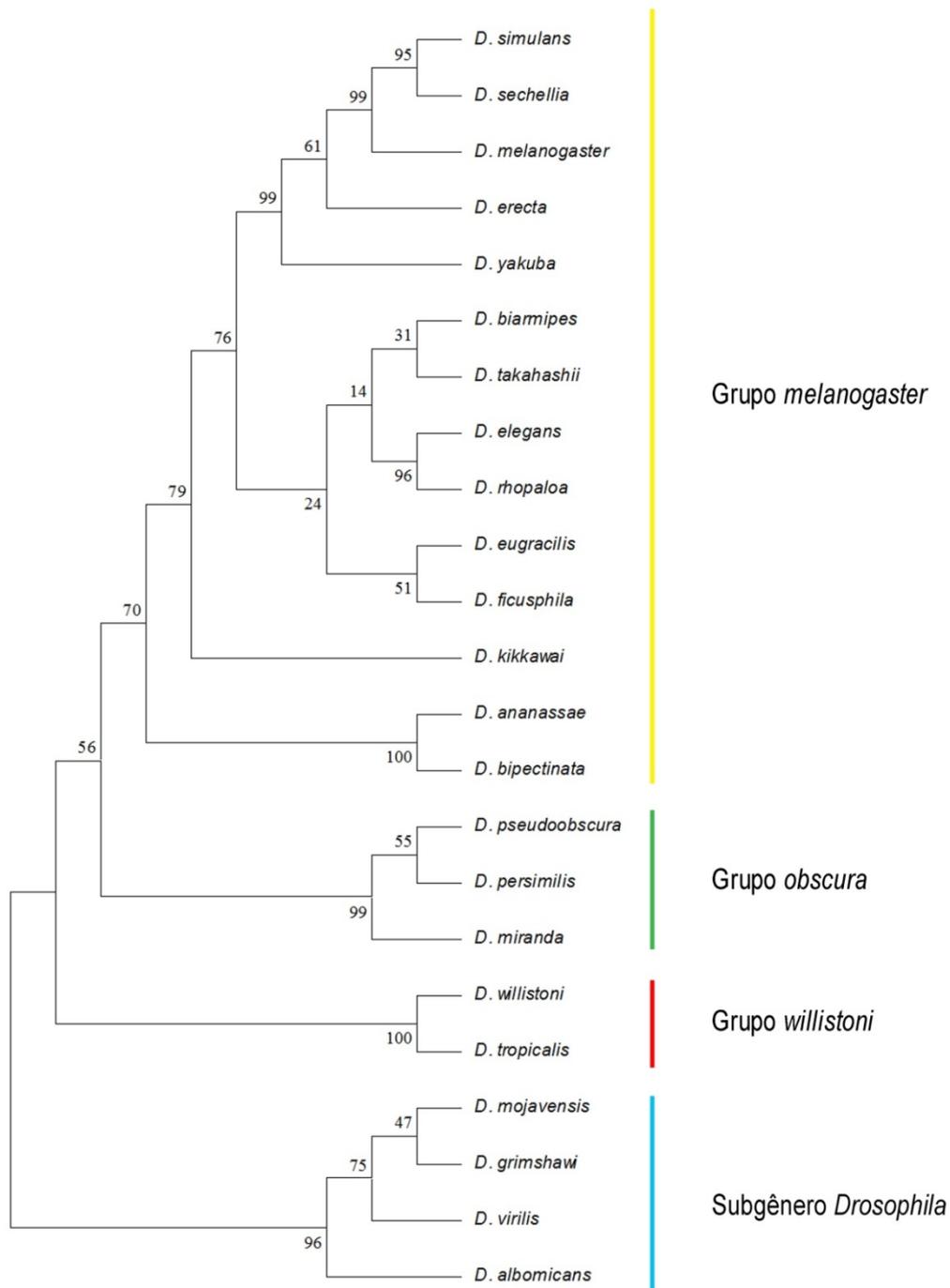


Figura 29. Análise filogenética pelo método de máxima verossimilhança das sequências de aminoácidos de DNMT2 de diferentes espécies de drosofilídeos. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas. O modelo evolutivo utilizado foi o de Le Gascuel (LG) com distribuição gama discreta com 4 categorias.

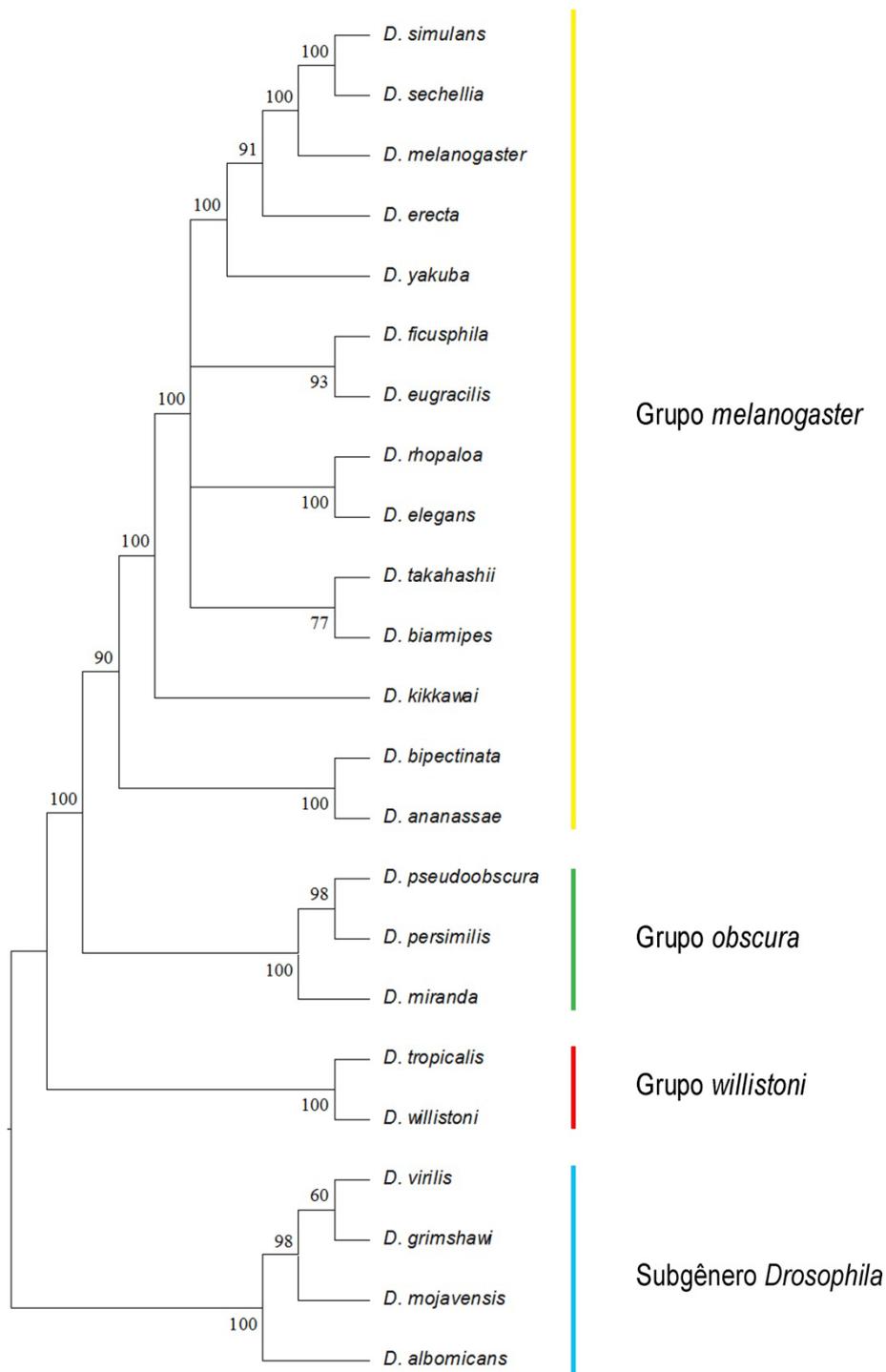


Figura 30. Inferência filogenética bayesiana das sequências de aminoácidos de DNMT2 das 23 espécies de drosofilídeos com genoma depositado no banco de dados Flybase, mais *D. tropicalis*. A árvore foi gerada utilizando o modelo evolutivo JTT com distribuição gama.

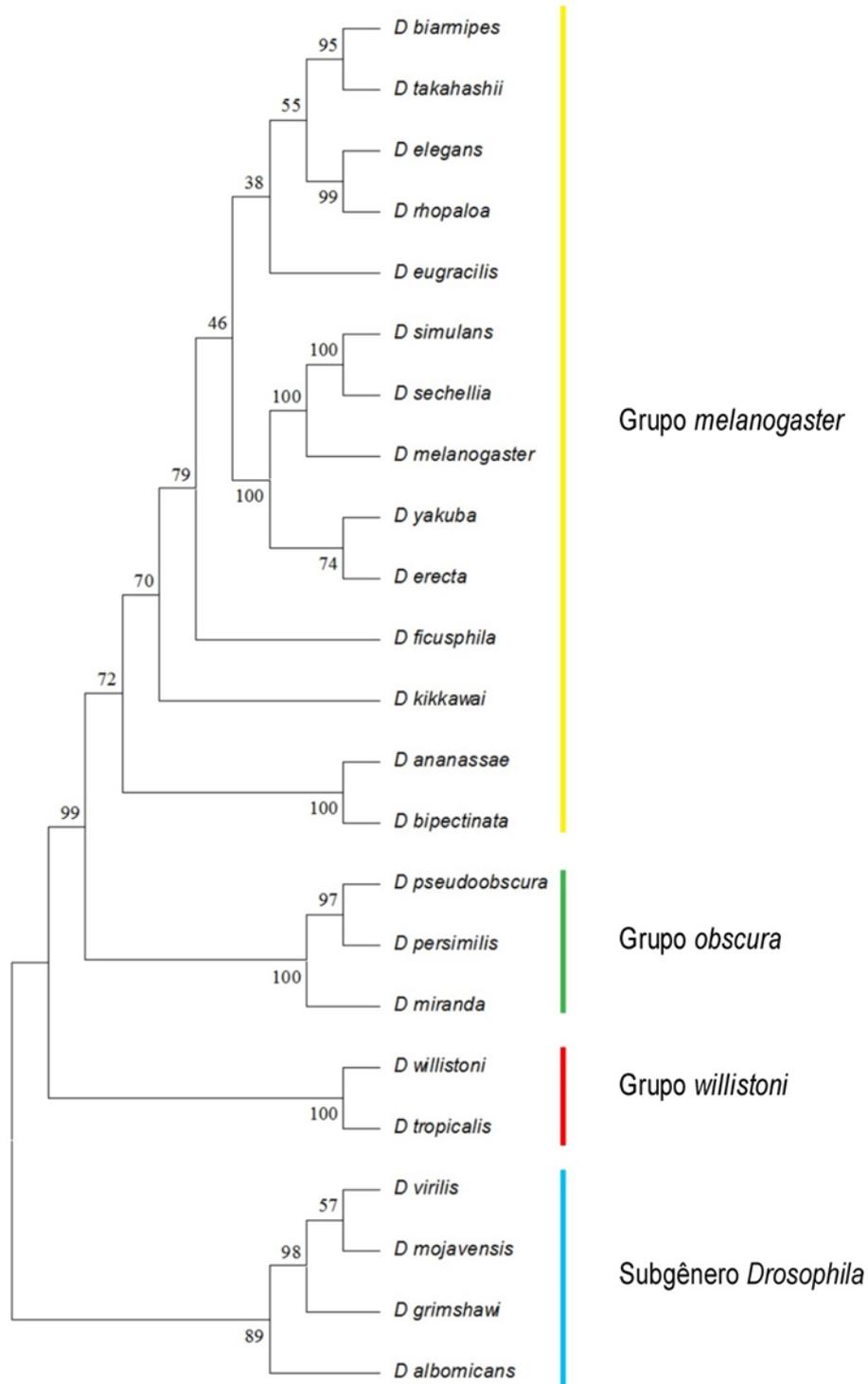


Figura 31. Análise filogenética pelo método de máxima verossimilhança das sequências de nucleotídeos de *Dnmt2*. A árvore consenso foi inferida a partir de um *bootstrap* de 10.000 réplicas. O modelo evolutivo utilizado foi o *General Time Reversible* (GTR) com distribuição gama discreta e alguns sítios invariáveis.

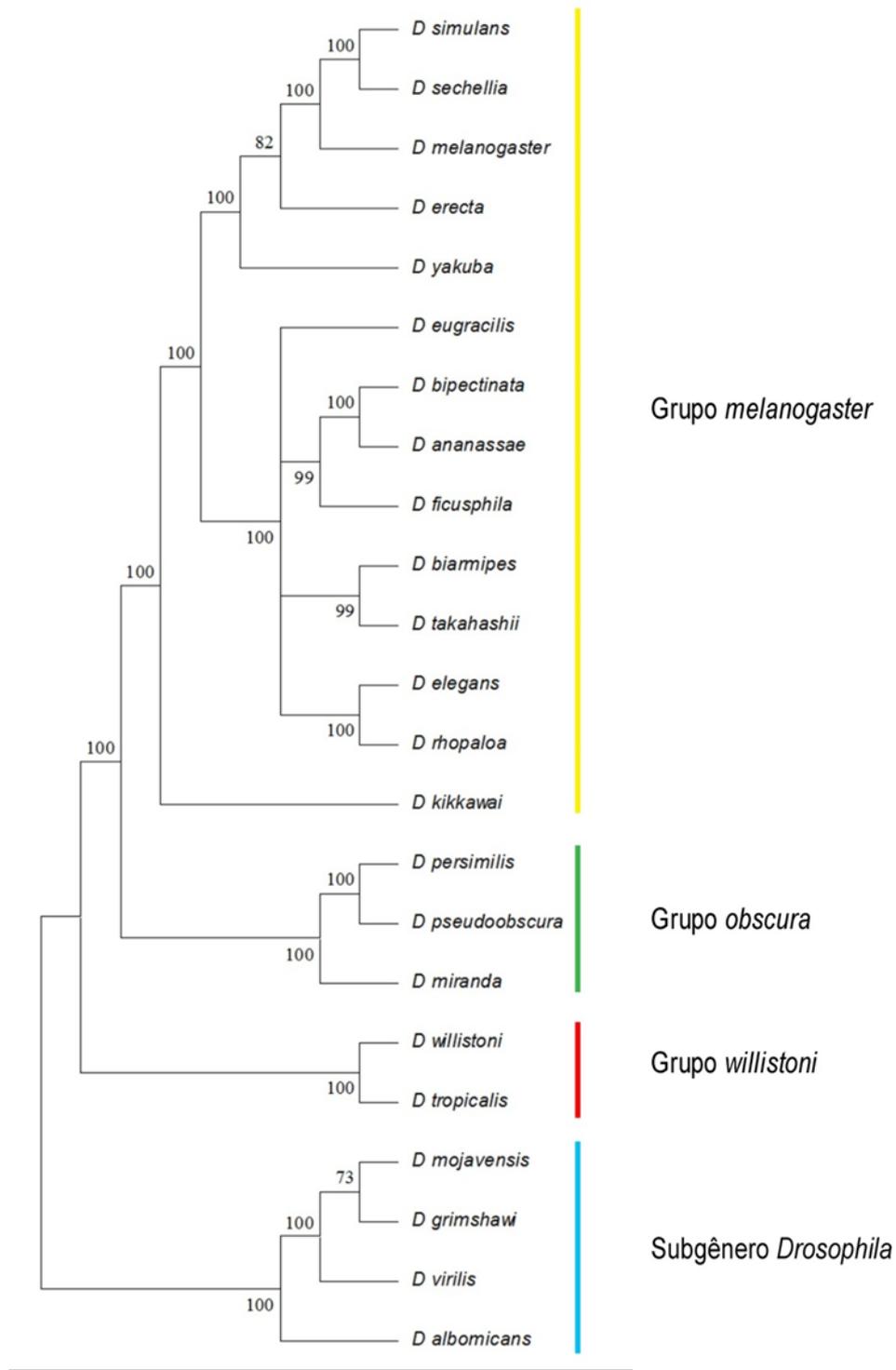


Figura 32. Inferência filogenética bayesiana das sequências de nucleotídeos de *Dnmt2* das 23 espécies de drosofilídeos com genoma depositado no banco de dados Flybase, mais *D. tropicalis*. A árvore foi gerada utilizando o modelo evolutivo General Time Reversible (GTR) com distribuição gama discreta e alguns sítios invariáveis.

CAPÍTULO IV

Imunodeteção da enzima DNMT2 em drosofilídeos do subgrupo *willistoni*.

Metodologia

Resultados e Discussão

4. Metodologia

4.1 Western blot

Western blot realizou-se com extratos proteicos de *D. willistoni*, *D. paulistorum*, *D. equinoxialis*, *D. tropicalis* e *D. melanogaster*, em diferentes fases de desenvolvimento: machos e fêmeas adultas, embriões de 0-6 horas, larvas e pupa. A extração de proteínas foi efetuado com tampão RIPA (150 mM NaCl, 0.1% Triton X-100, 0.5% desoxicolato de sódio, 0.1% SDS e 50 mM Tris-HCl pH 8.0) adicionando-se inibidores de protease à solução imediatamente. Em seguida as amostras foram maceradas, sendo mantidas no gelo para diminuir a ação das proteases.

Para a detecção, se utilizaram anticorpos policlonais de Dnmt2 (H-271) fornecidos pela Santa Cruz Biotechnology, inc., que possuem como epitopo de reconhecimento a sequência de aminoácidos 121 até 391 C-terminal da Dnmt2 de humanos. A membrana foi incubada com anticorpo primário numa diluição de 1:200. Para o anticorpo secundário, a diluição utilizada foi de 1:1000.

```
MEPLRVLELYSGVGGMHHALRESCIPAQVVAIDVNTVANEVYKYNFPHTQLLAKTIEGITLEEF
DRLSFDMLMSPPCQPFTRIGRQGDMTDSRTNSFLHILDILPRLQKLPKYILLENVKGFEVSSTR
DLLIQTIENCGFYQEFLLSPTSLGIPNSRLRYFLIAKLQSEPLPFQAPGQVLMFEPKIESVHPQK
YAMDVENKIQEKNVEPNISFDGSIQCSGKDAILFKLETAEEIHRKNQQDSDLSVKMLKDFLEDD
TDVNQYLLPPKSLRLRYALLLDIVQPTCRRSVCFTKGYGSYIEGTGSVLQTAEDVQVENIYKSLT
NLSQEEQITKLLILKRYFTPKEIANLLGFPPEFGFPEKITVKQRYRLLGNSLNVHVVAKLIKILYE
```

Figura 33. Sequência completa da Dnmt2 de humanos com a sequência imunógena em destaque.

Amostras com iguais concentrações de proteína foram separadas em gel de poliacrilamida-SDS 12%; transferência e revelação foram efetuados conforme procedimentos padrões de *western blot*.

4.2 Resultados e Discussão

As análises por *western blot* revelaram a expressão de DNMT2 na fase embrionária em espécies do subgrupo *willistoni* (*D. equinoxialis*, *D. paulistorum*, *D. willistoni* e *D. tropicalis*), assim como em *D. melanogaster* (**Figura 34**). Para analisar os padrões de expressão da DNMT2 ao longo do desenvolvimento do organismo, foram preparados extratos proteicos de diversos estágios de desenvolvimento como descrito anteriormente.

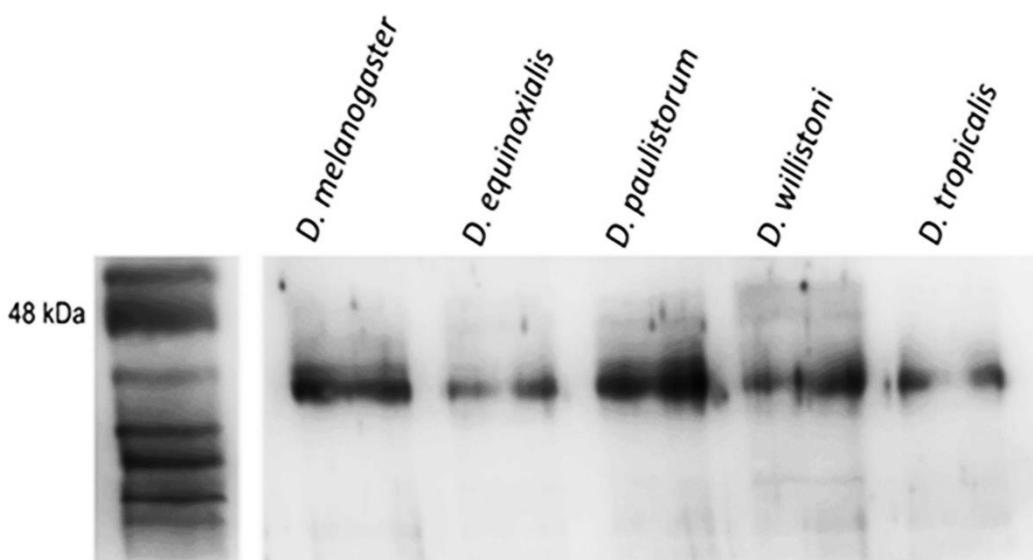


Figura 34. Padrões de expressão da DNMT2 em espécies do grupo *willistoni* de *Drosophila* na fase de embrião. Marcador indicando o peso molecular esperado para DNMT2 de drosofilídeos (entre 40 e 48 kDa, aproximadamente).

Não foram detectados níveis de expressão significativos de DNMT2 nas fases seguintes (larva e adultos) em *D. melanogaster* e *D. equinoxialis*. Entretanto, *D. willistoni*, *D. tropicalis* e *D. paulistorum* apresentaram níveis detectáveis de expressão de DNMT2 na fase de pupa (**Figura 34**). Kunert *et al.* (2003) encontraram quantidades significantes da proteína DNMT2 em *D. melanogaster*, mas somente em estágios embrionários. Nesse mesmo trabalho, o autor verifica que a depleção de DNMT2 nessa espécie não resultou em qualquer efeito deletério no desenvolvimento embrionário, apesar de registrar a perda

completa de citosinas metiladas no genoma do organismo.

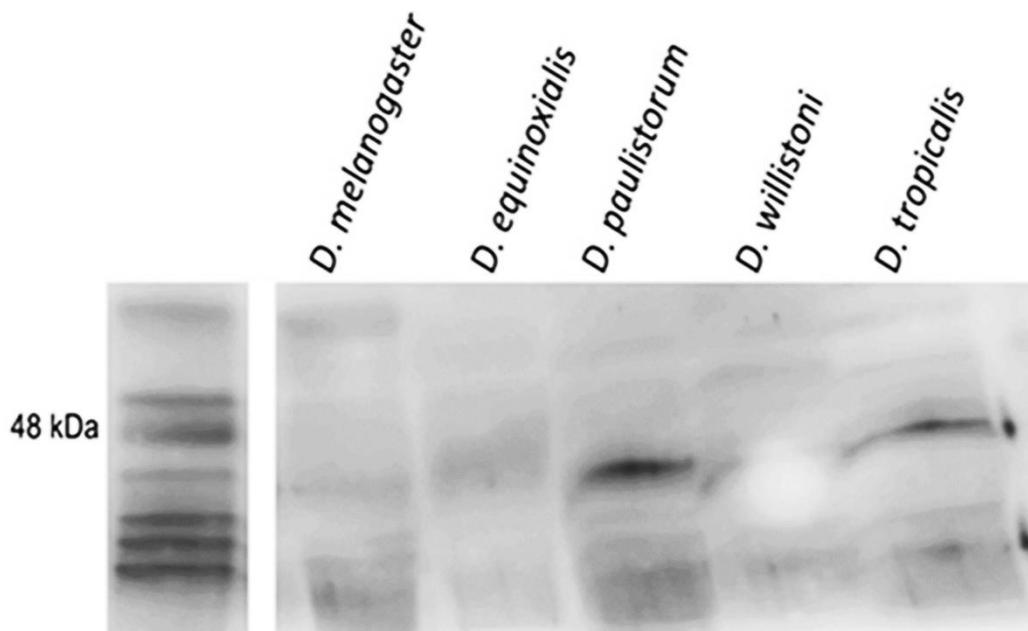


Figura 35. Padrões de expressão da DNMT2 em espécies do grupo *willistoni* de *Drosophila* na fase de pupa. Marcador indicando o peso molecular esperado para DNMT2 de drosofilídeos (entre 40 e 48 kDa, aproximadamente).

Em trabalho publicado em 2004, Marhold e colaboradores demonstraram grande conservação nas sequências de aminoácidos de DNMT2 em diversos dípteros, entre eles: *D. melanogaster*, *D. pseudoobscura*, *D. virilis*, *D. hydei* e *D. simulans*. Até então, todas as espécies de drosofilídeos analisadas apresentaram citosinas metiladas em seu genoma, mesmo que em baixos níveis. Lyko *et al.* (2000) demonstraram que os maiores níveis de metilação de citosina, em *D. melanogaster*, se concentram nos estágios embrionários iniciais, diminuindo ao longo do desenvolvimento. Além disso, os níveis mais altos de transcritos de *Dnmt2* estão presentes no tecido ovariano de fêmeas adultas. Dessa forma, as células *nurse* sintetizariam RNA maternal de *Dnmt2* para ser estocado no oócito em desenvolvimento.

Grande parte dos estudos epigenéticos em drosofilídeos está focada, principalmente, em espécies do grupo *melanogaster* e do subgênero *Drosophila*. A partir desses estudos, assume-se que tais eventos e padrões se estendem a

todas demais espécies de drosofilídeos. Entretanto, nosso grupo de pesquisa, que possui enfoque no conjunto de espécies Neotropicais já apresentou em diversos trabalhos, significativas diferenças de âmbito evolutivo, molecular e ecológico, com destaque para os grupos *cardini* e *willistoni* (Cordeiro *et al.*, 2014; Gonçalves *et al.* 2014; Pita *et al.*, 2014; Müller *et al.*, 2013; Garcia *et al.*, 2007). Entre as diversas questões que atraem a atenção para esse grupo de espécies, pode se citar o grande polimorfismo cromossômico, as diversas inversões existentes nos braços dos cromossomos politênicos (Da Cunha *et al.*,1950; Valente & Rohde, 2012) e a presença de padrões de metilação sexo-específica em espécies do subgrupo *willistoni* (D'Ávila *et al.*, 2010).

Os dados aqui apresentados são preliminares, necessitando maior aprofundamento e refinamento da técnica de extração de proteína e posterior imunodeteção. Entretanto, apresentam indícios importantes dentro do aspecto de fenômenos epigenéticos peculiares ao subgrupo *willistoni*, como a presença de níveis de expressão na fase de pupa em *D. willistoni*, *D. tropicalis* e *D. paulistorum*. Novos ensaios devem ser realizados para se poder fornecer um conjunto de dados suficientes para o entendimento dos mecanismos, fatores e funções que estão envolvidos na história evolutiva desses organismos.

CAPÍTULO V

Discussão geral

Conclusões

Perspectivas

5.1 Discussão Geral

O subgênero *Sophophora* compõe o gênero *Drosophila*, descrito por Sturtevant (1939), dividindo-se em dois grupos principais: espécies do Novo Mundo (Neotropicais) e espécies do Velho Mundo. As espécies do Velho Mundo estão distribuídas dentro dos grupos *melanogaster*, *setifemur*, *populi*, *obscura*, *fima*, *díspar* e *dentissima*. Já as espécies Neotropicais estão subdividas dentro dos grupos *saltans* e *willistoni*. Throckmorton (1975) classifica o subgênero *Sophophora* como parte de uma grande radiação de drosofilídeos que contem os gêneros *Chymomyza* e *Neotanygastrella*, colocando a radiação como basal e originária do que vem a ser o gênero *Drosophila*. Estudos biogeográficos descritos por Throckmorton (1975) indicam que o ancestral comum de *melanogaster* e *obscura* teve sua origem no sudeste asiático, a partir da metade do Oligoceno (por volta de 30-27 milhões de anos atrás), já Robe (2008) sugere que a separação ocorreu há aproximadamente 45 milhões de anos. Grande parte dos organismos que compõem as espécies do Velho Mundo habitam regiões tropicais, contudo há diversas espécies consideradas cosmopolitas (Lachaise et al., 1988). Já as espécies Neotropicais tem sua origem a partir de um ancestral *protomelanogaster* (Throckmorton, 1975) e teriam invadido a região Neotropical há cerca de 25-22 milhões de anos.

Diversas análises utilizando morfologia, DNA e filogenética indicam a monofilia do subgênero *Sophophora*. No presente trabalho, a partir das análises dos resíduos e da sequência de nucleotídeos de *Dnmt2*, essa monofilia é resgatada com valores robustos de suporte. Os resultados das análises filogenéticas também indicam monofilia fortemente suportada entre espécies Neotropicais e o clado *obscura-melanogaster*.

Ainda dentro do grupo de espécies Neotropicais, o subgrupo *willistoni* chama a atenção pelo seu elevado grau de polimorfismo cromossômico. Em *D. willistoni* foram registradas mais de 50 inversões cromossômicas em amostras coletadas do sul da Flórida até o sul da América do Sul, na região do rio

La Plata (Valente *et al.*, 2003). Quando analisadas as inversões cromossômicas, verificou-se uma correlação entre a geografia e a frequência dessas inversões. Populações que habitam regiões mais centrais do continente tendem a possuir uma alta frequência de inversões cromossômicas heterozigóticas, enquanto que as populações das áreas limítrofes de distribuição tendem a uma menor frequência dessas inversões. Segundo Da Cunha *et al.* (1959), parece existir íntima relação entre o nicho explorado e variabilidade genética desses organismos: maior variabilidade ecológica, maior polimorfismo populacional. Geralmente regiões centrais de distribuição populacional apresentam maior diversidade de nicho para exploração do que as regiões limítrofes. Entretanto, Valente *et al.* (2003) registraram alto grau de polimorfismo em *D. willistoni* coletados no Uruguai, o que vai de encontro ao que seria esperado de acordo com Da Cunha (1959). Inversões raras nas populações do Rio Grande do Sul acabaram sendo observadas com grande frequência em amostras coletadas no Uruguai. Esses e outros registros, como os estudos feitos no parque do Turvo e Itapuã, possibilitaram a Valente *et al.* (2003) sugerir que nesse grupo de organismos fatores bióticos e abióticos seriam primordiais para o estabelecimento de padrões diferentes nas frequências de inversões cromossômicas. Como regiões limítrofes de distribuição de uma população possuem condições ambientais menos estáveis, o polimorfismo cromossômico se torna uma ferramenta importante para sobrepujar as adversidades impostas pelo ambiente, sendo essa uma característica peculiar às espécies do subgrupo *willistoni*.

Contudo, o polimorfismo cromossômico presente em *D. willistoni* não se restringe apenas no contexto global de inversões, mas em como esses rearranjos se apresentam dentro do próprio conjunto cromossômico nas diferentes populações dessa espécie. Rohde & Valente (2012) sistematizaram as mais de 50 inversões encontradas nas três últimas décadas e verificaram que o cromossomo X e os autossomos parecem estar sobre pressões seletivas diferentes. Além do polimorfismo cromossômico, no nível molecular, *D. willistoni* também possui um alto percentual de polimorfismo enzimático, por volta de 60% (Ayala & Powell, 1972). Entre as espécies de *Drosophila*, *D. willistoni* também apresenta redução no desvio de uso de códons – que está relacionado com a

melhoria da síntese proteica - em relação aos demais drosofilídeos (Anderson *et al.*, 2003; Vicario *et al.*, 2007; Heger *et al.*, 2007). Essa redução também pode ser observada no presente trabalho, com o desvio da preferência de uso de códon pelas espécies pertencentes aos grupos *willistoni* em relação às espécies que constituem o grupo *melanogaster*.

Outra característica em *Drosophila* é a organização das regiões genômicas. O DNA satélite está, principalmente, localizado em regiões heterocromáticas. A heterocromatina de *D. melanogaster* é rica em sequências repetitivas em *tandem* (DNA satélite), grupos de genes rDNA, elementos relativamente repetitivos (Elementos Transponíveis - TEs) e alguns elementos de cópia única (Weiler & Wakimoto, 1995). Ashburner (1989), através de hibridização *in situ* verificou que DNAs satélites também podem estar presentes em regiões eucromáticas de *D. melanogaster*. A distribuição de TEs e sequências repetitivas nos diversos genomas de *Drosophila* variam bastante em relação aos percentuais encontrados: ~2.7% em *D. simulans*, ~16% em *D. willistoni* e ~25% em *D. ananassae* (Clark *et al.*, 2007). Essas sequências repetitivas podem estar associadas à organização da arquitetura dos genomas eucarióticos, formando compartimentos genômicos que podem, dinamicamente, se modificar em resposta a fatores transcricionais de ativação, ou desligamento de determinados genes por metilação, por exemplo. Nesse caso, a metilação de citosinas atua como um sistema de defesa do genoma frente às eventuais mobilizações de DNA invasores (TEs, por exemplo), ou no silenciamento de genes essenciais duplicados, que poderiam ser fatais na fase sexual destes organismos (Bender, 1998) – mecanismo de compensação de dose.

Além disso, sabe-se que fatores ecológicos, como clima e a quantidade de recursos disponíveis, por exemplo, podem ser desencadeadores de eventos estressantes, que, por conseguinte, induzem o aumento de mobilização de TEs no genoma do organismo. Além disso, é ampla a descrição de eventos de Transferência Horizontal (Loreto *et al.*, 2008). Os eventos de mobilização de TEs, por sua vez, podem estar ligados à ocorrência de quebras e rearranjos dos braços cromossômicos, modificações em sequências regulatórias

de genes, ou até mesmo a completa inativação de um gene. Da mesma forma, a mobilização de TEs cumpre importante papel na evolução molecular dos organismos, capaz de fomentar a emergência de novos genes, assim como novas vias na rede de sinalização celular. A forma como as transposições dos TEs são reguladas estão, principalmente, associadas aos mecanismos epigenéticos do organismo (Lerat *et al.* 2011).

A presença de metilação sexo-específica em espécies do subgrupo *willistoni* de *Drosophila* é outro fenômeno peculiar a esses organismos. A presença de metilação no genoma de *D. willistoni* foi primeiramente descrita por nosso grupo de pesquisa, em Garcia *et al.* (2007). Através da técnica MSRE (*Methylation-Sensitive Restriction Endonuclease*) foi possível detectar a metilação de citosinas em *D. willistoni*, além disso, os padrões de clivagem entre machos e fêmeas mostraram-se diferentes, o que não ocorre em *D. melanogaster*. As sequências oriundas do processo de clivagem de DNA por MSRE eram compostas por diferentes elementos do genoma e de rDNA. Posteriormente, D'Ávila *et al.* (2010) verificaram que *D. tropicalis* e *D. insularis* possuem os mesmos padrões de metilação sexo-específica de rDNA encontrados em *D. willistoni*. *D. paulistorum* e *D. equinoxialis* também apresentaram padrões de metilação sexo-específica, mas não em genes ribossomais. Interessantemente, o subgrupo *bocainensis* (grupo *willistoni*), que é considerado mais basal do que o subgrupo *willistoni* (Robe *et al.*, 2010), não possui metilação sexo-específica (D'Ávila *et al.* 2010), sugerindo que o peculiar fenômeno tenha surgido após a divergência dos subgrupos.

No presente estudo foi possível estabelecer um modelo preditivo de boa qualidade para as DNMT2 de drosofilídeos e para *M. musculus*. Com isso pode se analisar diferentes aspectos que estão envolvidos na funcionalidade das enzimas da família DNMT2, tendo como pano de fundo as suas relações evolutivas entre os diferentes grupos de espécies analisados. Como observado, as DNMT2 mantiveram-se conservadas estruturalmente em seu domínio catalítico, ao longo da evolução. Dentro do domínio catalítico, alguns motivos apresentam grande conservação, o que indica um processo de forte seleção purificadora

nesses sítios, eliminando-se qualquer variante que oferecesse a mínima diferença química nesses sítios (motivos ENV, dipeptídeo ProCys e o resíduo de arginina do motivo V, por exemplo, que são cruciais para a atividade catalítica da enzima). Entretanto, os resíduos que estabelecem a arquitetura do domínio catalítico apresentam-se com menor conservação de suas sequências, o que sugere que para essas regiões a seleção atua em nível estrutural. Kinch *et al.* (2002), apontam que as similaridades existentes nas estruturas terciárias das proteínas podem ocorrer sem que exista identidade das sequências, tal aspecto também pode ser observado mais claramente quando se compara os domínios da TRD das DNMT2.

Procariotos e eucariotos possuem baixa similaridade estrutural e de composição de resíduos da TRD. Essa baixa similaridade também é observada quando se analisa as proteínas dentro dos próprios grupos, como, por exemplo, entre as sequências e estruturas das DNMT2 de *H. haemolyticus*, *H. influenzae* e *M. penetrans*, onde cada uma das enzimas possui baixa conservação de resíduos da TRD e diferentes mecanismos de reconhecimento da sequência alvo a ser metilada. O mesmo parece ser verdadeiro para as DNMT2 de eucariotos, onde há marcante diferença entre a estrutura da TRD de protozoários (*E. histolytica*) e metazoários – além da baixa identidade e similaridade das sequências nessa região. Entre os metazoários, a aparente similaridade estrutural acaba contrastando com a grande divergência dos resíduos, o que se reflete no comportamento do domínio TRD ao longo das simulações de dinâmica molecular e a distribuição dos potenciais de superfície entre as moléculas.

As DNMT2 analisadas filogeneticamente no capítulo II resgatam as relações evolutivas das espécies envolvidas. Porém, através do refinamento dos alinhamentos, com a adição das informações a respeito de como se distribuem os resíduos nas estruturas terciárias, houve melhorias na resolução dos ramos evolutivos. A posição do ramo de *S. frugiperda* que possui 43% de identidade de aminoácidos com *D. melanogaster* e 40% com *M. musculus*, por exemplo, teve seu valor de suporte aumentado, quando inseridas as informações de homologia dos sítios levando-se em conta a estrutura 3D da proteína.

Outro dado interessante, é que o gene *Dnmt2* pode ser de grande utilidade como marcador molecular a fim de descrever, ou dar maior suporte para as relações filogenéticas de diferentes níveis taxonômicos – tanto espécies próximas como distantes, evolutivamente. Isto se percebe quando foram feitas as análises filogenéticas dos diferentes grupos de drosofilídeos, no capítulo III, em que foram resgatadas as relações evolutivas dos principais grupos do subgênero *Sophophora* e a monofilia com o subgênero *Drosophila*, conforme já descrito na literatura (Robe *et al.* 2010). Porém, o gene *Dnmt2* não serviria como um marcador universal, visto que algumas linhagens de organismos acabaram perdendo o gene ao longo da evolução.

Contudo, apesar das DNMT2 de drosofilídeos guardarem relações evolutivas próximas e apresentarem divergência de aminoácidos na ordem de 25%, em média, os dados oriundos dos potenciais eletrostáticos de superfície permitem inferir diferenças significativas nas áreas de cargas positivas e negativas das moléculas, em seus domínios: catalítico e de reconhecimento de substrato. Interessantemente, *D. willistoni* e *D. tropicalis* foram as que apresentaram um maior distanciamento de similaridade de potenciais com as demais espécies de drosofilídeos na matriz do epograma. A análise visual da distribuição de cargas de superfície das DNMT2 de drosofilídeos revelou uma maior área com perfil eletrostático positivo na TRD das duas espécies Neotropicais. Considerando que espécies do subgrupo *willistoni* apresentam fenômenos epigenéticos de metilação sexo-específico de citosinas do genoma (presente apenas em drosofilídeos pertencentes a este subgrupo), que apresentam um elevado grau de polimorfismos, além de uma ampla área de distribuição, onde as diversas populações são expostas a diferentes cenários ecológicos, pode se esperar que algum grau de modulação diferencial da função da DNMT2 em direção a uma maior afinidade à molécula de DNA tenha sido favorecido.

Assim, espera-se que ocorram nas espécies do subgrupo *willistoni* mecanismos que garantam a estabilidade genômica em níveis adequados, mas que ao mesmo tempo possuam plasticidade suficiente para permitir rápida

capacidade adaptativa em ambientes diversos. Isso parece ter ocorrido ao longo da evolução das espécies pertencentes a esse grupo. Porém, a proposição não exclui a possibilidade de ação da DNMT2 de espécies do subgrupo *willistoni* sobre moléculas de tRNA, ou na metilação de transcritos de TEs, justamente por preservarem uma arquitetura muito semelhante às DNMT2 de vertebrados, que possuem *in vivo* a função de metilação de citosinas de diferentes tRNAs.

Os estudos sobre as diversidades funcionais das macromoléculas podem levar a uma maior compreensão da diversidade biológica existente na natureza. Os aspectos macro (molécula como um todo e interações com outras moléculas) e micro (características físico-químicas, distribuição de potenciais eletrostáticos, comportamento em meio fisiológico e conservação de resíduos) são fundamentais para entendimento de quais e de que maneira diferentes mecanismos evolutivos possam ser cruciais na preservação ou na aquisição de novas funções de uma molécula. No caso das espécies Neotropicais do subgrupo *willistoni*, sugere-se que possa haver um entremeado sistema de fatores intrínsecos (presença de TEs e polimorfismo cromossômico, por exemplo) e extrínsecos (distribuição geográfica e nichos variáveis) que estariam atuando como forças evolutivas no surgimento e manutenção das diferenças observadas na DNMT2.

5.2 Conclusão

O presente trabalho apresenta um modelo estrutural de qualidade para as DNMT2 de *Drosophila melanogaster* e *D. willistoni*. A utilização de mais de uma técnica de construção de modelos e a utilização de diferentes validadores contribuem para um melhor desempenho no processo de modelagem de proteínas. Além de possibilitar reconhecimento de sítios importantes para o entendimento da funcionalidade biológica de uma proteína, o desenvolvimento de modelos de alta qualidade proporciona a chance de estudo de análises evolutivas comparadas de âmbito tridimensional.

As análises entre as diferentes DNMT2 (*Haemophilus haemolyticus*, *Haemophilus influenzae*, *Mycoplasma penetrans*, *Entamoeba histolytica*, *Drosophila willistoni*, *D. melanogaster*, *Spodoptera frugiperda*, *Mus musculus* e humana) demonstraram a marcante conservação do domínio catalítico das enzimas, sobretudo nos motivos envolvidos na atividade de metilação. Entretanto, não há grande conservação dos demais resíduos constituintes do motivo catalítico, o que sugere um processo de seleção em nível estrutural, nesses casos. Por outro lado, a região responsável pelo reconhecimento da citosina-alvo a ser metilada (TRD) apresenta grande diversidade tanto em suas sequências primárias, quanto na arquitetura do domínio. Mesmo em metazoários, que possuem arquitetura bastante semelhante nessa região, a composição de resíduos resulta na emergência de outro nível de diferenciação entre as enzimas: a distribuição de cargas na superfície da molécula.

Estudos de similaridade de potencial eletrostático de superfície apresentaram características marcadamente diferentes até mesmo entre DNMT2 muito próximas evolutivamente e com arquiteturas muito semelhantes. As DNMT2 de vertebrados diferenciam-se significativamente das DNMT2 de invertebrados em suas distribuições de potenciais eletrostáticos. Comparativamente, *S. frugiperda* e drosofilídeos também possuem dissimilaridades destacadas. As simulações de dinâmica molecular também dão suporte às diferenças

encontradas entre as 9 estruturas de DNMT2, onde enzimas que parecem ter dualidade de substrato (DNA e/ou tRNA) apresentam uma maior mobilidade entre seus domínios, enquanto aquelas que tem preferência por apenas um substrato, comportam-se de maneira mais estável ao longo da dinâmica.

De posse de modelos estruturais de qualidade de *D. melanogaster* e *D. willistoni*, foi possível realizar estudos mais aprofundados nas características evolutivas e funcionais das DNMT2 de drosofilídeos. No aspecto evolutivo, drosofilídeos apresentam grande conservação das sequências de resíduos, mas com maior variabilidade apenas na TRD. Inferências filogenéticas colocam *Dnmt2* como um possível marcador molecular para os estudos de evolutivos em drosofilídeos, resgatando com bom grau de robustez as relações evolutivas entre os diversos clados que compõem o gênero *Drosophila* – com destaque à confirmação da monofilia existente entre os subgêneros *Drosophila* e *Sophophora*.

Finalmente, cabe destacar que *D. willistoni* e *D. tropicalis* foram as espécies que exibiram menor similaridade de distribuição de cargas de superfície entre as 23 espécies analisadas, apresentando uma maior densidade de cargas positivas em suas TRDs. Assim, resultados por nós obtidos vão ao encontro das diversas evidências já descritas sobre as peculiaridades existentes em nível ecológico, citogenético, epigenético, evolutivo e molecular do conjunto de espécies que compõem o grupo *willistoni*, mais especificamente, o subgrupo *willistoni* de *Drosophila*. Nossos achados também lançam luz para futuros trabalhos que contribuam para a maior compreensão de como os diferentes níveis envolvidos na evolução dessas espécies estão interligados.

5.3 Perspectivas

Ao longo dos últimos anos, diversos estudos com intuito de entendimento dos processos e mecanismos envolvidos na atividade catalítica de metilação das DNMT2 foram desenvolvidos. A grande parte está focada na reação de metilação, propriamente dita, mas poucos abordam a dinâmica envolvida no reconhecimento e interação da enzima com o substrato (DNA e/ou tRNA), havendo apenas alguns trabalhos com DNMT2 de procariotos (*H. haemolyticus*, *H. influenzae* e *M. penetrans*). Dessa forma, com o advento de novos recursos tecnológicos, com programas que possibilitam simular as interações entre enzima e substrato, abre-se a oportunidade de ampliar as abordagens dos estudos sobre as relações existentes entre DNMT2 e os diferentes substratos em que a enzima pode atuar.

No presente trabalho desenvolveram-se modelos de DNMT2 de duas espécies de drosofilídeos (*Drosophila melanogaster* e *D. willistoni*) e verificou-se que elas possuem características diferentes na distribuição de potenciais de superfície. Essa característica pode estar envolvida com afinidades diferentes entre as duas enzimas em relação ao DNA. Assim, como perspectiva, pretende-se avaliar simulações destas enzimas a partir de interações tanto com a molécula de DNA como a molécula de tRNA. Além disso, pretende-se realizar novos estudos buscando isolar e caracterizar as regiões do genoma metiladas diferencialmente entre machos e fêmeas de espécies do subgrupo *willistoni*. Certamente, os resultados advindos, contribuirão de maneira significativa para a compreensão dos fatores evolutivos envolvidos nesse fenômeno.

5.4 Referências

Anderson CL, Carew EA and Powell JR (1993) Evolution of the *Adh* locus in the *Drosophila willistoni* group: the loss of an intron, and shift in codon usage. *Mol Biol Evol.*10(3):605-618.

Arnold K, Bordoli L, Kopp J and Schwede T (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22:195-201.

Ashburner, M and Novitski E (1989) The polytene chromosomes. Chapter 2 in *The Genetics and Biology of Drosophila*. Academic press. Vol. 1a: 31-66.

Ayala FJ and Powell JR (1972) Allozymes as diagnostic character of sibling species of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 69: 1094-1096.

Bachman KE, Rountree MR and Baylin SB (2001) DNMT3a and DNMT3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J Biol Chem* 276: 32282–32287.

Baker NA, Sept D, Joseph S, Holst MJ and McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98, 10037-10041.

Banerjee S, Fisher O, Lohia A and Ankri S (2005) *Entamoeba histolytica* DNA methyltransferase (EhMeth) is a nuclear matrix protein that binds EhMRS2, a DNA that includes a scaffold/matrix attachment region (S/MAR). *Mol Biochem Parasitol* 139:91–97.

Bender J (1998). Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. *TIBS*, 23: 252-256.

Benkert P, Biasini M and Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3):343-50.

Bhabha G, Ekiert DC, Jennewein M, Zmasek CM and Tuttle LM (2013) Divergent evolution of protein conformational dynamics in dihydrofolate reductase *Nature*

Structural & Molecular Biology 20, 1243–1249

Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16: 6-21.

Blomberg N, Gabdoulline RR, Nilges M and Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins: Str., Function and Genetics*. 37: 379-387.

Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 12;253(5016):164-70.

Campbell JD, Deol SS, Ashcroft, FM, et al. (2004) Nucleotide dependent conformational changes in HisP: molecular dynamics simulations of an ABC transporter nucleotide-binding domain. *Biophys J*, v.87, n. 6, pp. 3703- 15.

Cavasotto CN and Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. *Drug discovery today* 14:676-683.

Clark AG, Eisen MB, Smith DR, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 450:203-218.

Cordeiro J, De Toni DC and Valente VLS (2014) Chromosomal evolution in the *Drosophila cardini* group (Diptera: Drosophilidae): photomaps and inversion analysis. *Genetica*. 142(5):461-72.

D'Ávila MF, Garcia RN, Panzera Y and Valente VLS (2010) Sex-specific methylation in *Drosophila*: an investigation of the Sophophora subgenus. *Genetica* 138:907-913.

Da Cunha AB, Burla H and Dobzhansky T (1950) Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution* 4:212-235.

D'Alfonso G, Tramontano A and Lahm A (2001) Structural conservation in single-domain proteins: implications for homology modeling. *J Struct Biol* 134:246-256.

D'Ávila MF (2011) Estudo da DNA metiltransferase 2 (DNMT2) e Metilação do DNA no Gênero *Drosophila*. 105 f. Tese (Doutorado) - Departamento de Genética,

Universidade Federal do Rio Grande do Sul, Porto Alegre.

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM and Notredame C (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39 (Web Server issue), W13–W17. doi:10.1093/nar/gkr245.

Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH and Cheng X (2001) Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturantresistant binding to DNA. *Nucleic Acids Res* 29:439-448.

Dozmorov MG, Giles CB, Koelsch KA and Wren JD (2013) Systematic classification of non-coding RNAs by epigenomic similarity. *BMC Bioinformatics*. 14 Suppl 1(Suppl 14):S2. doi:10.1186/1471-2105-14-S14-S2..

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. doi: 10.1093/nar/gkh340.

Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U and Sali A (2006) Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc, Supplement 15, 5.6.1-5.6.30.

Farthing CR. et al. (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet*. 4, e1000116

Feinberg AP and Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301, 89-92.

Fiser A, Do RK and Sali A (2000) Modeling of loops in protein structures. *Protein Science* 9:1753-1773.

Fisher O, Siman-Tov R and Ankri S (2004) Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res* 32:287–297.

Fuks F, Burgers WA, Brehm A (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature* 24:1–4.

Garcia RN, D'Avila MF, Robe LJ, Loreto ELS, Panzera Y, Heredia FO and Valente VLS (2007) First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica* 131:91–105.

Goldsmith-Fschman S and Honig B (2003) Structural genomics: computational methods for structure analyses. *Protein Science*. v. 12, p. 1813-1821.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, et al. (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog DNMT2. *Science* 311: 395–398.

Gonçalves JW, Valiati VH, Delprat A, Valente VL and Ruiz A (2014) Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome. *BMC Genomics*.15:792.

Gowher H, Leismann O and Jeltsch A (2000) DNA of *Drosophila melanogaster* contains 5-methylcytosine. *The EMBO journal*. vol 19. 24: 6918-6923.

Grishin NV (2001) Fold Change in Evolution of Protein Structures. *Journal of Structural Biology* 134, 167–185.

Han H, Cortez CC, Yang X, Nichols PW, Jones PA and Liang G (2011) DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet* 20:4299–4310. doi: 10.1093/hmg/ddr356.

Hardy TM and Tollefsbol TO (2011) Epigenetic diet: impact on the epigenome and cancer. *Epigenomics* 3:503–518. doi: 10.2217/epi.11.71.

Heger A and Ponting CP (2007) Variable strength of translational selection among 12 *Drosophila* species. *Genetics*. 177:1337-1348.

Hermann A, Schmitt S, Jeltsch A (2003) The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J Biol Chem* 278: 31717–31721.

Hillisch A, Pineda LF and Hilgenfeld R (2004) Utility of homology models in the

drug discovery process. *Drug Discov Today* 9:659-669.

Humphrey W, Dalke A and Schulten K (1996) VMD - Visual Molecular Dynamics. *J. Molec. Graphics*, 1996, vol. 14, pp. 33-38.

Hung MS, Karthikeyan N, Huang B, Koo HC, Kiger J, et al. (1999) *Drosophila* proteins related to vertebrate DNA (5-cytosine) methyltransferases. *Proc Natl Acad Sci USA* 96: 11940–11945.

Jeltsch A, Nellen W, Lyko F (2006) Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 31: 306– 308.

Johnson ML, Holt JM and Ackers GK (2009) *Methods in enzymology: Biothermodynamics, Part B. Examining Ion Channel Properties Using Free-Energy Methods* by Carmen Domene and Simone Furini. London: Elsevier, xiv + 676 pages., 466: 157-159.

Jones DT, Taylor WR and Thornton J.M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 8: 275-282.

Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202.

Jurkowski TP and Jeltsch A (2011) On the Evolutionary Origin of Eukaryotic DNA Methyltransferases and DNMT2. *PLoS ONE* 6(11): e28104. doi:10.1371/.

Kakumani PK, Malhotra P, Mukherjee SK and Bhatnagar RK. (2014) A draft genome assembly of the army worm, *Spodoptera frugiperda*. *Genomics* 104: 134–143.

Kemena C and Notredame C (2009) Upcoming challenges for multiple sequence alignment methods in the high-throughput era. *Bioinformatics.* 25:2455–2465.

Keshet I, Lieman-Hurwitz J and Cedar H (1986). DNA methylation affects the formation of active chromatin. *Cell* 44: 535-543.

Kiani J, Grandjean V, Liebers R, Tuorto F, Ghanbarian H, et al. (2013) RNA–

Mediated Epigenetic Heredity Requires the Cytosine Methyltransferase DNMT2. PLoS Genet 9(5): e1003498. doi:10.1371/journal.pgen.1003498.

Kinch LN and Grishin N V (2002) Evolution of protein structures and functions. Curr Opin Struct Biol 12:400–8.

Krauss V and Reuter G (2011) DNA methylation in *Drosophila*--a critical evaluation. Prog Mol Biol Transl Sci 101:177–91. doi: 10.1016/B978-0-12-387685-0.00003-2.

Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, Eichinger L, Lyko F, Ambros V, So F et al. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. 33:6405–6417. doi: 10.1093/nar/gki952.

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ and Wilson G (1994) The DNA (cytosine-5) methyltransferases. Nucleic Acids Research, 22: 1-10.

Kumar, A. et al. (2008). Chromatin remodeling is a key mechanism underlying cocaine-induced plasticity in striatum. Neuron 48 (2): 303-314

Kunert N, Marhold J, Stanke J, Stach D and Lyko F (2003) A DNMT2-like protein mediates DNA methylation in *Drosophila*. Development 130: 5083–5090.

Laurent L et al. (2010) Dynamic changes in the human methylome during differentiation. Genome Res. 20, 320–331.

Lazaridis T, Karplus M. (1999) Discrimination of the native from misfolded protein models with an energy function including implicit solvation J Mol Biol;288:477–487

Lerat E, Burlet N, Biémont C and Vieira C (2011) Comparative analysis of transposable elements in the *melanogaster* subgroup sequenced genomes. Gene. 473(2):100-109.

Lin MJ, Lin-Ya T, Reddy MN and Shen CKJ (2005) DNA methyltransferase gene dDNMT2 and longevity of *Drosophila*. J Biol Chem 280:861-4.

Liu K, Wang YF, Cantemir C and Muller MT (2003) Endogenous assays of DNA methyltransferases: Evidence for differential activities of DNMT1, DNMT2, and

DNMT3 in mammalian cells in vivo. *Mol Cell Biol* 23: 2709–2719.

Lo P-K and Sukumar S (2009) Epigenomics and breast cancer. *Pharmacogenomics* 9:1879–1902. doi: 10.2217/14622416.9.12.1879.Epigenomics.

Loreto EL, Carareto CM and Capy, P (2008). Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 100, 545–554.

Lupas a N, Ponting CP and Russell RB (2001) On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J Struct Biol* 134:191–203. doi: 10.1006/jsbi.2001.4393.

Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature*, 5;356(6364):83-5.

Lyko F, Whittaker AJ, Orr-Weaver TL and Jaenisch R (2000) The putative *Drosophila* methyltransferase gene dDNMT2 is contained in a transposon-like element and is expressed specifically in ovaries. *Mech Dev* 95:215–217.

Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct* 29, 291-325, 2000.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F and Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291-325.

McGuffin, L. J. (2008) The ModFOLD Server for the Quality Assessment of Protein Structural Models. *Bioinformatics*, 24, 586-7.

Müller MJ, Dörr NC, Deprá M, Schmitz HJ, Valiati VH and Valente VL (2013) Reevaluating the infection status by the *Wolbachia* endosymbiont in *Drosophila* Neotropical species from the willistoni subgroup. *Infect Genet Evol.* 19:232-9

Nan X, Ng HH, Johnson C a, Laherty CD, Turner BM, Eisenman RN and Bird a (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2

involves a histone deacetylase complex. *Nature* 393:386–9. doi: 10.1038/30764

O'Gara M, Roberts RJ and Cheng X (1996) A structural basis for the preferential binding of hemimethylated DNA by HhaI DNA methyltransferase. *J Mol Biol* 263:597–606. doi: 10.1006/jmbi.1996.0601.

O'Sullivan O, Suhre K, Abergel C, Higgins DG and Notredame C (2004) 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *Journal of Molecular Biology* 340, 385–95.

Okano M, Bell DW, Haber D a and Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–57.

Okano M, Xie S and Li E (1998) DNMT2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 26:2536-2540.

Ooi SK et al. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714–717.

Ørom UA and Shiekhattar R (2011) Long non-coding RNAs and enhancers. *Curr Opin Genet Dev* 21:194–198. doi: 10.1016/j.gde.2011.01.020.

Park YJ, Claus R, Weichenhan D and Plass C (2011) Genome-wide epigenetic modifications in cancer. *Prog Drug Res* 67:25–49. doi: 10.1007/978-3-7643-8989-5_2.

Patel CV and Gopinathan KP (1987) Determination of trace amounts of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal Biochem* 164-169.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TEJ (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Comput Chem.* 25(13):1605-12.

Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G (2009)

Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet* 41:696-702.

Pikkemaat MG, Linssen ABM, Berendsen HJC and Janssen DB (2002) Molecular dynamics simulations as a tool for improving protein stability. *Protein Eng* 15:185–192. doi: 10.1093/protein/15.3.185.

Pita S, Panzera Y, Valente VLS, Garcia C, Garcia AC, Montes MA and Rohde C (2014) Cytogenetic mapping of the Muller F element genes in *Drosophila willistoni* group. *Genetica*. 142(5):397-403.

Reinisch KM, Chen L, Verdine GL and Lipscomb WL (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: An extrahelical cytosine and rearranged base pairing. *Cell*, vol. 82, 143-153.

Robe LJ, Cordeiro J, Loreto EL and Valente VL. (2010) Taxonomic boundaries, genetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). *Genetica*. 138(6):601-17

Rohde C and Valente VLS (2012) Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. *Genet Mol Biol*. 35(4 Suppl): 966–979

Ronquist F and Huelsenbeck JP (2003) MrBayes version 3.0: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572-1574.

Rössle SCS (2004) Desenvolvimento de um Sistema Computacional para a Modelagem Comparativa em Genômica Estrutural: Análise de Seqüências do Genoma da *Gluconacetobacter diazotrophicus*. D.Sc., Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

Roy A, Kucukural A and Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–38. doi: 10.1038/nprot.2010.5.

Roy A, Kucukural A and Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, vol 5, 725-

738.

Roy A, Yang J and Zhang Y (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* 40:W471–7. doi: 10.1093/nar/gks372.

Roy A, Yang J and Zhang Y (2012) COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research*, 40: W471-W477.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425

Sakamoto Y, Ishiguro M and Kitagawa G (1986). Akaike information criterion statistics. D. Reidel Publishing Company, Dordrecht, The Netherlands.

Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol* 234:779-815.

Sánchez R and Sali A (1998) Comparative protein modelling in genomics. *Journal of Computational Physics*. n. 151, p. 388-401.

Sankpal UT, Rao DN (2002) Structure, function, and mechanism of HhaI DNA methyltransferases. *Crit Rev Biochem Mol Biol* 37(3):167–197.

Santos-Filho OA., Alencastro RB (2003) Modelagem de proteínas por homologia. *Quim Nov* 26:253–259.

Schaefer M, Steringer JP and Lyko F (2008) The *Drosophila* cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS One* 3:e1414. doi: 10.1371/journal.pone.0001414.

Schulz EC, Roth HM, Ankri S and Ficner R (2012) Structure analysis of *Entamoeba histolytica* DNMT2 (EhMeth). *PLoS One* 7:e38728. doi: 10.1371/journal.pone.0038728.

Schwede T, Kopp J, Guex N and Peitsch Mc (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385

Sharma S, De Carvalho DD, Jeong S, Jones PA and Liang G (2011) Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS Genet.* 7, e1001286.

Sippl M (1995) Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 1995;5:229–235.

Spitale RC, Tsai MC and Chang HY (2011) RNA templating the epigenome: Long noncoding RNAs as molecular scaffolds. *Epigenetics* 6:539–543. doi: 10.4161/epi.6.5.15221.

Sturtevant AH (1939). On the subdivision of the genus *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 25, 137–141.

Sweatt JD (2009). Experience-dependent epigenetic modifications in the central nervous system. *Biological Psychiatry*, 65: 191-197.

Tamura K, Stecher G, Peterson D, Filipinski A and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30:2725-2729.

Tang LY, Reddy MN, Rasheva V, Lee TL, Lin MJ, et al. (2003) The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases. *J Biol Chem* 278: 33613–33616.

Thiagarajan D, Dev R and Khosla S (2011) The DNA methyltransferase DNMT2 participates in RNA processing during cellular stress. *Epigenetics* 6:1, 103-113.

Tovy A, Siman Tov R, Gaentzsch R, Helm M and Ankri S (2010) A New Nuclear Function of the *Entamoeba histolytica* Glycolytic Enzyme Enolase: The Metabolic Regulation of Cytosine-5 Methyltransferase 2 (DNMT2) Activity. *PLoS Pathog* 6(2): e1000775. doi:10.1371/journal.ppat.1000775.

Tsankova N, Renthal W, Kumar A and Nextler EJ (2007). Epigenetic regulation in psychiatric disorders. *Nature Reviews Neuroscience*, 8: 355-367.

Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M,

Stoecklin G and Lyko F (2012) RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol* 19:900–5. doi: 10.1038/nsmb.2357.

Tweedie S, Ng HH, Barlow a L, Turner BM, Hendrich B and Bird a (1999) Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat Genet* 23:389–90. doi: 10.1038/70490.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701-1718.

Vicario S, Moriyama EN and Powell JR (2007) Codon usage in twelve species of *Drosophila*. *BMC Evol Biol.* 7:226.

Vriend G and Sander C (1993) Quality control of protein models: Directional atomic contact analysis. *J.Appl.Cryst.* 26, 47-60.

Waddington CH (1942) The epigenotype. *Endeavour* 1: 18–20.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL and Schübeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37:853–862. doi: 10.1038/ng1598.

Weiler KS and Wakimoto BT (1995) Heterochromatin and expression in *Drosophila*. *Annu. Rev. Genet.* 29: 577-605.

Wojciechowski M, Czapinska H and Bochtler M (2013) CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.Mpel. *Proc Natl Acad Sci U S A* 110:105–10. doi: 10.1073/pnas.1207986110.

Wyszynski MW, Gabbara S, Kubareva EA, Romanova EA, Oretskaya TS, Gromova ES, Shabarova ZA and Bhagwat AS (1993) The cysteine conserved among DNA cytosine methylases is required for methyl transfer, but not for specific DNA binding. *Nucleic Acids Res.* 21(2):295–301.

Xiang, Z (2006) Advances in Homology Protein Structure Modeling. *Curr Protein*

Pept Sci.; 7(3): 217–227.

Yoder JA and Bestor TH (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet* 7: 279–284.

Yoder JA, Walsh CP and Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340.

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, vol 9, 40.

Zhang Y and Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins.* 57(4):702-10.